Neuropsychological Assessments in Children from a Longitudinal Perspective for the National Children's Study

Roberta F. White, Ph.D., ABPP/cn
Diplomate in Clinical Neuropsychology (ABPP/cn)

Professor and Chair, Department of Environmental Health
Boston University School of Public Health

Professor of Neurology (Neuropsychology) and Psychology
Boston University Schools of Medicine and Arts and Sciences

Staff Psychologist
Director, Environmental Hazards Research Center
VA Boston Healthcare System

With assistance from:
Richard Campbell, Ph.D.
Patricia Janulewicz, B.S.
Department of Environmental Health
Boston University School of Public Health

Fall 2004

Prepared for the National Children's Study
by Battelle Memorial Institute
Under Contract No. 282-98-0019

Table of Contents

Summary of Task

This white paper summarizes the process of reviewing and selecting outcome measures that assess developmental cognitive function and offers a testing strategy for their use in the National Child Study (Study). The Study is a proposed 20-year prospective study of 100,000 children. Its major goals include a description of the natural development of cognitive skills from birth through age 20 and the monitoring of environmental influences on American children (for example, diet, social, media, chemicals). Ultimately, the data collected through the Study will allow the investigation of the relationships between the exposures and health outcomes that have been investigated.

A basic assumption of the process described in this white paper was that the cognitive assessment strategy must comprehensively evaluate dimensions of intellectual development from 6 months through 20 years of age, ensuring that appropriate aspects of cognitive development are examined at appropriate ages and that the testing strategy detects vulnerable domains of function at specific stages of development. A further assumption was that the cognitive evaluation strategy recommended for the Study should include assessment instruments that are well imbedded in the clinical and scientific literature on cognitive development. This is important because it allows interpretation of Study findings with regard to possible etiologies of cognitive dysfunctions identified during the Study. Given the fact that cognitive development is an expression of brain development, the approach used in developing the recommended assessment strategy was significantly influenced by the neurocognitive and neuropsychological literature. This approach is especially appropriate for interpreting developmental cognitive outcome data, because its emphasis on defining the brain-behavior relationships revealed by cognitive outcome measures will facilitate interpretation of Study results with regard to the neural system basis of cognitive deficits and vulnerabilities that are revealed in the Study.

Given these assumptions, the development of recommended cognitive assessment strategies consisted of several steps. At the outset, the decision was made to focus on quantitative measures that detect subtle, preclinical cognitive dysfunction. Although outcome measures that depend on clinical diagnosis of neuropsychiatric disorders utilizing specific criteria would be a possible assessment approach, this strategy is likely to miss subtle manifestations of dysfunction in children at critical ages and has little power to describe cognitive development. In approaching the cognitive assessment process quantitatively, it was necessary to define the critical domains that should be evaluated and the stages of development during which the domains should be assessed. A list of tests and test batteries that evaluate these domains was then developed, and the tests were reviewed according to a defined set of criteria.

In carrying out this process, existing clinical and scientific knowledge about child development was central. Also consulted was the literature on developmental neurotoxicology, which describes the relationships between exposure to common environmental chemicals and their effects on brain function. Studies in this field have taken advantage of the power of cognitive developmental tests as measures of brain function and have thus produced important information on domains of cognitive function that are especially vulnerable to the subtle effects of environmental influences, ages at which particular vulnerabilities may appear in specific functional domains, effect sizes of subtle effects of environmental influences, and test instruments that are especially useful to detecting subtle cognitive deficits in children.

Based on review of the developmental and neurotoxicology literature and on specific criteria developed for test selection, a subset of tests was defined and further explored for inclusion in a proposed Study assessment battery. In addition, strategies for assessing children at critical stages during development were considered. These lines of thought converged in a battery of recommended tests presented in an age-by-domain-by-test paradigm, along with a list of alternative tests that can be used at each age in each domain. It should be noted that the development of a battery to be used over 20 years has the quality of being frozen in time, since such recommendations must be made on the basis of the state of the art in the field at present. Tests and strategies obviously must be adjusted at intervals as the Study progresses. These recommendations may be a starting point.

Assessment Domains

Omnibus "IQ" versus domain-specific tests. An initial issue addressed with regard to developing the most fruitful and definitive approach to the assessment of child cognition was the tension between "general intelligence" testing (intelligence quotient or IQ determination based on standardized omnibus tests) versus the neuropsychological approach (domain-specific tests). This issue has received considerable attention in the developmental neurotoxicology literature and scientific community. IQ tests have been used extensively in the study of certain types of toxicant exposure (especially lead and PCBs), and it has been argued that IQ tests are preferable to other kinds of tests for their public health value—that is, individuals besides psychologists believe that they understand the meaning of such tests, and the results are therefore more likely to be taken seriously and acted upon by the public health community and society at large. This argument has been proffered despite the fact that subtle low-level effects of toxicants such as lead are often in the three-point IQ range, a conclusion that has caused some to argue that lead effects at these levels are meaningless. It has been pointed out that a three-point IQ decrement in a population results in a significant shift in IQs such that more individuals fall into the "mentally impaired" range (2 SDs below average) and fewer people are classified as "very superior" (2 SDs above average (Bellinger). The argument has also been made that a three percent difference is, in fact, similar to the degree of effect seen for physiological changes associated with toxicant exposure that are widely accepted as important (Bellinger).

Domain-specific neuropsychological tests have received more attention in recent years in behavioral toxicology because of their sensitivity to prenatal exposure to toxicants such as methylmercury (White and Grandjean) and nicotine (Freed). In addition, it has been argued that these tests provide more insight into the underlying central nervous system (CNS) damage that may be associated with exposures, since there is a significant literature that links impaired performance within individual domains or patterns of impaired and intact performance across domains to specific types of brain damage (structural, neural system, neurotransmitter).

It seems obvious that the Study design allows consideration of both IQ and domain-specific testing at critical stages of vulnerability during child development. This will provide the advantages of both types of assessment approaches and has the potential to reevaluate the results of existing studies.

Recommended Study battery domains. Categorizing the dimensions of cognition into component parts is challenging even to cognitive psychologists and to neuropsychologists, who often divide or subdivide these dimensions in different ways. It is even more difficult to categorize existing cognitive and behavioral tests, since few were designed to be pure tests of a specific aspect of cognitive processing or a single domain. For the purposes of developing the recommended Study battery, domains were identified using the labels commonly applied in clinical neuropsychology (Lezak, Spreen, and Strauss). In addition, careful consideration was given to the key aspects of processing that should be assessed within each domain. The following list identifies the domains selected by this author and associated defining characteristics. The definitions are not meant to be exhaustive but to give the reader an overview of the types of skills subsumed under each domain and assessed by associated neuropsychological tasks.

- **General intelligence/mental abilities/omnibus cognitive skills tests.** These tests consist of subtests with various labels purported to measure particular aspects of cognitive function. Subtest scores are summed in order to obtain overarching measures such as IQ, verbal abilities (for example, Verbal IQ, Verbal Comprehension Index), visual-motor or visuospatial skills (for example Performance IQ, Perceptual Organization Index), attention/working memory (for example, Working Memory Index, Attention Index), or speed of processing (for example, Processing Speed Index).
- **Academic skills.** This domain includes skills that are generally learned in school, such as reading words or paragraphs, spelling, and completing arithmetic problems.
- **Attention.** This domain encompasses several processes, including the capacity to focus on and attend to stimuli over a period of time (sustained attention, often assessed by Continuous Performance Tests) and capacity to take in and report back stimuli immediately after presentation (for example, Digit Spans or Visual Pointing Spans).
- **Executive function/working memory.** This is a complex domain that historically includes the capacities to learn and manipulate stimuli (for example, Digit Spans Backwards, Visual Pointing Spans Backwards), to invoke strategies for manipulating novel stimuli (any task with a structure that enhances task completion if recognized), or to solve novel problems (problem solving tests). This domain includes skills such as the ability to acquire the "set" of new tasks and to maintain the set of the task while completing it as well as the ability to flexibly switch from one set of task requirements to another. Inhibition of dominant or distracting stimuli in order to attend to critical stimuli is also generally included in this domain.
- **Language/verbal skills**. This domain includes basic linguistic abilities such as the capacity to produce phonemes; lexical development and production of words; speech comprehension, and linguistic aspects of writing and reading. Language skills are often divided into *expressive* and *receptive* components. Applied verbal skills, such as reading comprehension and vocabulary definitions, are often included in this domain.
- **Visuospatial abilities.** These skills (sometimes referred to as "nonverbal abilities") generally invoke the processing and manipulation of visual designs, the spatial or physical aspects of environmental objects, or constructional skills. These abilities are assessed by tasks such as drawing designs, recognizing objects presented in degraded form or embedded in a more

complex visual array, or assembling puzzles or block designs. Constructional tasks involve a motor output, but there are visuospatial tasks that require simply the mental manipulation of spatial information (for example, identifying the correct outline of an object presented in cut-up form, matching faces, matching angles).

- **Learning and memory.** This domain encompasses several aspects of memory function. *Declarative memory* is generally divided into anterograde and retrograde memory function. *Anterograde memory* refers to the learning of new information, retention of information over shorter and longer delays, and capacity for retrieval of information from memory stores. It can be assessed using both recall and recognition paradigms (recall paradigms get at the individual's capacity to retrieve information at will, while recognition paradigms are often better at assessing capacity for learning and retention when retrieval problems exist). Anterograde memory functions are sometimes divided into verbal and visuospatial components, generally associated with dominant and nondominant memory function, though visuospatial memory skills are also frequently affected in individuals with basal ganglia and white matter dysfunction. Anterograde memory is measured in many ways, including the presentation of stories, lists of words, designs or objects for immediate learning, with delayed recall and recognition (multiple choice) conditions. *Retrograde memory* refers to the capacity to remember events or information from earlier stages of the individual's life. It can be tested using famous faces, questions about historical events or facts, or questions about the individual's personal history. *Procedural learning and memory* refer to the individual's capacity to learn and remember a problem-solving sequence (for example, reading words in a mirror) or a motor skill (for example, driving a car).
- **Motor skills.** These abilities refer to the individual's capacity to carry out manual motor activities. Using neuropsychological tests, they are generally assessed using the hands (manual motor dexterity), with evaluation of speed and accuracy. Tasks may be relatively simple (tapping a computer key or finger-tapping apparatus), complex and requiring coordination as well as speed (pegboard tasks), or integrative (writing or typing symbols to match digits on a coding task).

Other domains have been omitted from this list but could be included if deemed relevant or if well developed and standardized tests become available. These include expressive and receptive prosody, motivation/malingering, and tactile/kinesthesia function. Tasks assessing these domains—and conditions under which they might be applied—are considered later in this document.

Test Selection Process

Overview of tests and test batteries available. Many cognitive and neuropsychological tests have been published that directly or indirectly assess the domains described above. In order to develop the Study recommended battery, a list was made including as many tests as could be located that are reasonably well known to researchers and clinicians working with children. Table 1 includes approximately 135 of these tests and scales, listed by the domains defined above and including age ranges to which they have been applied. Also included in the table are the five most commonly used batteries that have been recommended for evaluating children and adults who have suspected or actual exposure to chemical toxicants. The listings for the domains and the toxicant batteries are not exhaustive, but include all tests that the author considered for inclusion in the recommended battery. Some other tests and some other types of tests are mentioned at other places in this document, including novel tasks and tests that may be informative but do not exist as published, standardized tests.

Criteria used in test review. In order to carry out a detailed review of the tests most likely to be recommended for inclusion in the Study prospective screening battery, the tests listed in Table 1 were reviewed by the author and a shorter list of tests developed for greater scrutiny as candidate screening battery tasks. In reviewing the Table 1 tests, the author considered a number of criteria, whichare listed and described briefly.

