Chapter 2
Standardized Assessment of Cognitive Development: Instruments and Issues

Kirsten M. Ellingsen

Abstract A standardized measure of cognitive functioning is often a primary component of a comprehensive early childhood psychological evaluation. Children's performance on cognitive assessments can have significant immediate and long-term implications. Treatment decisions and access to services or resources may be predicated on the information gained from a particular measure. This chapter will describe the applications of standardized instruments commonly used to measure cognitive development in infants, toddlers, and young children. It begins with a discussion of cognitive development theory to provide a foundation of the background and structure of these instruments. Second, factors associated with appropriate instrument selection are presented, including consideration of psychometric properties, norms, and limitations. The chapter also provides a framework for conducting and interpreting an assessment using these measures, including a practical checklist for clinicians. Finally, the chapter provides specific descriptions of standardized cognitive measures that are commonly used with infants, toddlers, and young children.

Keywords Early childhood cognitive assessment · Early childhood cognitive development · Bayley Scales of Infant and Toddler Development · DIAL-4 · Battelle Developmental Inventory (BDI-2) · Differential Ability Scales (DAS-II) · WPPSI-IV · Kaufman Assessment Battery for Children (KABC-II)

Introduction

A standardized measure of cognitive functioning is often a primary component of a comprehensive early childhood psychological evaluation. A child's performance during the cognitive assessment can have significant immediate and long-term

K.M. Ellingsen (✉)
University of South Florida, Tampa, USA
e-mail: drkirstenellingsen@gmail.com

A. Garro (ed.), *Early Childhood Assessment in School and Clinical Child Psychology*, DOI 10.1007/978-1-4939-6349-2_2

implications. Access to resources or treatment decisions may be predicated on the information gained from a particular measure. Scores may be used to make a diagnosis, determine eligibility for support services, or to examine the effectiveness of an intervention, plan treatment, or document consequences of disease, trauma, chronic conditions, and medical procedures. Therefore, it is essential to obtain reliable, valid, and meaningful assessment results and to interpret performance appropriately within the context of all other evaluation information.

For psychologists, successfully obtaining reliable, valid, and meaningful cognitive assessment results requires effective planning, appropriate training, and sufficient background knowledge about child development. It begins with carefully selecting an instrument based on the technical qualities and intended use and limitations of potential measures given the assessment purpose, referral concerns, and characteristics of a child. Practiced administration and accurate scoring are necessary, but not sufficient. Maximizing performance on standardized measures and accurate interpretation of results also depends upon comprehensive knowledge of child development and recognizing typical behavior for infants, toddlers, and preschoolers. Further, instrument selection, administration, and results interpretation should be grounded in developmental theory and informed by a general understanding about potential functional consequences of impairment, disability, and health conditions.

Assessment of early childhood cognitive ability is challenging for many reasons. Developmental change occurs most rapidly during the first five years of life and may be uneven across domains. Behavior can be variable and more susceptible to environmental and situational factors. The manifestation of different skills and abilities varies in infancy and "the rapid and often sporadic development of the CNS associated with the first year of life frequently results in dramatic changes in cognitive ability over a very short time" (Smith, Pretzel, & Landry, 2001, p. 188). Accurate assessment of concerns about cognitive functioning is also difficult due to the limited number of instruments available for young children and variability of child behavior within different contexts.

Literature Review

A cognitive assessment is conducted for different purposes including diagnostic, screening, research, program evaluation, and intervention planning. The method used to assess children's development and functioning reflects a particular theoretical framework. Identifying the theory and underlying assumptions behind evaluation decisions promotes responsible, informed, and purposeful practice. Before determining the method or instrument that will be used for the cognitive component of an early childhood assessment, it is also important to examine recommended practice in early childhood assessment and understand strengths and limitations of an identified approach.

Theories of Cognitive Development and Intelligence

It is important to recognize how theories are specifically related to understanding of cognitive development and definitions of intelligence. Discussions about cognitive ability constructs that are measured by intelligence measures for young children often take a developmental perspective, with research in this area using Piagetian and Information Processing Theories (Tusing & Ford, 2004). While an extensive discussion of theory is outside the scope of this chapter, select major contributions to present day understanding of cognitive assessment are presented below to ground discussion of evaluation methods and instrumentation.

Intelligence is a construct that is presumed to represent aptitude or effective application of cognitive functioning. Definitions of intelligence often refer to capacity or ability to learn, reason or understand, and apply knowledge (e.g., www.merriam-webster.com/dictionary, www.dictionary.com). During the past century, psychologists have proposed various theories about how mental processes can be observed, measured, and documented to be combined in a quantifiable way to reflect a general cognitive ability or "intelligence". While there is still no agreement about the definition of intelligence, particularly in early childhood, several commonly used tests of intelligence generally reflect the theory that there is a general factor (*g* factor) and specific factors or dimensions of intelligence building on the work of Alfred Binet, William Stern, Charles Spearman, and Louis Thurston. Intelligence is often assumed to be measurable and quantifiable in a total sum score and identified by different primary mental abilities (e.g., reasoning, verbal comprehension). However, the applicability of this construct to infants and toddler is questionable given their rapidly evolving cognitive development. The Cattell-Horn-Carroll (CHC) theory (Carroll, 1993; Horn & Cattell, 1966) has been applied in the design and interpretation of preschool measures of cognitive functioning. CHC posits that the general construct of *g* is comprised of a combination of *crystallized intelligence* (*Gc*), which represents knowledge acquired from experience and *fluid intelligence* (*Gf*), which is considered to be independent of acquired knowledge and reflect the ability to think logically, problem solve, and reason in novel situations.

Information Processing Theory has offered a leading strategy to study cognitive development (Meece, 2002). This approach focuses on developmental change of abilities in attention, memory, and problem solving. It represents cognitive development as a continuous process using a computer analogy with inputs, throughputs, and outputs to represent how the mind operates during memory, attention, and problem-solving activities (Puckett & Black, 2005). Changes in cognitive functioning for young children are considered to be the result of increases in memory, association, and use of cognitive strategies to process information (e.g., attending, rehearsing, coding information, forming mental images, or representational images). Cognitive development during early childhood includes a focus on salient features of objects in attention processes; memory scripts, which are mental representations of frequently repeated daily events; and simple strategies to remember experiences, with an ability to only keep a small amount of information in working memory at one

time (Meece, 2002). Early misperceptions and incomplete concepts are thought to be the result of a short attention span, unsystematic attending abilities, and limited memory (Puckett & Black, 2005). Many available intelligence tests and comprehensive evaluations of cognitive functioning emphasize these skills.

Vygotsky's theory has been helpful to frame learning and educational recommendations from cognitive assessments. One component of this theory that is reflected in recommended early childhood practice and assessment is emphasis on children's culture and history of experiences for understanding cognitive development (Meece, 2002). Ongoing criticism of intelligence testing with minority populations can be traced back to Vygotsky's work and the idea that different intellectual skills are developed according to what is needed in different societies. In addition, his ideas about the zone of proximal development, scaffolding, and the social context of learning can be used to address problems associated with use of standardized scores to determine "school readiness" for young children.