Place of test in child development literature. Tests were considered with regard to their place in the field of developmental psychology. Tests that have been in widespread use by clinicians and researchers were given preference. This was done for several reasons. First, such tests are generally feasible with regard to administration. Second, the existing literature in which these tests are mentioned provides a wealth of information for interpreting impaired and superior performances on the tests. This information includes the relationship of test outcomes to particular types of developmental disorders (for example, attention, learning disabilities, speech and language disorders, extremes in IQ, motor deficits), neurological diseases (for example, epilepsy, brain tumors, traumatic brain injury), neuropsychiatric disorders (for example, autism, childhood depression, personality disorders, posttraumatic stress disorder, anxiety), medications, and medical conditions (for example, genetic disorders affecting cognition, metabolic disorders, respiratory diseases). Such information contributes to the tests' capacity to assist in screening (for triaging children on to other evaluations at the same time they serve as outcome measures). In addition, structure-function relationships have been described for many of the tests, relating impaired performance on certain tests (or patterns of impaired and retained performance on groups of tests) to particular structures of the CNS. This knowledge is critical in that it may allow investigators to form hypotheses concerning the structural or functional elements of the CNS that may be affected by exposures being measured in the Study. These hypotheses can serve as the basis for further investigations (for example, using sophisticated neuroimaging).

Place in the neurotoxicant literature. Although the Study proposes to measure several types of environmental exposures in addition to chemical toxicants, the literature in behavioral toxicology is particularly rich with regard to the types of outcome measures that are the subject of this document. If tests have proven sensitivity to low level and subtle effects of chemical exposures, they are high on the candidate list for exploring exposure-outcome relationships in a large epidemiological screening study. They also may have value in examining the subtle effects of other types of exposure (stress, violence, medications, drugs, stimulus deprivation or overstimulation, undernourishment/overnourishment/malnourishment, etc.).

Construct validity. Tests were preferred if they were designed to assess specific domains (or have been shown to do so). In addition, tests that assess a specific cognitive process (or a set of processes) within a domain were preferred. It was also necessary to balance the types of tests used within and across domains at specific ages.

Demographics. Tests were preferred if effects of variables such as age (in months at the younger ages) and gender have been defined/quantified. It was critical to include tests with a wide age range because the Study is designed to follow individuals from birth to age 20. Effects of parental education and intelligence were also considered, if available.

Culture/ethnicity/language. Available data on the relationships between culture and language and test performance were also considered. Are there ethnicity/cultural effects on test performance, and if so, what are they? Are there special versions of the tests for children from specific subcultures? How "culture-fair" is the test? Also considered were effects of primary languages and multi-lingualism on test performance, including availability of the tests in languages other than English. Information was not available on these variables for many of the tests, but tests with such information received special consideration. For some of the tests, published culture- or language-specific tests are not available, but the author is aware that they have been successfully applied and feasible in specific cultures/subcultures.

Psychometrics. In order to be sensitive to subtle effects of exposures, tests work best if they possess certain psychometric characteristics. These include a sufficient range of outcome scores at both ends (high ceiling, low floor) in order to identify exposure-outcome relationships. They must also be reliable (especially with regard to test-retest reliability) and show reasonable variance. Demonstrated validity with regard to other tests and with regard to cognitive, developmental, and exposure variables is also key. Tests that are well standardized were also preferred. Availability of appropriate normative values for tests performance at different ages and for other variables was considered to be important for certain purposes (for example, characterizing cohort performance relative to the U.S. population). However, for data analysis purposes, raw scores are usually the outcome of choice in a study such as the Study.

Other factors. Other test characteristics that were considered included ease of administration, acceptability to children, acceptability to parents, reasonable difficulty levels without extreme frustration for most children, efficiency (amount of information acquired within a relatively brief time period), and capacity of the test to contribute to screening/triaging. Finally, it must be stated that the author's experience using the tests in research and clinical settings played a role in the test review.

Candidate tests. Once the comprehensive list of tests had been considered with regard to the criteria listed above, a list of 42 candidate tests was developed. These tests are listed in Table 2, where they are characterized with regard to available information on the following variables: applicable age range, types of outcome scores generated by the test, populations on which normative values were based, reliability, validity, administration time, cost, and existence of literature on exposure-effect outcomes for chemical toxicants in children. Attachment 1 details information on these and other variables for the tests listed in Table 2. Table 3 lists selected studies describing exposure-outcome studies of children utilizing these tests.

Neurotoxicant batteries. The last page of Table 1 lists five test batteries that have been developed for use in studying exposure-outcome relationships in behavioral toxicology research. This is only a subgroup of the many batteries that have been proposed for this work, but represent those judged by the author to be most common in the vanguard of the behavioral toxicology literature. Although some of the tests included in these batteries were considered for or included in the recommended and alternate lists of tests for the Study battery, no single battery was deemed appropriate for the Study protocol. This decision was made based on the criteria outlined above. The two computer-assisted batteries (BARS, NES) are not well embedded in the developmental neuropsychology literature, are incompletely characterized with regard to structure-function relationships, and have other drawbacks, including differences in examinee cooperation on computerized versus human-administered tests. The WHO battery (NCTB) is outdated, appears to be culturally biased, and does not seem to detect subtle exposure effects. The AENTB and the PENTB have similar drawbacks to the NCTB, and the PENTB includes some tests for which psychometric properties are limited with regard to assessing exposure-outcome relationships.

Battery Design and Testing Strategy

Strategies for prospective longitudinal testing. The literature on exposure-outcome relationships is spotty with regard to data on age at exposure and age at which exposure effects can be detected. For example, the neurotoxicology literature suggests that lead exposure in early childhood is associated with IQ changes and that prenatal methylmercury exposure is associated with domain-specific neuropsychological effects at age 7. However, systematic studies across ages of exposure, ages at which outcomes are measured, and specific toxicants does not yet exist. Therefore, it is difficult to pinpoint critical ages at which specific types of neuropsychological outcomes should be measured.

Given the existing knowledge, it appears that the optimal strategy is to acquire outcome data at as many ages as possible. Because practice effects are large and can overwhelm subtle exposure effects, it is not recommended that each child be tested every year. A strategy that would allow the collection of data in yearly age increments but prevent practice effects is to divide the Study cohort into four groups of 25,000 children each. Testing each child every 4 years beginning at age 3–6 will result in neuropsychological outcome data on a large group of children at each age. Before age 3, it is recommended that the four groups of children be tested at 6 months, 1 year, 1.5 years, or 2.0 years. This strategy is outlined in Table 4, an age x domain x test representation of a recommended battery. Each of the four subgroups of 25,000 children within the Study cohort is designated by the letters A, B, C, and D. This strategy also allows repeat testing with certain key tasks at widely divergent ages, facilitating follow-up on exposure-effect relationships.

The testing strategy outlined in Table 4 necessarily initiates testing of different domains at different ages. For example, different tests are introduced at different ages within domains at the 3 to 6 year age interval. Executive function testing is not introduced until age 7. Testing of learning and memory is limited before age 7. These recommendations reflect the developmental curve of domain-specific skills as well as the availability of tests appropriate for certain ages. These factors are discussed in greater detail below.

The testing strategy outlined in Table 4 also allows for the introduction of different tests that assess the specific domains at different ages. Using this strategy, domain-specific findings observed on one test can be evaluated somewhat later with a similar test from the same domain, allowing for a chance to evaluate convergent validity.

Using the proposed testing strategy, it is possible to examine more than one specific aspect of cognitive processing within each domain. In designing the battery, attempts have been made to evaluate parallel aspects of cognitive processing during each age range and within each domain.

A major consideration in developing the recommended battery was the opportunity to use the neuropsychological outcome measures as “triggers” prompting complete diagnostic evaluations in children who may have specific types of developmental disorders. The discussion of the recommended battery includes consideration of ages and criteria for using outcome measures to triage children into screening for mental retardation, disorders of attention and learning, motor coordination deficits, autism, and neurological disorders.

Finally, the author’s own experience using these tests unavoidably affected the test review. Her experience includes 30 years of work in research and clinical settings assessing individuals across the lifespan. This work has included prospective evaluation of children with environmental exposures during infancy and early childhood, cross-sectional research on environmental toxicant exposures in childhood, occupational exposure studies with adults, and the long-term evaluation of neurodegenerative disorders in the elderly (Framingham Heart Study). In all of this research, neuropsychological test techniques have been applied as a method for uncovering the underlying neuropathological mechanisms of action for cognitive development. The battery described below is viewed as a starting point in planning for the cognitive and neuropsychological assessment of the Study cohort at various ages.

Recommended Neuropsychological Outcome Battery and Alternative Tests

Recommended test battery. Table 4 summarizes the recommended test battery to be administered at each age level (6 months–2.0 years, 3–20 years) for each proposed domain. Table 5 lists the recommended tests at each age and estimated times the battery will require at the indicated age level. Table 6 suggests alternative tests for each age range and domain. The reader is also referred to Table 2, which includes other details about the tests described in this section. The rationale for choosing the tests for each domain will be reviewed, along with a summary of the advantages and disadvantages of each test, for both recommended and alternative tests. A brief section will follow describing special requirements of test administration during each of the six proposed age ranges and screening possibilities during some testing cycles.

The age ranges outlined in Table 4 cover 2 years at the lowest level (with 6-month intervals for testing) and 4 years after that until age 19–20 (so that 25 percent of the cohort is tested at each age). Thus the testing for each group is planned for the following ages:

- Group A: 25,000 children tested at 0.5, 3, 7, 11, 15, and 19 years of age
- Group B: 25,000 children tested at 1.0, 4, 8, 12, 16, and 19 years of age

- Group C: 25,000 children tested at 1.5, 5, 9, 13, 17 and 20 years of age
- Group D: 25,000 children tested at 2.0, 6, 10, 14, 18, and 20 years of age.

It is noteworthy that Table 4 can also be used with other testing strategies in mind. For example, if all Study children were to be tested at fixed ages, the set of tests recommended for each designated each age could be administered.

The following sections of this document and the tables list tests under their current revision as much as possible. For example, the most recently revised Wechsler Intelligence Scale for Children is called the WISC-IV. Further, it is likely that other versions of the tests will be developed by the time the tests would be used.

Domains

Omnibus intelligence (IQ) and abilities measures. These measures are included to assess exposure effects on omnibus measures of mental abilities at ages that have proven to be critical in prior studies (0–2 and 3–6), ages at which such measures are relatively stable and should reflect IQ across childhood (7–10) and at an age when long-term effects on IQ of earlier exposures can be evaluated (15–18).

- **Age 0.5–2**. For age 0.5–2.0, the Bayley Scales of Infant Development-II are recommended. Though other scales are possible for all or some of this age range (for example, Fagan test, Brazelton Scale), the Bayley is the best standardized instrument and has been used the most extensively in exposure-outcome research in the past. For example, it has been applied to assess the effects of lead, PCBs, methylmercury, and DDE (see Table 3). The Bayley is a fairly blunt instrument and may not pick up subtle deficits associated with exposures. The items are rather diverse and do not easily lend themselves to domain-specific analysis. However, it is the best option available. It is recommended that this test be given to the four Study cohort subgroups at one time point each (6 months, 1 year, 1.5 years, or 2 years).
- **Age 3–6.** At age 3–6, an abbreviated version of the WPPSI-III (Block Designs, Matrix Reasoning and Vocabulary) is recommended. The WPPSI-III was chosen over other possibilities for several reasons. The Wechsler scales have been the most extensively applied to IQ research on this age range in the past and they dovetail nicely with Wechsler subtests available at later ages. The subtests chosen will produce an IQ score, have parallel versions available at later ages, and can contribute some information to domain-specific function (although they are far from pure measures of specific domains). The major disadvantage of the Wechsler Scales is that they have somewhat abbreviated ranges, meaning that at the lowest and highest age ranges the tests can be too difficult or too simple and that persons with low IQs or who are gifted may be unable to meaningfully complete subtests or may be able to correctly complete all or virtually all items. Other tests that were considered for this age span include the McCarthy, which has also been used extensively in developmental research. The major disadvantage of this test is the limited number and heterogeneity of items in specific subscale areas. The Stanford Binet Intelligence Scale-V was also considered. The major advantage of this test is that it can be given across the lifespan, frequently using the same subtests. It is also excellent for measuring the higher and lower ends of intelligence. Disadvantages include its lesser use in developmental research, less

information on relationships between subtest performance and developmental outcomes, and the fact that a new version has recently been produced that was not available for review at this writing. Based on promotional material from the publisher, it appears that the test has been considerably revamped and may differ significantly from the S-B IV; an important omission from the new version appears to be the Copying Test, which proved to be very sensitive in several cultures at several ages to the effects of prenatal and childhood methylmercury exposure. Finally, the Kaufman scales (KABC-2, KBIT-2) were considered. These scales have also been used less extensively in developmental research and have recently undergone significant revision, raising questions about their comparability to prior versions of the tests. The Ravens Progressive Matrices Test has been used in the past in toxicant exposure studies. The test has been successfully applied in many cultures and appears to possess inherently less cultural and linguistic bias than other intelligence tests. However, the Ravens assesses intelligence in a unidimensional fashion (a type of nonverbal executive function), and supportive psychometric data for the test is limited with regard to norms, validity, and reliability.

- **Ages 7–10 and 15–18.** At ages 7–10 and 15–18, it is recommended that the Wechsler Abbreviated Scale of Intelligence (WASI) be employed to assess the general abilities domain. This test has four subtests (Block Designs, Matrix Reasoning, Similarities, and Vocabulary) that provide continuity with the WPPSI-3 subtests recommended for the 3–6 age group. In addition, the WASI can be used across the lifespan after age 6, using the same subtests. Another possibility for ages 7–10 is the Wechsler Intelligence Scale for Children (WISC-IV), a revised version of the prior WISC scales (WISC, WISC-R, WISC-III), which have been used extensively in developmental research. The WASI was recommended over the WISC-IV due to its greater brevity and its continuity across the developmental span. Also, the WISC-IV, which was not available for review at this writing, appears to be significantly different from its earlier counterparts and may or may not assess the same constructs or be as feasible as prior versions of the test.

  For the 15–18 age range, it would be necessary to use both the WISC-IV and the WAIS-III, switching tests at age 16 or 17. This would produce less continuity in the age range testing, and both scales can be problematic for 16-year-olds (too easy or too hard).

The NEPSY was considered as an alternative test for both omnibus scores and domain-specific assessment and is seen as an alternative instrument. The test seems to have a rather low ceiling and subtest length is somewhat limited, restricting the utility of outcome data. When discussing neuropsychological domains, subtests from the IQ tests and the NEPSY can always be considered to be possible alternative tasks. The pros and cons of these tests have been described above and will not be repeated.

Academic screening. A domain for brief academic testing was included as an assist to screening for learning disorders. This domain was not designed to serve as a full assessment of academic abilities as outcome measures, although the results can be used as a cursory evaluation of these outcomes at the age ranges in which they are included. During the 7–10 year age range, brief testing of basic academic skills can be combined with results of IQ and domain-specific testing in order to identify children who may have disorders of learning. The test recommended for this domain is the Wide Range Achievement Test-3 (WRAT-3), which assesses single-word reading, single-word spelling and arithmetic. This test was selected for ease of administration, time efficiency, and its acceptance in the field. It has seen limited use in exposure studies. Alternative tests include the Woodcock-Johnson, which was used in a methylmercury study, the Kaufman Test of Individual Acheivement-2, and the Wechsler Individual Achievement Test. These tests are more complex than the WRAT-3 and less suited to screening. It is recommended that the WRAT-3 be repeated during the 15–18 age testing in order to assess stability of any exposure-related changes in basic academic skills over time.

Attention/concentration. The cognitive processes subsumed under this domain have been widely described and evaluated in the cognitive psychology literature. Because it is not possible to assess all aspect of the domain, the assessment strategy recommended in this document focuses on behavior, sustained attention/reaction time, and spans of apprehension.

- **Age 3–6.** Tests recommended for this domain among 3–6 year-olds include the Conners Rating Scale-Revised and the Conners Continuous Performance Test (CPT—II). The Conners Rating Scale is used to assess behavioral characteristics that are associated with Attention Deficit Hyperactivity Disorder (ADHD) as defined by the Diagnostic and Statistical Manual-IV (DSM-IV). Outcomes include both a score that can be used as a quantitative outcome measure and provisional diagnoses of ADHD based on cut-off criteria. Thus, the test can contribute to screening for ADHD in this age range. Testing with the Conners CPT-II begins at age 4, and the test allows evaluation of lapses in attention (omission errors), overresponding (false positives) and reaction time. Reaction times have proved to be sensitive indicators of exposures to toxicants and medications. The Conners CPT-II was recommended because of its widespread use in child clinical neuropsychology. Span of apprehension testing (Wechsler Digit Spans Forward) is not recommended for this age group due to limited applicability at ages 3–5.
- **Age 7–10.** For the 7–10 age range, it is recommended that the Conners Rating Scale-Revised be repeated in order to acquire a second set of outcome scores on attentional behaviors and to allow a second chance to pick up possible cases of ADHD that were missed at prior testing. The Conners CPT-II is also recommended at all ages due to the sensitivity of reaction time data to many types of exposures/insults/disorders. Finally, the Wechsler (WISC-IV) Digit Spans Forward test is recommended as a span of apprehension task. This task provides data on the number of bits of information that the child can automatically register and repeat back. Such data are important as outcome measures (and have been related to specific types of exposures during development). They can also be used to estimate appropriate expectations for performance on learning tests. Other possibilities include the Neurobehavioral Evaluation System (NES) letter or animal CPT, which has also been used extensively and effectively in detecting subtle toxicant effects in children and adults. The NES is less widely available and more difficult to adapt to different testing situations than the Conners. Normative and psychometric data are also less extensive for it. A visual pointing span test is another alternative.
- **Age 11–14.** During the 11–14 age range, it is recommended that the Conners CPT-II and WISC-IV Digits Forward be repeated. At age 14, the Conners Rating Scale-R can be repeated to assess stability of scores and cut-offs for ADHD diagnosis criteria.

- **Age 15–18.** At age 15–18, it is recommended that the Conners CPT-II and Wechsler Digit Span Forward be repeated (WISC-IV, ages 15–16; WAIS-III, ages 17–18). The Conners Rating Scale-R can be repeated at ages 15, 16, and 17.
- **Age 19–20.** The recommended attention test for the brief battery suggested for ages 19–20 is the Conners CPT-II.

Executive function/working memory. This domain is a complex one and the related skills tend to develop somewhat later than those subsumed under other domains. Since it would be impossible to evaluate all aspects of this domain at every age level, testing is limited to screening for a few skills at each age level. The tests chosen for the age ranges between 3 and 18 were selected to include both visually and verbally mediated tasks.

- **Age 7–10**. Testing in this domain begins in the 7–10 year age range. The Wisconsin Card Sorting Test (WCST) is a widely used task assessing this domain. It taps inferential reasoning, working memory, capacity to attain and switch sets flexibly, and ability to inhibit distractions and perseverative tendencies. Performance on the test has been related to a wide variety of neurological and developmental disorders, neuropsychiatric syndromes, and exposures to chemical toxicants. A second test, the WASI Similarities test, is administered as part of the general intelligence testing and also taps aspects of executive function. This task assesses abstract reasoning, and performance on it has been related to aspects of normal and abnormal child development. The Children's Categories Test is an alternative test for this domain. It has two levels (for ages 5–8 and 9–13), so the test stimuli would be different for the children tested at ages 7 and 8 than those tested at 9 and 10. A few of the items are slightly problematic as well. Other tests that assess this domain that can be considered as alternatives for use in the 7–10 age range include the Children's Color Trails Test (age 8–10) and the Stroop Color Word Test. The Color Trails Test appears to be relatively culture fair but has (in the author's experience) been difficult to administer (many children do not understand it initially). There is also much less information about how performance on this test relates to other aspects of childhood cognitive development than is available for other tests of executive function. The Stroop Test is commonly used as an executive test and much more is known about its relationship to other variables. However, performance on the test varies widely, affecting reliability and psychometrics of the test for data analysis purposes.
- **Age 11–14.** The executive domain tests recommended in Table 4 for the 11–14 age are Wechsler (WISC-IV) Digit Spans Backward, a working memory task requiring the registration and manipulation of verbal information, and the Trail-making Test (TMT), which requires the examinee to track and connect visual information (A condition) and to alternate sets while tracking and connecting visual stimuli (B condition). Both tests have rich sources of scientific and clinical data to support their use and interpretation in a study such as this. Drawbacks to Digit Spans Backward include resistance to the task by examinees who feel that they cannot manipulate numbers as well as the need for considerable examinee cooperation in completing it. The Trail-making Test is a timed task and optimal performance is only elicited when the examinee is willing to work as quickly and accurately as possible. It also requires overlearned, automatic knowledge of the alphabet sequence and numbers. Alternative tests include the tests described above for examining this domain in 7–10-year-olds.

- **Age 15–18.** For testing in the 15–18-year age range, it is recommended that Wechsler Digits Backwards (WISC-IV for age 15–16; WAIS-III for age 17–18), Wisconsin Card Sorting Test (WCST), and WASI Similarities be repeated. Other possibilities include Children's Categories Test (Level 2), Children's Color Trails (1–17), and Color Trails (age 18). Pros and cons of these tests have been noted above. Another alternative is the Paced Auditory Addition Test (PASAT), a rather difficult task that was developed to demonstrate subtle brain damage associated with head injuries. A strength of this test is its sensitivity to subtle processing deficits, although it requires considerable examinee cooperation and performances tend to be rather variable. It is not as well represented in the general developmental cognitive literature as some of the other tests mentioned.
- **Age 19–20.** The recommended executive domain task for the brief testing of 19–20-year-olds is the Trail-making Test, repeated 7–8 years after initial presentation to the cohort. This is a highly sensitive and efficient test (takes about 5 minutes) that is also well investigated and appropriate for young adults.

Language/verbal skills. Goals for assessment of this domain include examination of lexical knowledge, simple verbal comprehension, and ability to define vocabulary words while allowing preliminary screening for speech disorders and verbally based learning disabilities at the younger ages. The group of tests recommended for each age range includes tasks assessing both expressive and receptive aspects of language skills.

- **Age 3–6.** During the 3–6 year assessments, it is recommended that simple naming of objects be evaluated. Both tests recommended for the assessment of naming require the child to name objects presented in drawings or pictures. The naming portion of the Expressive Vocabulary Test (EVT) is recommended for ages 3–5. The EVT was chosen because of its applicability for expressive language during childhood and the particular balance of naming/synonyms in assessment of language function over development. At age 6, a different naming test must be applied. The Boston Naming Test is recommended due to its feasibility from age 6 through adulthood and its known effectiveness in detecting subtle effects of prenatal exposure to toxicants such as methylmercury. It has been applied in widely diverse cultures and subcultures and translated into many languages. Another possible test that could be used to evaluate naming in this age range is the WPPSI-III Naming Test, which can be administered at all four ages (3–6), though it does not have a parallel version for use at later ages. The Peabody Picture Vocabulary Test-III is recommended as a test of receptive language or language comprehension for this age range. It has a number of advantages, including ease of administration, well-documented reliability and validity, extensive norms, well-defined psychometrics, and widespread use in the field. Similarly, the Token Test for Children has normalized scores for ages (3–12 in 6-month increments) that was developed as a rapid screening measure of language competence particularly for children with receptive language dysfunction that depress language scores. Both tests are appropriate but the Token Test is selected as an alternate to the preferred Peabody Picture Vocabulary Test because it is less process-specific in its task demands. It is recommended that assessment of ability to provide definitions of words be carried out with the WPPSI-III Vocabulary Test, a test with the same advantages as PPVT-III. The Clinical Evaluation of Language Fundamentals-III (CELF-3) includes both expressive and receptive subtests, one of which is recommended for use at a later age level. Most of the CELF-3 subtests appear to be less appropriate for screening during the 3–6 age level than those included in the recommended list.

- **Age 7–10.** At ages 7–10, the Boston Naming Test is recommended as an assessment tool for expression/naming, the PPVT-III for comprehension/receptive speech, and the WASI Vocabulary subtest for production of word definitions. Pros and cons of these tests and alternative tasks are noted in the last paragraph.
- **Age 11–14.** Recommended tests for the language domain in the 11–14 age range include two new tasks. This allows examination of a slightly different aspect of expressive and receptive processing and reduces practice effects across the age ranges. The EVT Synonyms task (requiring the examinee to produce words with the same meaning as those given orally by the examiner for a stimulus picture) and the CELF-3 Sentence Structure subtest (a receptive task requiring recognition of words appropriate to sentences) are recommended for this age range. These tasks are less well known than the language tests recommended for earlier age ranges but have been well standardized and normed and will provide information on more complex aspects of expressive and receptive language at this age level.
- **Age 15–18.** For the language/verbal assessment at ages 15–18, it is recommended that the BNT, PPVT-III, and WASI Vocabulary subtest be repeated. The recommended brief language domain assessment at ages 19–20 includes repetition of the EVT (Synonyms) and CELF-3 Sentence Structure subtest. It would also be possible to repeat the BNT or PPVT-III or to apply other language tasks.