Piaget's theory of cognitive development emphasizes that children progress through a predictable pattern of four qualitatively different developmental stages (i.e., Sensorimotor, Pre-Operational, Concrete Operational, and Formal Operational) that reflect increasingly sophisticated and purposeful cognitive processes during infancy through adulthood. Changes in cognitive processes occur within an expected sequence of demonstrated skills and are reflected in different reasoning processes. Piaget viewed children as active in their own development, having their own logic and different methods of reasoning that they use to make sense of the world (Meece, 2002). His research and observation of childhood behavior are reflected in the design and content of popular standardized measures of cognitive development for infants and young children (e.g., the Bayley Scales of Infant and Toddler Development) as well as expectations of skills in qualitative assessments and developmental checklists.

Piaget proposed that "mental development during the first eight months of life is particularly important, for it is during this time that the child constructs all the cognitive substructures that will serve as a point of departure for his later perceptive and intellectual development" (Piaget & Inhelder, 1969, p. 1). It is toward the end of an infant's first year that children were expected to engage in "acts of practical intelligence" or show behavior that represents seeking different ways to reach an end goal or understanding relationships between actions (Piaget & Inhelder, 1969). Piaget focused on individual and general stage differences in quality of thinking and reasoning processes of children in assessment rather than an emphasis on a score or quantifiable score to represent an internal trait (e.g., intelligence).

Measuring Cognitive Development and Intelligence in Early Childhood

Direct individual assessment using a standardized measure is one method of collecting information for diagnostic purposes and intervention planning and monitoring. When measures demonstrate discriminant validity, the results "may be used for the early

identification of children at risk for difficulties, to describe a child's skill acquisition compared to peers, to determine eligibility for services, and to assess a child's specific needs for intervention" (Mazer, Majnemer, Dahan-Oliel, & Sebestyen, 2012, p. 249). However, "the younger the child, the more difficult it is to obtain reliable and valid assessment data" and it is particularly difficult to accurately assess children's cognitive abilities before six years of age (Shepard, Kagan, & Wurtz, 1998, p. 5).

Standardized Measures of Cognitive Assessment

Standard assessment of any construct begins with a definition and identification of observable or measureable indicators. Advantages of standardized measures include well-written and tested items, established standard conditions of administration and scoring, and inclusion of norm tables (Black & Powell, 2004). Standardized instruments of cognitive functioning should be based on theory and have empirical support that their tasks and activities are relevant, adequately representative, and discriminating. Standardized, norm-referenced measures require an examiner to follow specific administration and scoring rules in an attempt to create a similar testing experience for all children. "Standardized developmental tests are often viewed as the gold standard for outcome assessment, providing an objective, valid and reliable evaluation of a child's development in comparison to the norm, and typically provides standardized scores that can be used to classify developmental level. Through the use of standardized administration and scoring criteria, measurement error is reduced, providing an objective, accurate evaluation of a child's abilities in various developmental domains (Mazer et al., 2012, p. 249)." While standardized measures of cognitive functioning for children younger than three years of age are not considered to yield a score of intelligence, preschool measures often refer to an intelligence quotient (IQ) or overall composite score that is used to represent general cognitive ability.

Issues, application, and limitations of standardized tests. Identified concerns about using standardized cognitive assessment and intelligence tests in early childhood include the low reliability and predictability of measures, challenges with trying to use standardized administration in young children, and poor utility of assessment data for intervention planning. These concerns have been particularly true for children with developmental delays or disabilities (Bagnato, 1992). However, these are precisely the children who are referred for diagnostic evaluations and developmental assessments to determine eligibility for early intervention. Bagnato and Neisworth (1994) have more specifically discussed the poor treatment and social validity of these measures. These forms of validity extend beyond psychometric properties of an instrument and focus on issues such as suitability/fit for children with functional limitations, practical use of measures to plan interventions that are applicable to children's everyday environments (i.e., ecological validity), and the feasibility of incorporating collaborative teamwork as part of assessment. Further, according to Smith et al. (2001) a limitation of standardized

individual assessment for infants is that the data represents "only a small sample of an infant's developmental repertoire" (p. 188) with these data "influenced greatly by current issues regarding the infant's motivation, mood, comfort, and responsiveness to the examiner and the evaluative process" (p. 188). Other concerns about the use of such measures with infants and toddlers include: (a) the inability of many measures to capture cognitive growth over brief intervals (McDermott et al., 2009); (b) variability in the way scores are obtained; (c) the absence of children with disabilities from norm groups; and (d) lack of an established unified definition of intelligence (Cornish, Sornberger, Dupasquier, & Wilding, 2012).

The aforementioned concerns have led to an ongoing debate about whether standardized, norm-referenced measures should ever be used with young children for diagnostic or eligibility decisions or if they should be used as part of a more comprehensive evaluation where the limitations of the measure are considered during the interpretation of results (Bradley-Johnson, 2001). Nevertheless, many psychologists report frequently using standardized tests of cognitive functioning and intelligence for early childhood evaluations; because of this, there is a need for training and awareness about the strengths and limitations of available measures for use with the early childhood population (Bagnato & Neisworth, 1994; Bradley-Johnson, 2001). The next section outlines important considerations for evaluating existing standardized measures and selecting an instrument.

Selecting and Evaluating Standardized Measures

Multiple issues are important for selecting a measure in early childhood assessments including purpose, conditions of testing, examiner's expertise, cost and the availability of materials (Grigorenko & Sternberg, 1999). Flanagan and Alfonso (1995) proposed the following criteria to consider for the selection of intelligence tests for preschoolers: size of normative sample; recency of normative data; match of demographic characteristics of the normative group to the US population; test internal reliability; test-retest reliability; subtest floors; validity evidence; and subtest item gradient violations. Subtest item gradient violations involve situations where a child's performance on a single or small number of items causes small changes in raw score points, which, in turn, generate disproportional effects on the child's standard scores and lead to reduced sensitivity of a test or part of a test (Bracken, Keith, & Walker, 1994; Campbell, 2005). Specific psychometric properties related to validity and reliability should be carefully evaluated as well as potential issues related to bias or limitations with specific populations. Qualitative characteristics should also play a role in selection of measures. These include cost of materials; time to administer, score, and interpret a measure; organization and ease of administration and scoring; attractiveness of materials for young children; and amount of training required to reliably administer a test/measure. The following sections will address factors related to assessment planning, including selection and evaluation of existing instruments.

Evaluating the Qualities of Early Childhood Cognitive Assessment Instruments

An examiner must understand how to select and evaluate instruments appropriate for an individual child based on assessment purpose, psychometric properties, and other factors. Although it is important to understand individual factors/circumstances surrounding an evaluation (e.g., reason for referral), it is equally important to understand how to generally evaluate a measure on psychometric properties and identified strengths and weaknesses. Such evaluation will ensure appropriate selection of a standardized measure for an individual child.

Purpose of the assessment. Selection of an appropriate instrument should initially be guided by identifying the purpose of the assessment. Purpose directs what should be measured and how it should be measured. It "determines the content of the assessment; methods of data collection; technical requirements of the assessment and, finally, the stakes or consequences of the assessment, which, in turn, determine the kinds of safeguards necessary to protect against potential harm from fallible assessment-based decisions" (Shepard et al., 1998, p. 6). This process begins with acquiring a basic understanding about a child's background, referral concerns, and intended use of scores. For example, determining eligibility for special education services, documenting level of specific cognitive functions for neurologically involved health conditions or determining effects of different medical conditions or procedures presents different data needs than program evaluations and documenting child outcomes for state IDEA accountability systems. Clinicians should examine the technical manual of an instrument to determine its intended purposes and look for evidence of validity connected to these purposes. If a test/measure is used outside its usual parameters, this factor needs to be considered when interpreting results.