Visuospatial abilities. The critical processes that must be evaluated in the assessment of visuospatial abilities have been less well defined than those of other domains, and there has been considerable overlap in the stimuli used across visuospatial tasks designed for children. In considering the visuospatial tests to be recommended for the Study battery, it was deemed important to include both traditional constructional tasks (with a motor component such as drawing or putting blocks or puzzle pieces together) as well as motor-free tasks that involve visuospatial processing and integration at a cognitive level only.

- **Age 3–6.** The visuospatial tasks recommended for use at ages 3–6 include the Visual Motor – Integration Test (VMI-5). This task is well embedded in the developmental literature and has recently been revised and renormed. It has been used in prior work involving environmental toxicant exposure and was chosen partially for its similarity to the Copying Test of the Stanford Binet-IV. The latter task was highly feasible in several cultures and able to detect subtle effects of early exposures to methylmercury. The Copying Test was also valuable because it could be administered across the lifespan, a property not associated with the VMI-5. However, the two tests have overlapping stimuli and test requirements. An alternative constructional test is the Bender Gestalt-II. The original version of this test has been used extensively in both clinical and research situations and has detected effects of toxicant exposures. However, for this age range the test could only be administered at ages 5 and 6. The WPPSI-III Block Designs and Matrix Reasoning subtests are recommended for the 3–6-age range. They provide a measure of visuospatial skills with a motor component (Block Designs) and a test without motor requirements (Matrix Reasoning). Both have a strong executive component (as do many visuospatial tests). They also contribute to the IQ score recommended for this age group.

- **Age 7–10.** The Block Designs and Matrix Reasoning subtests of the WASI are recommended for the 7–10 age range. Advantages and drawbacks are similar to those described for the subtests at ages 3–6, though the WASI subtests can be given across the lifespan after age 6. A second recommended visuospatial task for the children at ages 7–10 is the Bender Gestalt-II, a visual constructional task similar to the initial version used in prior research but with additional designs added for a better range of scores and difficulty level. Experience with the revised version is still somewhat limited. It has a recall condition, an advantage that contributes to the test's efficiency in a battery such as that to be used for the Study.
- **Age 11–14.** For the 11–14-age range, three tests are recommended. Repetition of the VMI-5 is suggested as a measure of visual constructions. The Hooper Visual Organization Test (HVOT) is recommended as a motor free task assessing visual integration. The test has a somewhat low ceiling in adults, but is one of the few tests available that allows examination of visual organization without drawing or assembling concrete objects. Finally, the Rey-Osterreith Complex Figure (ROCF) test is recommended for use in this age range. The ROFC is a relatively difficult construction that increases the range of assessment of the visuospatial domain, can be given quickly, is well known to clinicians and researchers in child development, and includes memory conditions. Scoring of the ROCF can be done simply, though complex scoring systems have been developed for it. This task also produces a wealth of qualitative information that may be useful for certain kinds of data analysis.
- **Age 15–18.** The recommended visuospatial tasks for the 15–18 year assessments include the Bender Gestalt-II and two WASI subtests (Block Designs and Matrix Reasoning).
- **Age 19–20.** The brief visuospatial battery recommended for 19–20 year-olds includes the motor free HVOT and the ROCF to assess constructional ability.

Learning and memory. As noted in the domain definitions provided earlier in this document, learning and memory function involves several key cognitive processes. Luckily, several tests of learning and memory have been developed that address all or most of these functional processes. For the recommended battery, this domain is focused on anterograde memory rather than retrograde memory or procedural learning. Because children differ in their verbal and visuospatial abilities and because the cerebral structures subserving the processing of verbal and visual information are different, visual and verbal memory tests are included at each age level. Similarly, within the verbal modality, learning lists of words or word-pairs can be differentiated on a neural system or neurofunctional basis from learning discourse or paragraph material. For this reason, both word list and discourse tasks were included as much as possible at each age level.

- **Age 3–6.** The battery for this domain is relatively limited in the 3–6 year age range. Few tests are available and administration of these kinds of tests is difficult for very young children. The recommended test battery takes advantage of the use of the WPPSI-III Coding Test at ages 4–6 (see Motor domain, below) to carry out incidental learning of the symbol-symbol pairs (visual memory task). At ages 5–6, administration of the California Verbal Learning Test (Children's version) (CVLT-C) is recommended. This is a list-learning test devised to comprehensively assess learning and memory at several levels, with learning immediate and delayed recall conditions as well as spontaneous and recognition test paradigms. Thus it is a very rich test that provides considerable information. It is somewhat time-consuming, and some individuals resist list-learning tests, but its advantages were judged to outweigh disadvantages. The Wide Range Assessment of Memory and Learning-2 (WRAML-2) Stories subtest is also recommended for ages 5–6. The WRAML (previous version) has been used in research of this type and appears to be solid with regard to psychometrics and standardization. It was recently revised but a review of the test suggests that the recommended subtests are not significantly changed.

- **Age 7–10.** At ages 7–10, the breadth of testing in this domain widens considerably. Recommended tests for the assessment of visual learning and memory include the recall condition of the Bender Gestalt-II and incidental learning of the visual pairs from the Wechsler-IV Coding subtest. For verbal learning, the CVLT-C list-learning test and WRAML-2 Stories test are again recommended (stimuli for the WRAML-2 Stories test change at age 9).
- **Age 11–14.** Recommended visual memory tasks at age 11–14 include the recall condition of the ROCF and incidental recall of the pairs from the WISC-IV coding test. Verbal memory tests included at this age range in the recommended battery include the WRAML-2 Stories subtest and the WRAML-2 Verbal Learning subtest (a list-learning task).
- **Age 15–18.** It is recommended that visual memory for the 15–18-year-olds be assessed using the immediate and delayed recall conditions of the Bender Gestalt-II and incidental recall from the Wechsler Coding subtest (WISC-IV at age 15–16, WAIS-III at age 17–18). Assessment of verbal list learning using the adult version of the California Verbal Learning Test II (CVLT-II) is recommended. For narrative or discourse learning, repetition of the WRAML-2 Stories subtest is suggested for 15–16 year-olds, and the Logical Memory subtest of the Wechsler Memory Scale-III (WMS-III) for the 17–18 year-olds.
- **Age 19–20.** The recommended brief battery for assessment of learning and memory at ages 19–20 includes the ROCF immediate and delayed recall, CVLT-II, and WMS-III Logical Memory.

Motor skills. There are few quantified fine motor tests for children or adults. In designing the batteries to assess motor function, the decision was made to recommend a simple test of motor speed, a more complex test that invokes both speed and dexterity, and an integrative task. All recommended tests evaluate manual motor speed with the hands, with at least one test at each age level allowing comparison of the right and left hands.

- **Age 3–6.** The recommended assessment of motor function at age 3–6 includes the Revised Purdue Pegboard. This test is a local adaptation and test instruments may have to be constructed for the Study, but it but may prove useful in assessing this domain in small children.
- **Age 4–10.** At age 4, administration of the Wechsler Coding subtest begins (using the WPPSI-III subtest). This task is highly sensitive to many brain insults and developmental conditions but is not very specific with regard to localization or diagnosis. The test itself is completed manually and timed, but performance improves if the child uses an effective strategy, has strong visual orientation and/or visual scanning skills, and can remember the stimuli. Thus, the task requires integration of many abilities. However, its inclusion is recommended based on its sensitivity. An incidental memory condition can be used to enhance the assessment of the memory domain as well (see above).
- **Age 11–20.** The recommended group of tests for evaluating this domain is the same from ages 11–20. It includes the Fingertapping Test for a simple assessment of manual motor

speed with each hand. Computerized versions provide especially precise data on Fingertapping and a system such as the Neurobehavioral Evaluation System (NES) might be used, but it is important that the testing be the same at all test sites. The mechanical and automated tappers are standard, but sometimes malfunction and may produce different data at different sites. Alternatives include the NEPSY Fingertapping Task (done with the examinee's fingers), but this test is much less precise. The Grooved Pegboard is recommended to assess motor speed involving dexterity/coordination with each hand. A standard form board can be purchased. This test is widely used. It is recommended over the Purdue Pegboard because it picks up more subtle brain damage. The Santa Ana Form board has a special place in the neurotoxicology literature and, like the Grooved Pegboard, is more challenging than the Purdue. Standard versions of it are not available, though they could be constructed for Study testing sites.

Other Categories

Novel tests. This category was included in the test battery planning matrix in order to indicate propitious times at which Study participants could undergo other kinds of tests when they had not been tested for a while. Thus, children from Group A, who are tested at age 3 and not scheduled for standard evaluation again until age 7, might undergo another test battery at age 5. Some kinds of tests that might be used at these times are described below.

Pilot testing. Tasks to be used later in the study or tests that are revised could be piloted at these time intervals.

Evaluation of domains not included in the standard battery. Test batteries to assess domains such as prosody (expression, comprehension), tactile and kinesthetic functions, retrograde memory, and procedural learning might be applied as deemed appropriate or useful. In addition, a test of purposeful test failure was not included in the recommended battery since there does not seem to be motivation for cohort members to fail tasks. However, under certain circumstances, the introduction of such a test might be warranted.

Animal test techniques. Tests adapted from animal studies could also be applied and piloted. These might include existing operant tasks and tests such as the Delayed Recognition Span Test (a learning task based on delayed nonmatching to sample methodology).

Special studies. Approved R01s assessing special abilities, tasks exploring the cognitive processes thought to subserve domains, and many other kinds of testing might be applied at these times.

Age ranges. As noted above, the six age ranges were selected to allow testing of each child every 4 years from age 3–6 to 15–18, with additional testing at age 0.5–2.0 and 19–20. It is recommended that the age band before and after the target age be restricted as follows: ± 1 month at age 0.5–2.0 and +0–3 months at ages 3–20.

Screening considerations. Test performance, especially at earlier ages, can be used as a signal that further evaluation is necessary for possible disorders or illnesses. At age 0.5–2.0, Bayley-II results could be used to tag children for evaluation of mental retardation (>2 SDs below average on the Mental Development Index), cerebral palsy or motor disorders (>2 SDs below average on the Psychomotor Development Index), or autism (<10 percentile on the behavior rating scale).

During the 3–6-age range, Conners Rating Scale-R scores below cut-off can tag children for evaluation of ADHD. Children scoring >2 SDs below average on the Conners CPT-II or making large numbers of errors can be considered for evaluation of other types of disorders (autism, neurological conditions such as epilepsy). Individuals with IQs below 70 on the WPPSI—III can be tagged for possible evaluation of mental retardation. Large discrepancies between verbal and visuospatial skills on the WPSSI-III subtests accompanied by similar discrepancies on the domain-specific tests could be used to identify children at risk for speech disorders, learning disabilities, neurological disorders, or autism.

During the 7–10 year evaluations, children with borderline intelligence or who are gifted can be identified using WASI data. Children with large discrepancies between verbal and performance measures on the WASI who also show large discrepancies between reading and arithmetic on the WRAT-3 can be considered for evaluation of possible learning disorders.

Considerations for Test Implementation and Review

The following suggestions apply to implementation and data collection of neuropsychological/cognitive data during the Study.

Pilot work. All tests should be piloted on a representative sample of children before they are given to each group.

Revised versions of tests. Virtually all of the tests included in a planned battery will be revised, sometimes changing the task demands and stimuli quite significantly. This will require piloting of new tests to evaluate their comparability to the task already in use. It may be necessary to decide whether to continue to use an “old” version of a test or to implement the revised version. Obviously, continuing to use a test in the same form that it has been previously given will enhance comparability of test results across development. However, revised versions of tests may be more useful for reporting or applying findings or may better fit new theories or knowledge about brain-behavior relationships.

Adapting tests. Tests and test scoring rules may not apply to specific cultural or linguistic groups. Care will be needed in addressing these issues within the cohort, again with pilot testing, review of protocols with regard to scoring, and translation or development of language- or culture-appropriate tests for certain groups of children.

This document has noted that some children will have special kinds of cognitive disorders that will affect their performance. There will be some children that cannot complete even the lowest levels of the tasks recommended in the present battery. For children who are identified with extremes in IQ (especially children with borderline IQ or mental retardation but possibly also children with very superior intelligence), it may be necessary to develop specialized batteries for assessing domain-specific functions. Children with learning disabilities or ADHD but normal intelligence can be presented with the battery tasks, though of course their disabilities will affect performance levels and types of errors.

Sensory deficits may also present difficulties. Tasks can be adjusted for children who cannot see or hear or who have motor deficits. These adjustments should be standardized for Study children and noted in data summaries.

Data collection. Verbatim recording of examinee responses and the use of answer sheets that allow the examiner to record approaches to tasks and other aspects of performance will enhance the data set. Generally, raw scores are most useful in studies of these kinds for data analysis, though conversion to normative scores may be useful for some purposes. It is generally impossible to record all qualitative data in databases, though the most relevant should be determined and included (for example, number of phonemic errors on a naming test). When data collection is carried out in a careful manner, qualitative findings can be reviewed and summarized later if specific questions arise.

Examiners. Educational requirements for examiners should be carefully considered. Licensed doctoral level psychologists with training in assessment might be preferred since they tend to respect the standardization rules and other critical aspects of testing. They are also more likely to note possible disorders in children as they are testing them. It seems unlikely that resources will be available to employ doctoral level examiners. However, master's level licensed psychometrists are an option as are highly trained bachelor's level research assistants. Extensive training and monitoring of examiners will be necessary, along with evaluation of intra- and interexaminer reliability of examiners. Videotaping of testing is recommended.

Examinee effort. Effort on the part of the examinee will affect scores. It may be useful to employ brief rating scales for indications of effort to be filled out by both the examiner and the examinee.

Attachment 1: Summary Characteristics of Individual Tests

Test 1

Test name. Bayley Scales of Infant Development-II
Target ages. 1–42 months
Scoring. Raw scores from a Mental Scale produce a standardized Mental Development Index (MDI), and raw scores from a Motor Scale produce a Psychomotor Developmental Index (PDI). The raw score for each scale consists of the number of items passed between basal and ceiling level, plus all items below basal level. Indexes are presented as standard scores. A 30-item Behavior Rating Scale (BRS) produces a composite total score that is converted into percentile ranks.
Normative comparison groups available. Standardized on 1,700 infants and children ranging in age from 1 year–42 months, with 17 age levels and 100 individuals (50 male and 50 female) at each level. The sample was representative of geographic, ethnic, and SES characteristics of the 1992 U.S. census.
Languages. English only
Reliability. Mental Scale: Internal consistency reliability coefficients:

- .78–.93 (.88) for Mental Scale
- .75–.87 (.84) for Motor Scale
- .82–.92 (.88) for Behavior Rating Scale

Test-retest reliability coefficients are more variable and lower for the BRS than for the Mental and Motor scales.
Validity. Concurrent validity for MDI with other scales is satisfactory to good: .49 for Preschool Language Scales, .79 with McCarthy Scales, and .73 with WPPSI. Correlation coefficients between BSID-II and BSID of .62 in national sample and .70 in a study of infants and toddlers from low-income families (Black et al., 2000)
Administration time. 25–30 minutes for <15 months; up to 60 minutes for 16–42 months.
Cost: $1,025 per complete kit.

Test 2

Test name. Bender Visual-Motor Gestalt-II
Target ages. 4–adult
Scoring. For Copy and Recall tests, subject drawings are rated on a 5-point scale, with zero points for no resemblance and four points for nearly perfect representation. A Motor Test is scored with one point for correct items and zero points for incorrect responses. Raw scores are calculated by summing correct responses, with a total maximum of 12 points. A Perception Test also assigns zero or one points, with a raw score maximum of 10 points.
Normative comparison groups available. Standardization sample of 4,000 individuals ranging from 4 years to 85+ years, representative of U.S. 2000 census on the basis of age, gender, race/ethnicity, geographic region, and socioeconomic status. Norms are available in single-age categories from 4 to 17 years of age, from 17 to 19 years, and from 20 to 24 years.
Languages. English only
Reliability. Interscorer reliability correlations ranged from .83 to .94 (average of .90) when five experienced scorers examined drawings of 30 individuals. Test-retest reliability coefficients in four age groups (4–7 years, 8–17 years, 18–49 years, 50 and older) with 2- to 3-week intervals

ranged from .80 to .88 for Copy phase and from .80 to .86 for Recall phase. Internal consistency evaluated by split-half procedure indicted overall reliability of .91, with standard error of 4.55.
Validity. Copy phase shows higher correlation with other visual-motor measures than Recall phase, which test authors argue is an indication that Recall phase measures a similar but distinct ability from visual-motor ability. Study of Bender-Gestalt II Global Scoring System in relation to Bender-Gestalt Test (with Koppitz scoring method) in 76 subjects produced correlations of .80 for Copy phase and .51 for Recall phase. A study of Bender-Gestalt II in relation to Beery-Buktenica Developmental Test of Visual-Motor Integration (Revised) produced correlations of .65 for Copy phase and .44 for Recall phase.
Administration time. 10–15 minutes
Cost. $130.00 complete kit.

Test 3

Test name. Boston Naming Test
Target ages. 5–adult
Scoring. Total naming score based on sum of correct responses between baseline and ceiling items, plus total number of test items preceding baseline.
Normative comparison groups available. Norms available in relation to gender, age, geographic region, education level, and living environment. Gap in normative data for 14–17 year-olds.
Languages. English, Spanish
Reliability. Correlation between test versions of .92
Validity. High correlation with verbal ability tests
Administration time. 10–20 minutes
Cost. $219.00 for kit.
Race/ethnicity. Cultural and regional differences found to significantly incorrect responses in comparison patients of different racial and ethnic backgrounds (Azrin et al., 1996). African-American children averaged 21 points below scores obtained on Beery Picture Vocabulary Test (Gioia et al., 1996).
Sensitivity/specificity. Successfully identified children with language or language and reading disorders, with sensitivity of .76 percent (Cooper and Rosen, 1997). Impairment on BNT in children with dysphonetic dyslexia (Cohen et al., 1988).

Test 4

Test name. California Verbal Learning Test, Children's Version (CVLT-C)
Target ages. 5–16:11
Scoring. Hand-scored. Scoring software is available.
Normative comparison groups available. Nationally normed age-based sample of 920 children, closely matching race/ethnicity, parental education, and geographic region data from 1988 U.S. census.
Languages. English
Reliability. Reliability coefficients averaged .88 for split-half and .85 coefficient alpha across age spans. Test-retest scores (28 day average interval) ranged from .31 to .90.
Validity. Principal component factor analysis, with six-factor solution similar to adult CVLT, indicates construct validity.

Administration time. 30 minutes
Cost. $159.00 for kit
Gender differences. Performance found to favor girls, but effect sizes were small and limited to age clusters rather than specific age groups (Kramer et al., 1997).
Illness/injury. Compared to controls, children with mild head injuries were able to recall fewer words following delay, and children with severe head injuries made more intrusion errors and showed poorer new learning, delayed recall, and recognition results (Yeates et al.,1995).

Test 5

Test name. Children's Category Test
Target ages. 5–16:11 (Level 1: 5–8; Level 2: 9–16:11)
Scoring. Raw scores (from 5 subtests for younger children and 6 subtests for older children) converted to standard scores, with mean (M) of 50 and SD of 10. Single *T* score results from administration of entire test.
Normative comparison groups available: Standardized on sample of 920 children, with stratified random sampling procedures utilized to ensure age, gender, race/ethnicity, parental education, and geographic representation, consistent with 1988 U.S. census. Age- and conormed with CVLT-C.
Languages. English
Reliability. Internal consistency correlations averaged .88 across 1-year ages levels for Level 1 (.86–.91) and averaged .86 for Level 2 (.81–.89). Test-retest results for 8, 12, and 16 year olds (at 10- to 42-day retest intervals) averaged .75.
Validity. Low negative correlations in comparison with WISC-R Vocabulary and CVLT-C, and low positive correlations (.14–.27) found in relation to WISC-III Full Scale, Verbal, and Performance scores. Significant correlations for Level 1 (.881) and Level 2 (.716) total scores in comparison with original Category Test, but 25 percent of Level 1 variance and 50 percent of Level 2 variance not explained by correlation between current and original forms. Recent research found questionable criterion validity of Subtest IV and suggests caution in interpreting composite scores that may be influenced by Subtest IV errors (Moore et al., 2004).
Administration time. 15–20 minutes
Cost. $375.00 for kit
Sensitivity/specificity. Findings mixed on CCT as test of frontal lobe dysfunction in children, with evidence for and against this association (Baron, 2004).

Test 6

Test name. Children's Color Trails Test
Note: Form K is standard test for clinical evaluation and for which normative data are available. Forms X, Y, and Z are designed for research purposes.
Target ages. 8–16
Scoring. Based on time in seconds for completion of test parts.
Normative comparison groups available. Normative data collected from 678 healthy children ranging from 8 to 16 years of age and more than 500 children from clinical groups participating in validation studies.
Languages. Language-free test

Reliability. Alternate form reliability between Form K and CCTT-1 of .85 and between Form X and CCTT-2 of .90. Test-retest reliability scores from .90 to .99 at various time delays.
Validity. High correlation with Trail Making Test. Discriminant validity suggested by comparisons between clinical and normal control groups (Williams et al., 1995).
Administration time. 5–7 minutes
Cost: $135.00 for kit
Sensitivity. Sensitive in distinguishing altered neuropsychological functioning (Williams et al., 1995).

Test 7

Test name. Clinical Evaluation of Language Fundamentals 3
Target ages. 6–21
Scoring. Subtests age scaled scores and scaled scores, expressive and receptive scaled scores. Normative comparison groups available: Normed on 2,450 students, 200 each year for ages 6–16 and 50 each year for those between ages of 17 and 21:11. Normative sample stratified to be representative of U.S. population with respect to age, gender, race, ethnicity, geographic region, and parental education.
Languages. Spanish
Reliability. Internal consistency for Receptive Language, Expressive Language, and Total Language range from .83 to .95. Internal consistency for individual tests scores range from .54 to .91. Test-retest results found consistently higher scores on second administration. Interrater reliability based results from two scorers produced correlations from .70 to .99.
Validity. Concurrent validity supported by strong correlation with WISC-III (r = .75), indicating that general verbal ability is being measured.
Construct validity. Factor analysis indicates that a single factor accounts for approximately 50 percent of variance, leading authors to suggest that the test measures one dominant skill, called "language ability." Misclassification of 29 percent when scores of age- and gender-matched children who had been diagnosed with language disorder were compared to scores of children who had no such diagnosis.
Administration time. 30–45 minutes for complete battery
Cost. $348.00 for kit

Test 8

Test name. Color Trails Test
Target ages. 18+
Scoring. Based on time in seconds for completion of test parts (Trails 1 and Trails 2). Also includes a qualitative scoring component, based on number errors, near-misses, corrections, and prompts.
Normative comparison groups available. Normed on 1,528 subjects ranging from 18 to 89 years, with African-American and Hispanic subsamples.
Languages. Language-free test. Administration instructions available in Spanish.
Reliability. 2-week test-retest reliability of .64 for Trails 1 and .79 for Trails 2.
Validity. Convergent validity via correlation with Trail Making Test, but shared variance was modest. Criterion validity indicated through comparison of results with normal controls and brain-injured group.

Administration time. 3–9 minutes
Cost. $104.00 for kit
Culture fairness. Uses colored circles rather than letters to minimize dependence upon knowledge of English language. Found to be generally culturally fair in study of Chinese- and English-speaking adults (Lee et al., 2000).

Test 9

Test name. Conners' Continuous Performance Test-II
Target ages. 4–adult
Scoring. Raw scores, t-scores, and percentiles for indicators of response accuracy, pattern, and reaction time (including errors of omission and commission).
Normative comparison groups available. Normed on 2,686 clinical and nonclinical subjects, with data organized by gender and age.
Languages. Language-free test
Reliability. Test-retest correlation coefficients (3-month interval) generally range from .55 to .84, but were as low as .05 for the Hit Standard Error Inter-Stimulus Interval Change score.
Validity. Has demonstrated ability to distinguish individuals with ADHD from general population and sensitivity to effects of treatment (Conners, 1995).
Administration time. 14 minutes
Cost. $595.00 for kit
Gender differences. Gender effects reported, and scores should be evaluated relative to same-gender peers in general population (Spreen and Strauss, 1998).