Psychometric properties of the instruments. An examiner should have sufficient knowledge about technical adequacy and psychometric properties when selecting a standardized measure of cognitive functioning. Technical adequacy/qualities are considered strong when internal consistency and stability are at or above 0.90 for total test scores and 0.80 for subtest scores and composites (Lichtenberger, 2005). Instruments should be responsive to subtle and rapid changes in demonstrated ability; this includes investigating the sensitivity (e.g., adequate item gradients, floors) and stability of a measure.

Reliability and validity are generally lower in tests for infants than for tests designed for older children. This is due, in part, to the fast changes in cognitive and other developmental domains. Low test-retest reliability of infant and preschool measures has been identified as a major concern (Bradley-Johnson, 2001). In general, the younger the child, the less stable the scores in relation to later IQ or academic functioning. Regarding item gradients, criteria established by Bracken (1987) suggest that a one-unit increase in raw score points should generate a standard score change that is no greater than one third of a standard deviation (SD). There are different types of validity that can be examined in relation to early childhood cognitive measures. Predictive validity for such measures is often low

due to rapid and/or uneven development during this age range. Establishment of construct validity can also be challenging since individual test items might measure different constructs at different ages. With respect to construct validity, it is also important to keep in mind that infant measures of cognitive development tend to rely more on children's sensorimotor functioning, whereas preschool measures include many activities that require verbal responses and often reflect expressive and receptive language skills (Mazer et al., 2012). Inadequate numbers of ceiling or floor items can have significant negative effects on validity; the latter especially has implications for assessing children with developmental delays and/or disabilities.

Norming procedures. Standardized measures provide raw scores on scales that are compared to same age peers for norm-referenced interpretation. For most measures, the normal distribution has a mean of 100 and SD of 15. For many total, composite, and subtest scores of a given measure, percentiles are provided to explain a child's performance/score relative to the norm group of same age peers. Standardized measures should be reasonably representative of the general population and normed using data that is no more than a decade old. The comparison of scores to what falls in the norm table indicates if a child is functioning at the expected level for his or her age or if the child demonstrates a significant difference. Therefore, the sample that is used for the norm group should match the children who are given the test. As such, "it is always critical to examine the makeup of the normative sample in order to determine whether the norms are applicable to the population of interest" (Mazer et al., 2012, p. 250). Outdated norms might result in inflated standard scores and, consequently, prevent a child from meeting eligibility requirements to access early intervention services.

Bias. Diverse populations may perform differently on measures of cognitive/intellectual functioning. Selecting an appropriate measure includes examining how the instrument was normed and if the children used to obtain norms match the background of the child. As such, "when evaluating ethnically diverse preschool children, clinicians must be aware of the standardization procedures for the measure they decide to use, as well as its interpretive quality" (Dale, McIntosh, Rothlisberg, Ward, & Bradley, 2011, p. 485). Recognition of potential bias is necessary because test scores are often used as criteria to access resources. It is important to examine the appropriateness of instruments for children from different cultural and language backgrounds. If the scores do not represent the ability or current functioning of a child because of significant cultural, linguistic, or disability factors, validity can be significantly compromised. APA (2012) recommends that a psychologist read the test manual or contact the test's publisher for additional information to determine if a measure is appropriate for children with specific disabilities, including review of validation studies. APA's most recent Guidelines for Assessment of and Intervention with Persons with Disabilities includes the specific principle to "strive to apply the assessment approach that is most psychometrically sound, fair, comprehensive, and appropriate for clients with disabilities." When a measure does not have relevant disabilities in norm groups, the psychologist is advised to find instruments that "maximize the collection of valid information" (APA, 2012).

Implications for Practice

Conducting and Interpreting a Standardized Cognitive Assessment

Before administering a standardized cognitive test it is important to have background information about the child and presenting problems. Selecting an appropriate measure based on the assessment purpose, child background, and technical qualities is important. Several administrative factors can influence the validity of results. For example, knowledge about developmental milestones and the sequence of skill acquisition is necessary to engage young children, plan an appropriate evaluation strategy, schedule and arrange the environment to optimize child performance, and interpret individual performance. Understanding sequences of development and growth in the cognitive domain provides valuable information about test construction, the intended purpose of different types of assessment tasks and activities, and underlying skills or abilities that are being assessed with particular items. Early cognitive abilities include alertness, visual scanning, problem solving, reasoning, comprehension of directions, concept formation, object permanence, understanding of cause and effect and size and spatial relationships. "It is important to note that the various facets of development are often interdependent and therefore, to successfully complete items in a specific developmental domain may require additional skills in other domains" (Mazer et al., 2012, p. 250). Clinical judgment and interpretive skill is enhanced when an examiner knows generally what to expect and the order in which skills emerge.

Performance on standardized measures may be affected by several factors other than a child's ability, including an examiner's skill in engaging a child in different activities and maintaining proficiency in administering items with different manipulatives. Knowing the standardized administration and scoring procedures is essential, particularly during early childhood assessments when each item may have different materials, administrative procedures, and scoring criteria. The objectivity and validity of the results depend on strict adherence to the standardized procedures (Mazer et al., 2012). Standard administration procedures need to be rehearsed adequately for verbal instructions and the correct presentation of manipulatives to facilitate smooth transitions between objects or types of tasks, while observing behavior for accurate scoring decisions and maintaining child engagement. This often involves learning verbal instructions verbatim and adequately rehearsing the specific presentation and placement of a variety of test materials. Many standardized instruments include a variety of manipulatives and tasks to facilitate engagement in the testing process. Transitioning between these items or different tasks must occur in a manner to keep the child engaged while making accurate observations of the child's behavior and functioning to better interpret results from the assessment. While adhering to standardized procedures is important, examiners must also demonstrate flexibility due to the variable nature of young children's behavior. Such flexibility might require additional breaks from testing, neutral

praise, use of tangible reinforcers, and administering items on a caregiver's lap for infants and toddlers or on the floor or a child-sized table for preschoolers.

Performance on standardized cognitive measures may be significantly influenced by various child characteristics and conditions. Specific characteristics that can impact young children's performance on cognitive measures include high levels of activity and distractibility, low attention span, stranger anxiety, or temperament anxiety. Language differences must also be taken into account since assessment results, particularly in verbal areas, might be confounded for children who come from backgrounds with limited exposure to English. As with other evaluation components, matching the temperament of a child and facilitating a positive and comfortable environment will optimize performance and engage a child. A shy infant may need more time getting comfortable playing with toys while the examiner talks with a parent, while an active 24 month old might need to be administered items quickly and with enthusiasm.