Test 10

Test name. Conners' Rating Scales-Revised
Target ages. 3–17
Scoring. Hand-scored scales, in long and short forms. Computer programs available to calculate standardized T-scores from raw scores.
Normative comparison groups available. Normative sample of 8,000 cases of parents, teachers, and students from more than 200 schools in over 45 states and 10 Canadian provinces. Separate norms for boys and girls, provided in 3-year intervals. Median parental income of participating parents is higher than median income in United States.
Languages. English, Spanish, and French Canadian versions
Reliability. Internal reliability coefficients for long form ranged from .73 to .94 and for short form from .86 to .94. Test-retest reliability (6- to 8-week delay) range from .47 to .92.
Validity. Factor analyses and inter-correlations among long- and short-form subtests indicate acceptable validity.
Administration time. 20 minutes or less
Cost. $203.00 for complete CRS-R Parent, Teacher, and Adolescent Users Package

Test 11

Test name. Delayed Recognition Span Test
Target ages.
Scoring.
Normative comparison groups available.

Languages.
Reliability.
Validity.
Administration time.
Cost.
Sensitivity/specificity. Evidence of sensitive indicator of deterioration of visual memory capacity, with a recent study finding decline in performance from beginning to end of shift in emergency room interns (Rollinson et al., 2003).

Test 12

Test name. Developmental Test of Visual-Motor Integration, 5th Edition Revised
Target ages. 3–17:11
Scoring. Pass/fail (one point or zero points) for copies of geometric designs
Normative comparison groups available. Standardized on national sample of 2,614 individuals from 3 to 17 years of age, representative of the 1990 U.S. census. Normative data provided as standard scores with M of 100 and SD of 15 through 17 years, 11 months. Visual-motor (VMI) raw score age equivalents also provided.
Languages. English only
Reliability. .82 coefficient alpha and .88 odd-even split-half correlation (internal consistency). Interrater reliabilities of .94 for VMI, .98 for Visual subtest, and .95 for Motor subtest.
Validity. Construct validity indicated by correlations with WISC-R non-verbal intelligence (performance IQ) of .66 for VMI, .58 for VMI Visual, and .55 for VMI Motor. Concurrent validity shows moderately high correlations with copying subtest of Developmental Test of Visual Perception (DTPV-2), with R of .75, and drawing subtest of Wide Range Assessment of Visual Motor Abilities (R = .52). Concurrent validity of .60 with WPPSI-R (Aylward and Schmidt, 1986).
Administration time. 10–15 minutes/test for Short or Full Format tests; 5 minutes/test for Visual Perception and Motor Coordination subtests.
Cost. $99.50 for kit

Test 13

Test name. Expressive Vocabulary Test
Target ages. 2–4 Naming; 5–90 Synonyms
Scoring. Scaled scores, percentiles, age equivalents
Normative comparison groups available. Normative data collected from 2,725 individuals.
Languages. English
Reliability. Alpha reliability = .90–.98; split-half reliability = .83–.97; test-retest reliability = .77–.90.
Validity. Correlation with PPVT-2 = .61–.88
Administration time. 10–20 minutes
Cost. $159.99 for kit
Sensitivity/specificity. Sensitive for speech impairment, language delay, language impairment, MR, and LDs (reading).

Test 14

Test name. Finger Tapping Test (manual)
Target ages. 6–adult (norms, but can use for younger)
Scoring. The mean number of taps from five consecutive 10-second trials
Normative comparison groups available. Children's norms—120 right handed children. The representativeness of the sample is not discussed (Finlayson and Reitan, 1976).
Languages. Language-free test
Reliability. Test-retest reliability coefficients = .58–.93. Reliability coefficients between dominant and nondominant hands = .50–.70.
Validity.
Administration time. 10 minutes
Cost. $95.00
Gender differences. In general males outperform females
Sensitivity/specificity. Sensitive to the presence and laterality of brain lesions (Barnes and Lucas, 1974; Reitan and Wolfson, 1994, 1996). The test can distinguish patients with motor dysfunctions of cerebellar, basal ganglia, and cerebral origins from the normal population (Shimoyama et al, 1990). There is evidence that this test can predict daily living skills in geriatric patients referred for possible dementia (Searight et al., 1989) and in traumatic brain injured patients (Prigatano et al., 1990).

Test 15

Test name. Grooved Pegboard Test
Target ages. 5–adult
Scoring. Scores based on time to completion
Normative comparison groups available. Normative data in the manual include norms based on 2,500 adults and children, but no information provided on IQ, socioeconomic status, educational level, or ethnicity (Trites, 1977). Norms also available for 5- to 14-year-old children (Knights and Moule, 1968).
Languages. Language-free test
Reliability. Limited information available. Test-retest reliability coefficients from clinical trial (second test followed a 6-week trial of methylphenidate) were .80 for preferred hand and .81 for alternate hand. Reliability coefficients for number of errors were .20 for preferred hand and .21 for alternate hand (Knights and Moule, 1968).
Validity. Test manual indicates that validity supported by studies discriminating head injured children from those in normal population and in poorer performance by children with head injuries or minor cerebral dysfunction, but the studies are not cited as sources of test validation (Trites, 1977). A subsequent study found that the GPT was able to discriminate between children with severe head injuries from those with mild or moderate injures (Knights et al., 1991).
Administration time. Trial discontinued after 5 minutes
Cost. $129.00
Gender differences. Females have been found to perform faster than males (Bornstein, 1986; Ruff and Parker, 1993). Finger size has been reported to be a crucial factor in gender performance differences, with smaller fingers of females argued to ease handling of pegs (Peters et al., 1990).

Sensitivity/specificity. Sensitive indicator of general cognitive slowing due to illness, including Parkinsonism (Matthews and Haaland, 1979) and HIV infection (Miller et al., 1990). Has also shown utility in identifying lateralized impairment (Haaland and Delaney, 1981).

Test 16

Test name. Hooper Visual Organization Test
Target ages. 5+
Scoring. Total raw score by summing correct items (.5 credit for 11 of 30 items), with raw score then converted into corrected raw score and T-score, adjusting for age and education.
Normative comparison groups available. Two studies provide norms for children, one based on 434 children from five–13 years of age (218 boys and 216 girls) and the second (in Eastern Canada) based on 207 children age 5–11 years.
Languages. English
Reliability. Split-half reliability coefficients of .82 for original nonclinical sample. Reliability of .78 for psychiatric inpatient sample and .80 for a group comprised of hospital employees and neurologically impaired, psychologically disturbed, and normative medical patients.
Validity. Has loaded significantly with WAIS-R Block Design (Johnstone and Wilhelm, 1997), Trails A and B, and Raven's Standard Progressive Matrices in recent research. Highly correlated with WISC block design subtest in pediatric populations (Seidel, 1994).
Administration time. 10–15 minutes
Cost. $185.00 for kit
Gender differences. Kirk (1992) found 13-year-old boys approached adult levels, whereas girls' scores were significantly lower. Seidel (1994) found no gender differences up to 11 years of age.
Sensitivity/specificity. Does not discriminate well between specific neurological populations (Baron, 2004), and considered a measure of global visual-spatial impairment in adults (Johnstone and Wilhelm, 1997).

Test 17

Test name. Kaufman Assessment Battery for Children 2
Target ages. 2.5–12.5
Scoring. Raw scores consist of ceiling items minus errors, which can be translated into scaled scores (M = 10, SD of 3) for mental processing subtests and standard scores (M = 100, SD of 15) for achievement subtests.
Normative comparison groups available. Normative sample of 1,981 children representative of 1980 U.S. census on basis of geographic region, gender, socioeconomic status, race/ethnicity, and community size. Variety of supplementary norms available.
Languages. English
Reliability. Odd-even correlations averaged in .80s and .90s within 1-year age groups for global scores. Test-retest reliabilities ranged from .77 to .97 for global scores.
Validity. Predictive validity for standard school achievement 12 in future ranged from .21 to .70 for mental process composite score. Correlation of .74 between test composite/mental processing with Stanford-Binet-IV in children with learning disabilities.
Administrative time. 35–75 minutes
Cost. $470.00 for full kit

Test 18

Test name. Kaufman Test of Educational Achievement-2
Target ages. 4.5–25 (Comprehensive Form); 4.5–90 (Brief Form)
Scoring. Raw scores can be converted into age- and grade-based standard scores, with M of 100 and SD of 15. Norms updated in 1995–1996, mirroring 1994 U.S. census data with respect to gender, race/ethnicity, and parental education, though no indication of geographic representation.
Normative comparison groups available. Age and grade equivalents, percentile ranks, normal curve equivalents, and stanines.
Languages. English only
Reliability. Reliability coefficients not updated for new normative sample. Internal consistence coefficients range from .77 to .85 by grade level and from .82 to .88 by age group. Test-retest coefficients (1- to 35-day interval) were generally .90 or higher. Reliability coefficients between K-TEA Brief and Comprehensive forms ranged from .87 to .96 for different grade levels and from .90 to .97 for separate age groups.
Validity. Correlation with other achievement tests (including WRAT, Stanford-Binet, and K-ABC) range from .83 to .88.
Administration time. Comprehensive Form. 25 minutes for Pre-K to K; 50 minutes for grades 1–2; 70 minutes for Grades 3+. Brief Form. 20–30 minutes.
Cost. $299.99 for full kit

Test 19

Test name. Kaufman Brief Intelligence Test-2 (K-BIT-2)
Target ages. 4–90
Scoring. Raw scores (0 for incorrect, 1 for correct) are converted to standard scores. Age-based standard scores for each subtest and composite are provided, with M of 100 and a SD of 15. Scores also reported by percentile rank, normal curve equivalents, stanines, and descriptive categories.
Normative comparison groups available. Norms based on national non-random sample of 2,022 subjects from 4 to 92 years, tested at 60 sites in 29 states. Sample stratified by sex, geographic region, SES, and race/ethnicity, but northeast region of country slightly underrepresented.
Languages. English only
Reliability. Split-half reliability coefficients in norming sample:

- .89–.98 (M = .93) for Vocabulary subtest
- .74–.95 (M = .88) for Matrices subtest
- .88–.98 (M = .94) for IQ Composite

Test-retest reliability coefficients:

- .86–.97 for Vocabulary subtest
- .80–.92 for Matrices subtest
- .92–.95 for IQ Composite

Validity. Concurrent validity studies indicate correlation of K-BIT scores with established school achievement and intelligence tests.
Administration time. 15–30 minutes
Cost. $162.00 for full kit

Test 20

Test name. McCarthy Scales of Children's Abilities
Target ages. 2.5–8.5
Scoring. Hand scored. Six scale scores, verbal, perceptual-performance, quantitative, memory, motor, and composite (General Cognitive Index). Raw scores are summed to create composite raw scores.
Normative comparison groups available. Standardized on sample of 1,032 children stratified by race, geographic region, father's occupational status, based on 1970 U.S. census data distributions. Children excluded if they had severe emotional/behavioral problems, known brain damage, physical defects, or could not speak and understand English. Note: Dated nature of normative data has led to concern that the MSCA is a limited assessment tool relative to more currently normed instruments (Baron, 2004).
Languages. English only
Reliability. Split-half reliability ranged from .79–.93 for subtests and General Cognitive Index (GCI). Test-retest (1-month interval) ranged from .75–.90.
Validity. Concurrent validity with Stanford-Binet and WPPSI IQ scores. Concurrent validity also indicated by high correlation with PPVT and with reading and mathematics subtests of California Achievement Test. Factor analytic studies provide support for construct validity.
Administration time. 45–75 minutes (manual); 60–90 minutes (09 Mental Measurements Yearbook).
Cost. $630.00
Sensitivity. Sensitive to effects of prenatal adversity (Moe and Smith, 2003)

Test 21

Test name. NEPSY
Target ages. 3–12
Scoring. Raw scores converted into scaled scores for domains and subtests. Computer scoring program available to generate derived scores.
Normative comparison groups available. Stratified random sample of children based on 1995 U.S. census.
Languages. English, Finnish
Reliability. Moderate stability coefficients, ranging from .67 to .76. Moderate high internal consistency scores range from .69 to .91.
Validity. Criterion validity assessed in comparison to Finnish (1988) version, with high correspondence among most subtests. Content validity assessed by literature and review by expert panel. Preliminary quantitative indication of validity in recent study that differentiated between controls and children with scholastic concerns or neurological conditions.
Administration time. 1 hour (45 minutes for core assessment) for ages 3–4; 2 hours (65 minutes for core assessment) for ages 5–12.
Cost. $639.00 for kit

Test 22

Test name. Neurobehavioral Evaluation System
Target ages. 7+
Scoring. Automatic time and error scores computed by NES administration program.

Normative comparison groups available. Limited
Languages. English, Spanish, Polish, Faroese
Reliability. Test-retest correlations in a variety of laboratory studies range from .87 to .92 for subtests with one-day interval, .68 to .88 for two-day interval, .55 to .87 for 2-week interval, and .51 to .88 for a.m.–p.m. intervals. In field studies, test-retest coefficients from .86 to .87 for various subtests with 3-day interval and from .66 to .85 in a.m.–p.m. intervals.
Validity. Demonstrated sensitivity to variety of known neurotoxicants. Moderate correlations with manually administered standardized tests.
Administration time. 50 minutes for total battery; 5–10 minutes per subtests.
Cost. $100.00
Sensitivity/specificity. NES 2 CPT is particularly sensitive to methylmercury toxicity (Murata et al, 1999).

Test 23

Test name. Peabody Picture Vocabulary Test-III
2 parallel forms (Form IIIA and IIIB)
Target ages. 2.5– 90
Scoring. Standard scores ranging from 40 to 160. Hand scored while test is administered. Computerized scoring coming. Raw scores converted into normative scores.
Normative comparison groups available. Conormed with Expressive Vocabulary Test (Williams, 1997). Normed on 2725 individuals from 2 to 90 years of age, stratified to match the 1994 U.S. census. Age-based norm groups available. Hispanic norms available.
Languages. English, Spanish
Reliability. Alpha reliabilities for 25 standardized age groups ranged from .90–.98, with median .95. Split-half reliability for 25 age groups ranged from .83–.97, with median of .94. Test-retest reliability coefficients ranged from .77–.90. Internal consistency of .95 and temporal stability of .92 (Campbell, 1998).
Validity. Correlation with WISC-III VIQ of .91 for Form A and .92 for Form B and correlation of .82 for Form A and .80 for Form B with K-BIT Vocabulary Score. Full-scale IQ correlation with WISC-III was .90 for Forms A and B, and correlation with K-BIT composite was .78 for Form A and .76 for Form B. Correlation with WISC-III PIQ was .82 for Form A and .84 for Form B, and correlation with K-BIT matrices score was .65 for Form A and .62 for Form B.
Administration time. 11–12 minutes
Cost. $154.95 for complete kit
Parental SES/Cultural Bias. Lower scores for minority children in lower SES brackets on PPVT-R suggest limited utility in these groups (Washington and Craig, 1992). An earlier study found evidence of bias in relation to item difficulty when comparing Anglo-American and Mexican-American children, but found no clear pattern of item difficulty and that the difficulty was restricted to a relatively small number of items (Argulewicz and Abel, 1984).
Parental education. Higher scores on PPVT-R scores significantly associated with higher maternal education in study of children with very low birth weight (Ment et al., 2003).
Sensitivity. Evidence of utility as screening measure of intelligence and achievement.
Injury/illness. In very low birth weight (VLBW) children, those with early onset intraventricular hemorrhage and subsequent CNS injury had lowest initial PPVT-R scores, which declined rather

than improved over time, in contrast to VLBW children without complications (Ment et al., 2003).

Test 24

Test name. Purdue Pegboard Test
Target ages. 5–adult
Scoring. Score for pin placement = number of pins inserted in the time period for each hand. Score for bimanual condition = the total number of pins inserted with both hands.
Normative comparison groups available. Norms available from a sample of 1334 U.S. normal school aged children aged 5–15 (Gardner and Broman, 1979). Norms are also available for adults aged 15–40 years from a Canadian population of 225 largely right-handed adults with above average IQ (Yeudall et al., 1986). Norms are also available for children aged 2.5–6 using a revised pegboard (Wilson et al., 1982).
Languages. Language-free test
Reliability. Test-retest reliabilities = .63–.82. (Reddon et al., 1988). In subjects aged 25–33 practice effects were found when tested at 2- to 4-week intervals over eight test sessions (Feinstein et al., 1994). Right-left difference scores have correlations ranging from .22–.61 (Reddon et al., 1988).
Validity.
Administration time. Trial discontinued after 5 minutes
Cost. $109.00
Gender differences. In general, females do better than males (Agnew et al., 1988).
Sensitivity/specificity. Patients with Parkinson's disease, cerebellar disease, and Huntington's disease all show decreased execution (Brown et al., 1993). The Purdue Pegboard also appears to be sensitive to the presence of brain damage and it may be able to provide information about laterality of the injury (Costa et al., 1983).

Test 25

Test name. Raven's Standard Progressive Matrices
Target ages. 6–adult
Scoring. Raw scores of total correct items
Normative comparison groups available. Available for ages 6–68, with over 15 independent reference norms.
Languages. English only
Reliability. Internal consistency ranges from .60 to .98, with median of .90. Test-retest coefficients of .88 for ages 13 to 30.
Validity. Considered to come close to Spearman's g (general intellectual ability). Concurrent validity coefficients with Stanford-Binet and Wechsler range between .54 and .88, with majority in .70s and .80s. There are generally higher correlations with mathematics than with language subjects (Raven et al., 1998).
Administration time. 45 minutes
Cost. $150.00
Gender differences. No statistically significant differences observed in testing of Icelandic children (Pind et al., 2003).

Floor/ceiling effects. Ceiling effect noted in Icelandic school-age children, 6 to 16 years of age, indicating that utility as a measure of intelligence limited to children in first 7 grades (Pind et al., 2003).

Test 26

Test name. Revised Token Test
Target ages. 6–adult
Scoring. Scoring system assigns 1–15 points for responses to commands. Scoring system is complex, with score categories related to dimensions of accuracy, responsiveness, completeness, promptness, and efficiency. Formal training required for scoring proficiency. Although averaging scores is encouraged to obtain M subtest scores for each of 10 subtests, no clear method provided for interpretation of average scores.
Normative comparison groups available. Standardized on sample of 90 normal subjects, 30 right hemisphere damaged subjects, and 30 left hemisphere damaged subjects between 20 and 80 years of age
Languages. English only
Reliability. Test-retest reliability (2-week interval) of .90, but only five test subjects were used and no significance levels were reported. Intrascorer reliability in which three judges scored three videotaped sessions twice produced a subtest correlation of .99. Interscorer reliability based on three judges scoring the same videotaped administration produced an average subtest correlation of .98.
Validity. Examination of percentile scores in normative samples were able to differentiate between normal and left hemisphere subjects with 91 percent accuracy, between normal and right hemisphere subjects with 74 percent accuracy, and between right and left hemisphere subjects with 55 percent accuracy. Correlation of .63 with Northwestern Syntax Screening Test and .71 with Peabody Picture Vocabulary Test.
Administration time. 30–75 minutes
Cost. $175.00 for kit

Test 27

Test name. Rey Osterrieth Complex Figure
Target ages. 6–89
Scoring. Several scoring systems available
Normative comparison groups available. Children and adults
Languages. Language-free test
Reliability. Variety of reliability studies with different scoring systems. Meyers and Meyers scoring system indicated median interrater reliability of .94 and temporal reliability coefficients of .76 for Immediate Recall, .88 for Delayed Recall, and .87 for Recognition in sample of 12 subjects tested twice over average 184-day interval.
Validity. Variety of data from correlational and factor analytic studies are seen to support test validity as measure of visuoconstructional ability (copy test) and memory (recall and recognition test). Scores from normative and clinical samples able to discriminate among brain-injured subjects, psychiatric patients, and normal controls.
Administration time. 10–15 minutes
Cost. No cost

Gender differences. Some research has found that males perform better than females, but differences are small and no significant gender differences were found in a study of 6- to 11-year olds (Demsky et al., 2000).
Racial differences. No significant race differences found in study of 6- to 11-year olds (Demsky et al., 2000).
Sensitivity/specificity. Utility as diagnostic tool of cognitive impairment sensitive to mild head injury and array of central nervous system problems. May be useful in differentiating between different disorders and right-left hemisphere dysfunction (Spreen and Strauss, 1998; Baron, 2004).

Test 28

Test name. Stanford-Binet-5th Edition
Target ages. 2–85
Scoring. Raw scores converted to scaled scores. Scaled scores summed for non-verbal, verbal, and full-scale IQ, and for five factor index scores.
Normative comparison groups available. Sample of 4,800 subjects, 2–85+ years, stratified by age, sex, race/ethnicity, geographic region, and SES. Representative of 2000 U.S. census.
Languages. English only
Reliability. Reliability coefficients of .98 for Full Scale Score, .96 for Verbal Score, and .95 for Nonverbal Score. Five Factor Index scores all above .90.
Validity. Correlation of .90 with Stanford-Binet IV. High correlation between composite scores and those of major IQ batteries. Correlations from .78 to .84 with WPPSI-R, WISC-III, WAIS-III, and Woodcock-Johnson-III.
Administration time. 1 hour (time varies by age and functional level of examinee)
Cost. $1,544.50

Test 29

Test name. Stroop Color and Word Test
Target ages. 7+
Scoring. Raw Word Scores, Raw Color Scores, and Raw Color-Word Scores. Interference Score is derived from difference between Color-Word *T* Score and Color *T* Score.
Normative comparison groups available. Heterogeneous body of normative data. Normative data available for Spanish version administered to Mexican children from 6 to 12 years of age (Armengol, 2002).
Gender differences. Color word subtest indicates females have superior color-naming skills.
Languages. English, Spanish. Also available in Chinese, Czechoslovakian, German, Hebrew, Swedish and Japanese (Homack and Riccio, in press).
Reliability. .69–.89
Validity. Discriminant validity suggested in one study utilizing Stroop Test to discriminate between neurologically impaired subjects and normal controls (87 percent correct classification in combined results). Poor discrimination in second study seeking to describe location of brain injury (56.7 percent accuracy of classification). Factor analysis supports construct validity through loading on same factor of WAIS-R Block Design, Digit Symbol, Similarities, and Digit Span subtests and with serial subtraction subtask. Other validity evidence is embedded in diverse research studies and is frequently limited in nature.

Administration time. 5 minutes
Cost. $99.50 for kit
Sensitivity. Sensitive to discriminating between children with learning disabilities and control groups (Homack and Riccio, in press).

Test 30

Test name. The Token Test for Children
Target ages. 3–12.5 years
Applicability across age range of interest. 9.5 years.
Scoring. Test items scored as correct or incorrect, with one point assigned for correct response to test command and maximum score of 61. Raw scores are converted to standard scores. Standard scores are provided for five subtests and total score, with M of 500 and SD of 5.
Normative comparison groups available. Standardized on 1,304 native speakers of American English in northeastern Pennsylvania, 48 percent boys and 52 percent girls. No other normative data reported. Concern has been raised that there may be inadequate numbers in younger age groups of standardization sample.
Languages. English (verbal commands)
Reliability. Not available
Validity. Correlation of .71 with PPVY indicates validity as receptive language measure (Lass and Golden, 1975).
Administration time. 10 minutes
Cost. $125.00 for kit
Sensitivity/specificity. Sensitive measure of impairment in presence of increasingly complex grammar or verbal abstraction. Useful as screening assessment of auditory/verbal comprehension.

Test 32

Test name. Trail Making Test
Target ages. 9–14; 15+
Scoring. Time in seconds to complete two test parts (Part A and Part B). A difference score may be tabulated to remove speed element (caused by time required for examiner to notice error and examinee to comprehend and correct)
Normative comparison groups available. Norms available for adults and for 8–15 year olds (Spreen and Gaddes, 1969). Normative data also available in blocks for ages 15–17 and 18–23 (Fromm-Auch and Yeudall, 1983).
Languages. Language-free test
Reliability. Reliability coefficients in .80s and .90s. Alternate form reliability from .78 to .92. Part B may be less reliable in 9–14 year olds. Interrater reliability of .94 for Part A and .90 for Part B.
Validity. Construct validity for visual search
Administration time. 5–10 minutes
Cost. $90.00 for complete kit
Gender differences. In testing of high school athletes, adolescent girls outperformed boys on Trail Making Test Part B (Barr, 2003), while a second study found that boys took longer to

complete Part B than girls (Williams et al., 1995). However, other research has found no significant gender differences (Kennedy, 1981; Heaton et al., 1986).
Sensitivity. Sensitive to subdomains of inhibitory control and shift and sustain in children (Kelly, 2000).