When it comes to assessing young children with delays, health conditions, or disabilities, it is also crucial to understand how these factors might affect performance in one or more domains of a standardized measure. According to Simeonsson and Rosenthal (2001), comprehensive and accurate assessment of children with disabilities or chronic health conditions is important in order to: (a) facilitate diagnostic efforts; (b) ensure that a match is made between the needs of the child and appropriate intervention and (c) evaluate the impact of individualized treatments or interventions. While assessment of young children with already identified or suspected disabilities or health conditions can be challenging, there are some general guiding principles. For example, Hodapp (1998) notes that "on the basis of data from research on children with disabilities as well as cross-cultural research, it has been concluded that all children develop early cognitive or language skills in the same sequences" (p. 174). From a practical perspective, this means that young children with significant delays or disabilities might not be able to complete item sets at their chronological level, but can do better with item sequences at a lower level. This issue, in turn, speaks to the need to have adequate floor items for young children with delays or disabilities. According to Shultz and Chase–Carmichael (2001), chronic health conditions, in general, do not call for deviation from standardized procedures unless the child has additional impairments (i.e., sensory, motor, orthopedic). If a physical disability or delay is present, degree of severity is important to take into account; overall, a higher level of severity calls for more significant adaptations to assessment and greater caution in interpretation of results (Shultz & Chase–Carmichael, 2001). As a guideline, Schultz and Chase–Carmichael recommend that examiners note the effects of chronic health conditions and any concomitant impairment on the child's performance. Alternately, examiners can adapt assessment procedures to reduce the impact of the condition or impairment on the child's performance. When young children with disabilities do require adaptation for standardized procedures, examiners might provide accommodations or modifications. Modifications involve changes to a test's format or content that wind up altering the underlying construct(s) that are being measured. Thus, modifications affect the validity of an assessment. Examples might include

removing time requirements from subtests that are supposed to be timed, allowing examinees to point instead of using language on verbally-based tasks, or using multiple choice formats instead of having examinees describe/explain responses. Accommodations in assessment, on the other hand, involve changes in test format, content, or administration, which makes it more accessible to individuals who would otherwise not be able to complete the measure, including individuals with disabilities (APA, 2012). Accommodations are not expected to change the construct being measured. Accommodations might include moving test materials closer for a child with visual impairment, other physical rearrangements of the testing environment, allowing extra time for responses on non-speed-related tasks, and providing breaks between testing tasks/activities.

As is the case with administration of cognitive measures or intelligence tests, interpretation and reporting of results should be carried out carefully. Scores/results should be interpreted in light of the child's background, including family context/history, social history, and developmental/medical history, as well as other assessment information (e.g., other test results, observations, interviews/reports from parents, teachers, etc.). When interpreting scores, the child's behavior, including effort; mood; compliance; and levels of interest, attention, persistence, engagement, and motivation should be evaluated across different tasks, subtests, etc. In addition, Lichtenberger (2005) notes that a number of variables can contribute to performance differences across test domains. These variables include neuromuscular problems; language learning; or visual-motor deficits; and/or internal factors such as fatigue and inattention. When discussing results of cognitive assessment with parents/families or teachers, the aforementioned factors should be included as part of the discussion. In addition, it is vital for clinicians to seek input from parents/caregivers and teachers about whether the child's performance during the assessment and his/her test results are representative of what is seen in real-life situations. Finally, standardized assessment results should be interpreted in terms of inter-individual differences, meaning performance relative to the norm group of the instrument, and for intra-individual differences. The latter considers the range of skills and characteristics unique to a particular child, including strengths and weaknesses. This is especially important for younger children in light of the rapid changes in their cognitive and other domains of development.

The following table provides a checklist and guidelines for practitioners in selecting, administering, and interpreting early childhood cognitive assessments.

Checklist to Evaluate Early Childhood Cognitive Assessment

	Yes	No	Comments
Instrument selection			
Was the measure appropriate?			
Used for purpose intended			

(continued)

(continued)

	Yes	No	Comments
Documented use with children with particular characteristics (e.g., disabilities, different racial or ethnic groups, etc.)			
How comprehensive is measure?			
What are the psychometric properties?			
Adequate reliability			
Adequate content and construct validity			
Recent norms			
Normed on population with similar child characteristics (e.g., racial/ethnic/cultural/language background or disability status)			
Effective in discriminating among children with different diagnoses, characteristics, etc.			
Adequate floor and ceiling			
Test administration and scoring			
Was the measure administered using standard protocol?			
Were materials organized?			
Instructions read verbatim			
Prompts used correctly			
Accurate presentation of materials. Presentation of trials as required			
Were modifications or accommodations necessary? If so, were they documented?			
Did the examiner establish rapport and obtain best performance from child?			
Successfully keep child engaged in tasks (consider activity level, affect, eye contact, communication exchanges, responsivity to people and assessment environment, etc.)			
Make the assessment as enjoyable as possible			
Provide necessary breaks			
Minimize distractions			
Provide smooth transitions between materials and tasks			
Effectively and neutrally use praise and reinforcement			
Were parents/caregivers appropriately engaged in assessment? Did parents/caregivers interfere with assessment?			
Were items accurately scored?			
Recorded correctly			
Added correctly and checked for accuracy			
Correct scoring criteria applied			
Does the interpretation of results appropriately synthesize all data?			
Were scores examined across different scales and subtests?			
Do the patterns of scores appear to reflect child's functioning?			
Were there marked or unusual inconsistencies across different domains, scales, etc.?			

(continued)

(continued)

	Yes	No	Comments
Was assessment performance influenced by factors such as shyness, anxiety, fatigue, etc.? Were these factors taken into account when interpreting results?			
Did child's health status, disability(ies), or impairment(s) limit the ability to respond to demands of the assessment and impact results?			
Were various aspects of child's behavior (e.g., response to examiner) recorded/noted and taken into account when examining and interpreting results?			
Were the results interpreted considering functioning and performance on other measures and in other settings? Using multiple measures and methods?			

Implications for Practice: Select Measures of Cognitive Functioning

Although there are limitations and challenges involved with the use of standardized norm-referenced instruments of cognitive functioning in young children, these measures continue to provide key data for a variety of purposes. This section summarizes the purpose, content, application, and critique of select early childhood cognitive assessments.

The Bayley Scales of Infant and Toddler Development™—Third Edition (Bayley-III; Bayley, 2006)

Description/Background. The Bayley-III provides a standardized assessment of the current developmental functioning of infants and young children from birth to 42 months. The first version of the Bayley Scales of Infant Development (BSID) was published in 1969 and revised in 1993 (BSID-II) (Black & Matula, 1999). The Bayley-III provides norm-referenced scores (i.e., scaled scores, composite scores, percentile ranks). It was designed to be consistent with current scholarship on child development and federal and professional standards and to be used to identify suspected developmental delay and plan treatment and intervention services (Weiss, Oakland, & Aylward, 2010). The Bayley-III contains three main scales: Cognitive, Language, and Motor. These take 30–90 min to administer, depending upon age of the child. Examiners can also administer the supplemental Social-Emotional and Adaptive Behavior Scales to caregivers. Tasks on the Cognitive, Language and Motor scales of the Bayley-III are completed in a standardized manner, using manufacturer-supplied testing toys, verbatim task instructions, and very specific

scoring criteria. The Cognitive scale includes items that assess sensorimotor development, exploration and manipulations, object relatedness, concept formation, memory, habituation, visual acuity, visual preference, cause and effect, problem solving, representational and pretend play, and object permanence. Early learning is also assessed such as early numeracy skills (e.g., one to one correspondence), matching colors, and discriminating patterns.

Raw scores are converted to scaled scores that range from 1 to 19 with a mean of 10 and SD of 3. Within each domain, (Cognitive, Language, and Motor), composite scores are also calculated. Composite scores have a mean of 100 and SD of 15 with a range from 40 to 160. Percentile ranks range from less than the first percentile to greater than the 99th percentile. Scores that fall two SDs or more below the mean are considered to reflect developmental delay. Composite scores ranging from 90 to 109 are considered average and within normal limits of functioning. A score at or below 70 is considered to represent a significant delay. Children do not typically qualify for early intervention services unless their scores are two SD below the mean. Descriptive classifications are also sometimes used to describe Composite scores from the Bayley-III using 10 point increments around the mean. These are as follows: 130 and up- "Very Superior"; 120–130- "Superior"; 110–120- "High Average"; 90–110- "Average"; 80–90- "Low Average"; 70–80- "Borderline/Low"; and Below 70- "Extremely Low." The Cognitive scale yields a composite score which allows for the comparison of a child's functioning to same age peers. Raw scores from children born full-term (i.e., 37 weeks gestation and above) are compared to chronological age peers. For children born prematurely (i.e., less than 37 weeks gestation), examiners first adjust for the child's weeks of prematurity and use this adjusted age to calculate standardized scores. Children who were born prematurely are compared to adjusted age peers until they are 24 months adjusted age and then compared to chronological age peers.

There are several practical considerations for clinicians who use the Bayley-III. During item administration, an infant may sit on a caregiver's lap, while toddlers and preschoolers may sit in a supported child-size chair. It is important to politely and respectfully discuss caregiver expectations and involvement before the administration of any items to prevent spoiling items and disruption of standardized procedures. Ask a parent or caregiver who is present (or holding a child) not to prompt, encourage, guide, or assist with any items. It is also imperative to practice the standardized administration for each item several times before attempting to use with a child. This includes rehearsing the exact language for instructions, proper placement of materials (e.g., handing the block to a child vs. putting it on the table in front of the child), and steps to present materials. It is important to organize the materials so that those that will be needed first are convenient and close, but out of view of the child to facilitate smooth transitions between items. Using a transition object may also be helpful to keep a child engaged between items if your pace is not yet quick, or if a child does not want to release a preferred object to avoid a child becoming upset. Knowing what materials are needed and having quick access to them will help with maintaining engagement and allow an examiner to watch subtle behaviors for accurate scoring. Keeping notes on the scoring protocol is also helpful.

Applications, Strengths and Limitations. The Bayley-III is generally considered the gold standard measure for infant and toddler development and is used in early childhood intervention evaluations, hospital developmental follow-up programs, community clinics, and as part of developmental outcome research protocols. It is often used to determine eligibility for early intervention services for children (birth to three) and preschool special education services within school districts. The Bayley-III has been described as an internationally recognized tool that is comprehensive in nature and is well suited for assessing the development of young children (Macow, 2008; Pinon, 2010). Because the measure includes three directly administered scales (i.e., cognitive, motor, and language), it is useful for multi/interdisciplinary early intervention assessments where a team of professionals administer the different components. For example, the Bayley-III might be used by a speech and language pathologist who administers the language scale, an occupational or physical therapist who administers the motor scale, and a psychologist who administers the cognitive scale. The Bayley-III has also been used clinically at hospitals. For example, multidisciplinary NICU clinics often conduct follow-up visits using the Bayley with children born prematurely and/or who have health conditions. In such clinics, a psychologist often administers all three components of the Bayley at set intervals and then refers to a physician, physical therapist, or speech/language pathologist for additional assessment or consultation for specific developmental concerns or for medically complicated children and atypical performance profiles. The Bayley has been used extensively in multiple educational, early intervention, and medical studies with a range of clinical populations, including premature infants and children with Down syndrome, cerebral palsy, language impairment, and/or suspected autism spectrum disorder. The Bayley-III has strong reliability and validity with extensive studies on validity with previous versions and standardized measures of language, adaptive behavior, and motor skills. The items are based on theory. The format is useful for multidisciplinary teams. Hand scoring is relatively quick and straightforward, as is determining ceiling or stopping points. Tasks are interactive and include a variety of materials to engage young children. The items are intended to be administered as fun and playful activities. Unlike the developmental screening and surveillance tools typically used in pediatric practices, which are not diagnostic and identify children only as "at risk," the Bayley-III can be used as part of a diagnostic assessment for developmental delay. The main limitation of the Bayley-III is that it is time intensive to learn. It requires both standard verbal instructions and specific steps to administer items; some items require timing or depend on performance on earlier items. Caregivers or parents may interfere and unintentionally spoil an item by prompting or guiding a child. Consistent with other measures, scores in infancy are not good predictors of later cognitive functioning unless very delayed. Scoring criteria for the Bayley sometimes make it difficult to determine whether skills are still emerging or have been mastered. Scoring might also be complicated when a child refuses an item or is not motivated to perform an activity. In these situations, the examiner might need to drop back to administer several additional items extending the time to complete the test.

The Developmental Indicators for the Assessment of Learning™, Fourth Edition (DIAL™-4 Mardell & Goldenberg, 2011)

Description/Background. The DIAL-4 is an individually administered screening test to identify children in need of interventions or further diagnostic assessment. The targeted age range of examinees for the DIAL-4 is 2–6 years through 5–11 years. Total administration time for the DIAL-4 is between 30 and 45 min. The DIAL-4 directly measures motor, conceptual, and language skill areas. Specifically, these include: (1) Gross Motor (e.g., catching, skipping) and Fine Motor (e.g., building with blocks, cutting, copying shapes) items, (2) Expressive (e.g., naming) and Receptive items (identifying objects) and (3) Concepts (e.g., naming or identifying colors, rote counting, sorting shapes). There are 21 subtests that comprise the three scales. There are two age ranges for items (i.e., 2 years 6 months–3 years 11 months and 4 years–5 years, 11 months). The DIAL-4 includes a variety of age-appropriate manipulatives and tasks for young children. It includes a total standard score with an accompanying percentile rank; standard scores and percentile ranks are also provided for the Motor, Concepts, and Language domains. Scores falling at or below the 16th percentile are considered to be indicative of a potential delay. In terms of procedures, the DIAL-4 involves direct assessment with an operator's handbook and bag of test materials for each domain/area. The handbooks include exact wording that an administrator should use in bold red type; instructions for when additional prompting can be used are also specified. An administrator records scores on record forms and takes a child's final response if more than one response is given for an item. Self-help Development, including personal care skills, and Social Development, (e.g., rule compliance, self-control, and empathy) can be assessed through caregiver report.

Applications, Strengths, and Limitations. The DIAL-4 is designed for use in preschools and kindergartens as well as early childhood education programs such as Head Start. According to the publishers, the DIAL-4 is an individually administered global screener for assessing large groups of children quickly and efficiently. Training for the DIAL-4 recommends setting up a room with a registration area, play table, and three separate tables for each of the direct assessment areas. There is a brief version of the DIAL (i.e., the Speed DIAL) that is advertised as appropriate for "quick screening in smaller settings such as departments of public health, pediatric offices, health fairs, homes, and classrooms." Strengths of the most updated version of the DIAL (DIAL-4) include new norms, a lowered floor of items (from 3 years to 2 years 6 months), additional items that are related to academic success, and improvements in the handbooks and record form for easier administration and scoring, including reformatting and simplified instructions and scoring rules. There is also a corresponding teacher questionnaire to obtain additional information about the child's functioning. The main limitation of the DIAL-4 is that it is not a diagnostic test or intelligence test. It should be applied as a screening instrument to identify if a child has a potential developmental delay.

Battelle Developmental Inventory, Second Edition (BDI-2; Newborg, 2005)

Description/Background. The BDI-2 is a measure used to screen and evaluate early childhood developmental milestones. It is a standardized assessment of a child's development that can be used from birth through age 7 years 11 months. According to the developer's manual, the four primary purposes of the instrument include: (a) assessing typically developing young children, including screening for school readiness; (b) assessing or identifying developmental delay or disability; (c) planning instruction and intervention; and (d) evaluating early childhood programs. In addition, the BDI was designed to align with Head Start and OSEP early childhood outcomes and many preschool curriculums. The BDI was designed to be a comprehensive test of development across five domains- Motor, Adaptive, Cognitive, Personal/Social, and Communication and is both norm-referenced and criterion-referenced. Domains can be administered separately. The BDI-2 (2005) is comprised of 450 items and involves multiple administration methods (e.g., structured play activities with scripted formats; observation of activities in natural environments; and interviews with parents/caregivers and teachers). Start points for items are determined by child age or estimated ability level. Data can be documented with a mobile data solution system. There is also a screening version, the BDI-2 ST that includes 100 total items, with 10 levels by age range. Scores for the BDI-2 include a total and domain scores; these can be provided as standard scores, percentile ranks, age equivalents, z scores, and T scores. The BDI-2 also offers change sensitive scores.

Applications, Strengths, and Limitations. The BDI-2 has been used as the uniform tool for collecting child outcome indicators across several states including Florida, Mississippi, New Jersey, and South Dakota (Elbaum, Gattamorta, & Penfield, 2010). While the BDI-2 may be used to describe developmental delay or typical development, it was not intended as an instrument to diagnose specific disabilities. The BDI-2 ST has also been found to have acceptable sensitivity, specificity, and accuracy in classifying developmental delay when compared to the complete BDI-2 (Elbaum et al., 2010). One of the strengths of the BDI-2 is its strong psychometric characteristics. It uses norms established by a standardization sample that had the same distribution of ethnicity reported in the US Census and was developed to be culturally sensitive and include accommodations for children with diverse disabilities (Elbaum et al., 2010). In fact, the BDI-2 meets or exceeds traditional standards for reliability at the domain and full test composite levels. The developers provide data on the sensitivity, specificity, and classification accuracy for different clinical samples for the BDI-2 and BDI-2 ST (Newborg, 2005). Administration of the BDI-2 does not require extensive training, although an examiner should have familiarity with child development, thorough understanding of the BDI-2, and adequate experience working with young children (Mazer et al., 2012). One limitation of the BDI-2 is that it was not normed for non-native English speakers and still needs to be validated with a larger, more diverse population of

children. It does have a Spanish language version. As of the date of this writing, norms for the BDI-2 are over ten years old and, thus, require updating. The BDI had problematic item gradients for children birth through 23 months, and the sample of items in the original test was too limited to inform intervention planning for the youngest children (Bradley-Johnson, 2001). The multiple administrative formats of the BDI can be problematic from a standardization perspective, but are regarded as a strength with respect to flexibility (Berls & McEwen, 1999).

The Differential Ability Scales: Second Edition (DAS-II, Elliott, 2007)

Description/Background. The DAS-II is a widely used measure used to evaluate cognitive development in preschool-age children. It was "developed to emphasize specific individual cognitive strengths and weaknesses, as well as general intelligence" (Keith, Low, Reynolds, Patel, & Ridley, 2010, p. 676). The DAS was originally designed from an eclectic theoretical orientation, although research has demonstrated that its factor structure is generally consistent with the CHC theory of intelligence (Keith et al., 2010; Lichtenberger, 2005). Practitioners are currently advised to interpret results for children aged 4 years and older applying CHC theory using the Verbal, Nonverbal, and Spatial clusters (Keith et al., 2010). The DAS can be used with children and youth ages 2 years 6 months through 17 years 11 months. Subtests are grouped into the Early Years and School-Age cognitive batteries with a few subtests that are common to both batteries. The Early Years core battery includes verbal, nonverbal, and spatial reasoning subtests appropriate for ages 2:6 through 6:11. This battery is further divided into two levels: one for children ages 2–6 through 3–5 and the other for children ages 3–6 through 6–11. Younger children are administered four core subtests to obtain a general ability composite score (GCA), while children ages 3–6 through 6–11 take six core subtests to acquire a GCA. The GCA is considered the general ability of a child to perform complex mental processing and is comprised of the following three domains: the Verbal Cluster, which measures acquired verbal concepts and knowledge, the Nonverbal Cluster, which represents complex nonverbal mental processing abilities, and the Spatial Cluster, which is a measure of complex visual-spatial processing. In addition to providing a GCA, the DAS-II yields standard scores for these clusters.

Applications, Strengths, and Limitations. The DAS–II is a useful tool for providing a profile analysis of children's cognitive strengths and weaknesses and yields reliable subtest and cluster scores (Reddy, Braunstein, & Dumont, 2008). One of the main advantages of the DAS-II is that it enables clinicians to conduct complete comparisons of test performance across time. Since both the Early Years and School Years batteries were standardized with children ages 5–0 through 8–11 and have overlapping norms for this age range, examiners can administer subtests that fit with child ability level. Scores for the DAS-II include age-based standard scores, percentile

ranks, age equivalents, and T scores for subtests. The DAS-II has a lower basal and higher ceiling range compared to the original DAS, which allows for standardized scores for children who may demonstrate very delayed development or advanced skills for their chronological age The DAS-II has supplementary diagnostic subtests that allow clinicians to obtain additional information about the child's skills in other areas (e.g., working memory and processing speed) (www.pearsonclinical.com). Another main strength of the DAS-II is that it includes clinical samples for a variety of disabilities and has been applied and studied with diverse groups of children, including those with learning disorders, language disorders, and ADHD. The psychometric properties of the instrument are considered sound, including good internal reliability and construct validity. The DAS-II has effectively differentiated children with learning disabilities and preschoolers at risk for LD from typically developing control groups (Reddy et al., 2008). Another advantage of the DAS-II is that it includes a Spanish translation and American Sign Language translation of the nonverbal subtest administration instructions. The main limitations of the DAS-II involve administration and scoring. Since the measure includes a range of manipulatives, complex organization of materials, various standardized directions, and required verbatim wording and queries, it is challenging to learn. In addition, the scoring and conversion system is more difficult and complex than that of other cognitive measures.

The Kaufman Assessment Battery for Children, Second Edition (KABC-II; Kaufman & Kaufman, 2004)

Description/Background. The KABC was designed as a measure of cognitive abilities and processing skills in children and adolescents. It was designed to be used with children and youth aged three through 18 years 11 months. The KABC-II is organized into three levels (i.e., for age 3, age 4–6 years, age 7–18 years). The basis of the KABC included two neuro-psychological theories characterized by a dual-processing approach: Sperry's 1968 cerebral specialization approach and the Luria-Das successive simultaneous processing dichotomy (Lichtenberger, 2005). The KABC-II test construction allows examiners to decide between two theoretical models: the CHC model and the Luria model. Before administering the KABC-II, the psychologist should decide which theoretical model to apply as "the theoretical model will influence the administration of subtests; different subtests are deemed "core" or "supplementary" depending on the model chosen, and the scoring of scales is also different between the CHC and Luria models." (Cornish et al., 2012, p. 50). For young children, the Age 3 battery yields one scale, which represents a global measure of ability using either five subtests (Mental Processing Index-MPI) or seven subtests (Fluid-Crystallized Index-FCI). The Age 4–6 battery organizes subtests into three scales (Luria model) or four scales (CHC model) (Lichtenberger, 2005). The four scales for the CHC model include Short Term Memory (Gsm), Visual Processing (Gv), Long-Term Storage

and Retrieval (Glr), and Crystallized Ability (Gc). According to Kaufman and Kaufman (2004), the CHC model is recommended for children with intellectual disabilities or disabilities in reading, written expression, or mathematics and for children with emotional, behavioral, or attentional problems. According to Cornish et al., the Luria model is recommended for children with autism and language disorders. The KABC-II generates a global score and subscale scores. Scores are provided as age-based standard scores, age equivalents, and percentile ranks.

Applications, Strengths, and Limitations. Test items of the KABC were designed to have little cultural content in order to provide a more fair assessment for children of diverse backgrounds. The measure has been translated and standardized in several different countries. According to several researchers (e.g., Dale et al., 2011; Kaufman & Lichtenberger, 2002; Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005), the KABC-II is more likely than other tests to show comparable performance between samples of African-American and White children, including preschoolers. Dale et al. (2011) praise the emphasis that KABC-II developers placed on diversity as part of the standardization process. Overall, the representative nature of the KABC-II's standardization sample and its foundation in CHC theory are considered to be strengths of this measure. The KABC-II also provides a Spanish language version. With respect to limitations, researchers (e.g., Dale et al.) have noted that the KABC-II is best interpreted at the composite level with an overall score that can provide a general view of performance. Analysis of strengths and weaknesses through subtest scores is not considered as valuable because these subtests were developed as a complementary aspect of the theoretical constructs represented at the composite level (Dale et al.). Finally, due to the age of the norms for the KABC-II (i.e., over ten years old), it requires updating in this area.

Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI-IV; Wechsler, 2012)

Description/Background. The WPPSI was first published in 1967, with one of its primary applications as a method to evaluate Head Start programs. Since its original version, the WPPSI has been revised three times- in 1989, 2002, and 2012. Initial versions of the test reflect Wechsler's view that intelligence "is a global entity that is multidimensional and multifaceted, with each ability being equally important" (Gyurke, Marmor, & Melrose, 2004, p. 57).

The WPPSI-IV is an individually administered measure designed to assess overall cognitive functioning in children aged 2 years 6 months through 7 years 7 months. According to Pearson, publisher of the WPPSI-IV, primary purposes of the test include: (a) identifying and qualifying children with cognitive disabilities, developmental delays or learning disabilities for special services; (b) identifying cognitive problems and recommending interventions; and (c) determining the impact of TBI on cognitive functioning. Tasks administered are determined by the child's

age. For the younger age subset (aged 2–6 through 3–11), the FSIQ is derived from the Verbal Comprehension, Visual-Spatial, and Working Memory Indices. For the older age subset (aged 4–0 through 7–7), the FSIQ is also comprised of these three Indices along with the Fluid Reasoning and Processing Speed Indices. The FSIQ is considered the most representative estimate of global intellectual functioning. Performance of cognitive functioning is compared to same age peers. Thus, an average FSIQ is 100, at the 50th percentile. Standard scores on composite measures are based on the mean of 100 and a SD of 15, with average scores ranging from 90 to 109. A score at or below 70 is considered to represent a significant delay.

Applications, Strengths, and Limitations. Overall, the WPPSI is widely used and regarded as a measure of young children's intelligence. It has been applied in both research and clinical practice, and is frequently employed as a means to qualify young children for early intervention or special education services. It has been used with children with diverse disability diagnoses. The WPPSI-IV includes several revisions from the WPSSI-III. According to Pearson, the WPPSI-IV psychometric properties are improved with increased accuracy of measurement for extremes of ability. The WPPSI-IV has better coverage of the construct of Working Memory as compared to its predecessor, and its Processing Speed subtests are revised so that they are less reliant on a young child's fine motor skills. In addition, the subtests are intended to be more "game-like" for young children. There are improvements to administration procedures over previous editions, including more clear-cut instructions for examiners and examinees. According to Thorndike's review (2014), the WPPSI-IV has many good psychometric properties. Thorndike highlights the interpretative section of the test manual as another strength since it provides clear, step-by-step guidelines for interpreting and reporting scores. Canivez (2014) also reviewed the WPPSI-IV and cited many positive characteristics of the revision, including an excellent standardization sample and good evidence of score reliability. One of the limitations of the WPPSI-IV described by Canivez is that some key information is missing from its technical manual such as exploratory factor analyses, which should have been conducted due to the rather significant revisions in subtests from the previous version of the WPPSI. Secondly, Canivez notes that the interpretation system of the WPPSI-IV, as well as similar instruments, which relies on analysis of subtest strengths and weaknesses and profile analyses, is lacking in empirical support and, thus, detracts from the clinical utility of the instrument.

Summary of Practice Guidelines for Cognitive Assessment of Young Children

- Accurate developmental and functional assessment of infants and young children is an inherently complex process that requires considerable knowledge, skill and experience. There are many unique aspects of assessing younger

children including lower attention span, higher activity levels, as well as the potential for stranger anxiety.

- Clinicians need to examine data about child cognitive abilities relative to functioning across other developmental domains and within different contexts. More specifically, scores on standardized tests and performance across different test components can provide valuable information about aptitude, relative strengths and weaknesses, and behavior, but must always be considered in light of functioning in other developmental domains (e.g., language, social-emotional, motor) in order to develop diagnostic impressions.
- The responsibility of clinicians extends beyond learning a standard assessment protocol and reliable administration. It also includes understanding of major theories and updated research about cognition and development that have been used to inform assessment methods and the design of specific instruments.
- Cognitive assessment should be grounded within an ecological framework that considers the wide range of contextual factors that influence children's functioning (e.g., current and past health status, relationships with caregivers, cultural and linguistic background, among others). Use of such a framework supports meaningful and accurate interpretation of results for diagnostic decisions and treatment planning consistent with recommended practice in early childhood assessment.
- Performance on standardized measures of cognitive abilities should be regarded as a picture of current functioning and used to develop current early intervention plans and not long-term prognosis about intelligence.

Case Study

Peter is a 31 month male who was referred to a developmental assessment clinic for concerns about social interaction and language delay. Background information indicates that Peter was born at 34 weeks gestation and spent two weeks in the NICU due to respiratory distress and difficulties with feeding. According to Peter's parents and medical records, these difficulties resolved by the time he was one year old. However, as a baby and toddler, Peter experienced repeated episodes of otitis media and took multiple courses of antibiotics. At 26 months, Peter underwent surgery for placement of PE tubes (tympanostomy tube) and had his adenoids removed. Peter's parents reported that his motor milestones were attained on time with respect to sitting up, walking, running, and using his hands to manipulate toys and utensils. Peter has demonstrated some delays in language development. He began using single words at 18 months of age. Presently, he uses two-word combinations and these are limited to toys and people that are most familiar to him. Peter does not attend daycare or preschool. He has some contact with other children through play groups. According to Peter's mother, he tends to play by himself during these groups but shows interest when other kids bring over toys that he likes.

For Peter's assessment, he was administered the Bayley-III and received the following scores:

Bayley-III scale	Composite score	Percentile rank	Description
Cognitive	85	16th	Low average
Language	73	4th	Borderline/low
Motor	92	30th	Average

Peter primarily displayed neutral affect throughout the assessment. While he appeared to be generally content, he did not reciprocate social smiles or spontaneously seek interactions with the evaluator or his mother. Eye contact was minimal. Peter's response to his own name was variable.

Discussion Questions

1. Based upon the data that has been gathered thus far, what working hypotheses do you have regarding Peter's functioning and areas of need?
2. How have the results from the Bayley-III contributed to your understanding of Peter's functioning?
3. What other information would you like to acquire to better understand Peter's functioning? What other instruments/measures might be beneficial in measuring Peter's skills and development?

References

American Psychological Association. (2012). Guidelines for assessment of and intervention with persons with disabilities. *American Psychologist, 67*(1), 43–62.

Bagnato, S. (1992). The case against intelligence testing in early intervention. *Topics in Early Childhood Special Education, 12*, 1–20.

Bagnato, S. J., & Neisworth, J. J. (1994). A national study of the social and treatment "invalidity" of intelligence testing for early intervention. *School Psychology Quarterly, 9*(2), 81–102.

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development*, (3rd ed.). Administration manual. San Antonio, TX: Harcourt Assessment.

Berls, A. T., & McEwen, I. R. (1999). Battelle Developmental Inventory. *Physical Therapy, 79*, 776–783.

Black, M. M., & Matula, K. (1999). *Essentials of Bayley Scales of Infant Development II assessment.* New York City: Wiley Publishing.

Black, M. M., & Powell, D. (2004). *Commonly used assessments and screening instruments*. Tampa, FL: University of South Florida. Retrieved from http://floridahippy.fmhi.usf.edu/grant/Instruments_FinalRevision_206.doc

Bracken, B. A. (1987). Limitations of preschool instrumentations and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment*, *5*, 313–326.

Bracken, B. A., Keith, L. K., & Walker, K. C. (1994). Assessment of preschool behavior and social-emotional functioning: A review of thirteen third-party instruments. *Assessment in Rehabilitation and Exceptionality, 4*, 331–346.

Bradley-Johnson, S. (2001). Cognitive assessment for the youngest children: A critical review of tests. *Journal of Psychoeducational Assessment, 19*, 19–44.

Campbell, G. M. (2005). Diagnostic assessment of Asperger's disorder: A review of five third-party rating scales. *Journal of Autism and Developmental Disorders, 35*(1), 25–35.

Canivez, G. L. (2014). Test review of Wechsler Preschool and Primary Scale of Intelligence—Fourth edition. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook*. Retrieved from http://marketplace.unl.edu/buros/

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Cornish, K. M., Sornberger, M. J., Dupasquier, M., & Wilding, J. (2012). Global mental functions: Intellectual (B117). In A. Majnemer (Ed.), *Measures for children with developmental disabilities: An ICF-CY approach* (pp. 49–59). London, UK: MacKeith Press.

Dale, B. A., McIntosh, D. E., Rothlisberg, B. A., Ward, K. E., & Bradley, M. H. (2011). Profile analysis of the Kaufman Assessment Battery for Children, Second Edition, with African American and Caucasian preschool children. *Psychology in the Schools, 48*(5), 476–487.

Elbaum, B., Gattamorta, K. A., & Penfield, R. D. (2010). Evaluation of the Battelle Developmental Inventory, 2nd edition, screening test for use in states' child outcomes measurement systems under the Individuals with Disabilities Education Act. *Journal of Early Intervention, 32*(4), 255–273.

Elliott, C. D. (2007). *Differential Ability Scales* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Flanagan, D. P., & Alfonso, A. C. (1995). A critical review of the technical characteristics of new and recently revised intelligence tests for preschool children. *Journal of Psychoeducational Assessment, 13*, 66–90.

Grigorenko, E. L., & Sternberg, R. J. (1999). Assessing cognitive development in early childhood. Great Britain: World Bank. Retrieved from http://files.eric.ed.gov/fulltext/ED453950.pdf

Gyurke, J. S., Marmor, D. S., & Melrose, S. E. (2004). Wechsler Preschool and Primary Scale of Intelligence—Revised. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hodapp, R. M. (1998). *Development and disabilities: Intellectual, sensory and motor impairments*. Cambridge, UK: Cambridge University Press.

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.

Kaufman, A. S., & Kaufman, N. L. (2004). *Manual for Kaufman Assessment Battery for Children—Second edition (KABC-II)—Comprehensive form*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Lichenberger, E. O. (2002). *Assessing adolescent and adult achievement* (2nd ed.). Boston: Allyn & Bacon.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. (2005). *Essentials of KABC-II assessment*. Hoboken, NJ: Wiley.

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales-II: Consistency across ages 4 to 17. *Psychology in the Schools, 47*, 676–697.

Lichtenberger, E. O. (2005). General measures of cognition for the preschool child. *Mental Retardation and Developmental Disabilities, 11*, 197–208.

Macow, G. (2008). *Bayley Scales of Infant and Toddler Development* (3rd ed.). Powerpoint slides. Retrieved November 15, 2015 from http://images.pearsonclinical.com/images/PDF/Bayley-III_Webinar.pdf

Mardell, C., & Goldenberg, D. S. (2011). *DIAL-4 training packet Co*. Bloomington, IL: NCS Pearson, Inc.

Mazer, B., Majnemer, A., Dahan-Oliel, N., & Sebestyen, I. (2012). Global developmental assessments. In A. Majnemer (Ed.), *Measures for children with developmental disabilities—An ICF-CY approach* (pp. 249–264). West Sussex, UK: MacKeith Press.

McDermott, P. A., Fantuzzo, J. W., Waterman, C., Angelo, L. E., Warley, H. P., Gadsden, V. L., et al. (2009). Measuring preschool cognitive growth while it's still happening: The learning express. *Journal of School Psychology, 47*, 337–366.

Meece, J. L. (2002). *Child and adolescent development for educators* (2nd ed.). New York, NY: McGraw-Hill.

Newborg, J. (2005). *Battelle Developmental Inventory* (2nd ed.). Itasca, IL: Riverside Publishing.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York, NY: Basic Books.

Pinon, M. (2010). Theoretical background and structure of the Bayley Scales of Infant and Toddler Development. In L. G. Weiss, T. Oakland, & G. P. Aylward (Eds.), *Bayley III clinical use and interpretation* (3rd ed.). London: Elsevier.

Puckett, M. B., & Black, J. K. (2005). *The young child: Development from pre-birth through age eight* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Reddy, L. A., Braunstein, D. J., & Dumont, R. (2008). Use of the Differential Ability Scales for children with attention-deficit/hyperactivity disorder. *School Psychology Quarterly, 23*(1), 139–148.

Shepard, L., Kagan, S. L., & Wurtz, E. (Eds.). (1998). Principles and recommendations for early childhood assessments. *The National Education Goals Panel*. Retrieved from http://govinfo.library.unt.edu/negp/reports/prinrec.pdf

Shultz, J., & Chase-Carmichael, C. (2001). Assessment of children with chronic illness. In R. Simeonsson & S. Rosenthal (Eds.), *Psychological and developmental assessment of special children* (p. 292).

Simeonsson, R. J., & Rosenthal, S. L. (Eds.). (2001). *Psychological and developmental assessment: Children with disabilities and chronic conditions*. New York, NY: Guilford Press.

Smith, T., Pretzel, R., & Landry, K. (2001). Infant assessment. In R. Simeonsson & S. Rosenthal (Eds.), *Psychological and developmental assessment of special children* (pp. 176–205).

Tusing, M. E., & Ford, L. (2004). Examining preschool cognitive abilities using a CHC framework. *International Journal of Testing, 4*(2), 91–114.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). San Antonio, TX: Pearson.

Weiss, L. G., Oakland, T., & Aylward, G. P. (2010). *Bayley III: Clinical use and interpretation*. Cambridge, UK: Academic Press-Elsevier.

http://www.springer.com/978-1-4939-6347-8

Early Childhood Assessment in School and Clinical Child Psychology

Garro, A. (Ed.)

2016, VIII, 297 p. 13 illus., Hardcover
ISBN: 978-1-4939-6347-8