Test 33

Test name. Wechsler Individual Achievement Test-II (WIAT - II)
Target ages. 4–85
Scoring. Provides composite scores for four areas of educational achievement: reading, mathematics, written language, and oral language. Raw scores converted to standard scores. Standard scores by age and grade, with M of 100 and SD of 15.
Normative comparison groups available. Normed on 5,586 individuals in 2 groups. a sample of school-aged children from pre-K to grade 12 (4–19 years, 11 months) and a sample of college students and adult, stratified on basis of gender, race-ethnicity, geographic region, and parental education to be representative of 1988 U.S. census. Conormed with WISC, WPPSI, and WAIS. Norms include fall, winter, and spring age- and grade-based standard scores, percentile ranks, stanines, and normal curve equivalents.
Languages. English and French-Canadian
Reliability. Internal consistency of subtests generally above .85, with Written Expression and Listening Comprehension subtests above .70 in school-aged sample, and Written Expression and Listening Comprehension subtests above .70 in college/adult sample. Internal consistency of composite scores above .90, except for Oral Language composite (above .85). Test-retest correlations (10-day interval) were above .85 and Composite Score test-retest scores were above .90 in school sample, and between .75–.85 in college/adult sample.
Validity. Strongly correlated with WIAT subtests (above .80). Correlations with WRAT-3 range from .68 to .77, with Differential Abilities Scales from .32 to .64.
Administration time. 45 minutes for pre-K to K, 90 minutes for grades 1–6, and 90–120 minutes for grades 7–16.
Cost. $375.00

Test 34

Test name. Wechsler Intelligence Scale for Children-IV
Target ages. 6–16.11
Scoring. Four Index scores and one Full-Scale score. Raw scores from sub-tests are converted into scales scores that are summed into four index scores.
Normative comparison groups available. Normative sample of 2200 children from 6 to 16.11, with samples from special groups. Sample stratified on basis of age, sex, race/ethnicity, parental education, and geographic region, consistent with 2000 U.S. census data.
Languages. English. Spanish, and Puerto Rican Spanish versions to be released in late 2004.
Reliability. Internal consistency reliability coefficients of subtests ranged from .79 to .90, with median of .86. Index scores range from .88 to .97, with median of .92.
Validity. Strong correlations between WISC-IV and equivalent metrics from WISC-III, WPPSI-III, WAIS-III, WASI, WIAT-II, and Children's Memory Scale. Matched samples of clinical and non-clinical children provided evidence of construct validity.
Administration time. 65–80 minutes

Cost. $950.00 for complete kit
Illness/Injury. Lower digit span scores WISC-R associated with elevated systolic blood pressure (Lande et al., 2003).

Test 35

Test name. Wechsler Preschool and Primary Scale of Intelligence-III
Target ages. 2:6–7.3
Applicability across age range of interest. 5 years
Scoring. Deviation IQ (M = 100, SD = 15) for Verbal, Performance and Full Scale IQs, and scaled scores for two supplemental subtests (M = 100, SD = 3).
Normative comparison groups available. Normative sample of 1,700 children in 9 age groups was representative of U.S. population on basis of sex, race/ethnicity, parental education, and geographic region in comparison with 1986 U.S. census.
Languages. English only
Reliability. Reliability coefficients for subtests range from .83 to .95 and for composite scales from .89 to .96. Test-retest reliabilities (M interval of 26 days) for 2:6–3:11 year olds were .90 for Verbal, .84 for Performance, .92 for Full Language, and .92 for General Language Scores; for 4–7:3 year olds, scores were .92 for Verbal, .87 for Performance, .93 for Processing Speed, .92 for Full Language, and .90 for General Language.
Validity. Validity information is generally derived from studies of the WPPSI-R, predecessor to the WPPSI-III. Concurrent validity is indicated by correlation of test scores with those from WISC-R, Stanford-Binet-IV, and the McCarthy Scales of Children's Abilities. Construct validity is supported by factor analysis, and studies of gifted, mentally impaired, learning disabled, and speech/language-impaired children have provided an indication of WPPSI-R discriminant validity.
Administration time. 30–50 minutes
Cost. $799.00

Test 36

Test name. Wechsler Adult Intelligence Scale-III**Target ages.** 16–89
Scoring. Verbal IQ, Performance IQ, Full-Scale IQ, with 7 subtests for each of the Verbal and Performance scales, though one of these (Object Assembly) is an optional subtest for the Performance scale. Four index scores (Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed) are generated from 11 of the subtests.
Normative comparison groups available. Normative sample consistent with contemporaneous census data and stratified for key variables. Oversampling was undertaken to facilitate research on educational level and cognitive abilities and to perform item bias analysis for African-American and Hispanic individuals. Norms expanded from WAIS-R based on contemporary sample and with extended upper range age limits.
Languages. English only
Reliability. Split-half reliabilities are generally excellent. Test-retest stability is good to excellent.
Validity. Criterion validity of .88 with Stanford-Binet, Fourth Edition and a range of .53–.81 with Wechsler Individual Achievement Test composites.
Administration time. 60–90 minutes

Cost. $799.00 for complete kit
Gender differences. In testing of high school athletes, adolescent girls found to perform better on WAIS-III digit symbol subtest, indicating advantage in information processing (Barr, 2003).

Test 37

Test name. Wechsler Memory Scale-III
Target ages. 16–89
Scoring. Verbatim recording of responses for subjective scoring on many subtests. Scoring software available.
Normative comparison groups available. Normed on sample of 1,250 individuals from 16-89 years, stratified to be representative of general population on basis of age, sex, race/ethnicity, educational level, and geographic region. Normative data based on 9- to 15-year-old Canadian children available for WMS-R logical memory and visual reproduction subtests (Paniak et al., 1998).
Languages. English only
Reliability. Internal consistency reliability coefficient for subtests average approximately .81 and for indexes approximately .87. Stability coefficients range from .62 to .82 across subtests over a 1-month interval.
Validity. Correlations with WMS-Revised. .72 between WMS-III Auditory Immediate and WMS-R Verbal Memory, .68 between WMS-III Auditory Delayed and WMS-R Verbal Memory, .65 between WMS-III General Memory and WMS-R Verbal Memory, .73 between WMS-III Auditory Immediate and WMS-R General Memory, .69 between WMS-III Auditory Delay and WMS-R General Memory, and .67 between WMS-III General Memory and WMS-R General Memory, .34
Administration time. 30–35 minutes (manual); 45–60 minutes (Mental Measurements Yearbook 14)
Cost. $499.00 for complete kit
Sensitivity. Evidence of sensitivity of WMS-R to memory disorders in a number of patient groups due to injury, illness, alcoholism, and toxic exposure (Spreen and Strauss, 1998).
Injury/illness. Lower scores on all composite indices of the WMS-R but the Attention/Concentration Index in patients with moderate to severe head injuries (Crossen and Viens, 1998).

Test 38

Test name. Wide Range Achievement Test 3
Target Ages. 5–75
Scoring. Raw scores, standard scores (M of 100 and SD of 15), absolute scores, and grade scores (grade scores are mean performance for particular grade level).
Normative comparison groups available. Standardized on sample of 4,473 subjects from 5 to 75 years of age, stratified by gender, ethnicity, geographic region, and socioeconomic status. Norms broken down by age.
Languages. English only
Reliability. Median test coefficients for nine subtests range from .82 to .95. Alternate form correlations of .92 for reading .93 for spelling, and .89 for arithmetic subtests.

Validity. High correlation with WRAT-R (.79–.92) and reasonable correlation with WAIS-R (.66–.73).
Administrative time. 15–30 minutes
Cost. $150.00 for kit
Prenatal exposures. Deficits on WRAT-R arithmetic and reading scores associated with prepregnancy binge drinking (Streissguth et al., 1990). Deficits in WRAT-R reading and spelling scores at 10 years of age associated with exposure to daily marijuana use during the first trimester. Reading deficits associated with marijuana use and with binge drinking during second trimester (Goldschmidt et al., 2004).

Test 39

Test name. Wide-Range Assessment of Memory and Learning-2
Target ages. 5–90
Scoring. Standard scores, scaled scores, and percentiles
Normative comparison groups available. WRAML was normed and standardized on 2,363 children ranging from 5 to 16 years of age, approximately 78 percent White, 12 percent Black, 7 percent Hispanic, and 3 percent other. The sample was representative of national population statistics based on the 1980 U.S. census and the 1988 Rand McNally Commercial Atlas and Marketing Guide.
Languages. English
Reliability. Alpha reliabilities:

- .93 (General Index)
- .92 (Core Battery Verbal Memory Index)
- .89 (Visual Memory Index)
- .86 (Attention-Concentration Index)

Validity. The WRAML has moderate correlations with WISC-R, indicating that it is not simply acting as a measure of general cognitive ability.
Administration time. <60 minutes
Cost. $519.00 for introductory kit
Prenatal exposures. Children with fetal alcohol syndrome or fetal alcohol effects showed impaired performance on WRAML word list learning and recalled significantly less information after delay relative to controls, although the exposed children were able to retain an equivalent proportion of the visual and verbal information relative to controls (Kaemingk et al., 2003).

Test 40

Test name. Wisconsin Card Sorting Test
Target ages. 6.5–89:11
Scoring. Scoring rules are complicated and scoring errors are reported to be common. Performance measured in several ways, taking into account number of categories completed, number of trials to complete the first category, percentage of "perseverative" errors in relation to overall test performance, errors made after completion of series of correct responses ("failure to maintain set"), percent of consecutive correct responses in series of three or more, and average change in conceptual efficiency in successive categories. Raw scores for 9 (of 13) variables can be converted to T scores, standard scores, and percentile ranks. Remaining variables can be converted to percentile ranks.

Normative comparison groups available. Normed on 899 individuals drawn from six distinct samples. Reliability and validity of normative data reported to be unclear. Majority of subjects from Southeast and Southwest regions of the United States, with race data provided for only one sample. Younger adults underrepresented in sample, compared to 1995 U.S. census. Normative data based on 5- to 12-year old Columbian children available for original WCST (Rosselli and Ardila, 1993).
Languages. English only
Reliability. Interscorer reliability coefficients for sample of children and adolescents ranged from .90 to 1.0 and for adults from .88 to .93. Intrascorer coefficients ranged from .83 to 1.0 for children and from .91 to .96 for adults.
Validity. Evidence indicates that WCST performance is valid indicator of executive ability, with poor performance observed in subjects with schizophrenia, Parkinson's disease, and Attention Deficit Disorder.
Administration time. 20–30 minutes
Cost. $305.00 for kit
Sensitivity/specificity. Original WCST seen to be sensitive to detecting frontal lesions in adults (Pendleton and Heaton, 1982), but ability to do so in children is less clear (Chase-Carmichael et al., 1999; Romine et al., 2004). Limited ability to differentiate between different clinical conditions (Romine et al., 2004)
Floor/ceiling. Performance found to be equivalent to that of adults by 10 (Chelune and Baer, 1986) or 11-12 (Rosselli and Ardila, 1993) years of age.

Test 41

Test name. Woodcock-Johnson-III (Test of Achievement)
Target ages. 2–90+
Scoring. Raw scores totaled and converted into age and grade equivalents, percentile ranks, and equivalency scores.
Normative comparison groups available. Normative samples from preschool to adult, selected to be representative of U.S. population, on basis of geographic distribution, community size and type, SES, sex, and race. School norm group included public, private, and home-schooled students.
Languages. English. Spanish version scheduled for late September, 2004 release.
Reliability. Median reliability for standard battery subtests range from .81 to .94 and for extended battery from .76 to .91. One-day test-retest reliabilities of speeded tests range from .69 to .96.
Validity. Confirmatory factor analyses provide validity evidence of CHC cognitive model relative to other batteries.
Administration time. 60–70 minutes
Cost. $466.50 for each of two forms (Form A and Form B)

References

Agnew J, Bolla-Wilson K, Kawas CH, Bleeker ML. Purdue pegboard age and sex norms for people 40 years old and older. *Developmental Neuropsychology* 1988; 4: 29–35.

Argulewicz EN and Abel RR. Internal evidence of bias in the PPVT-R for Anglo-American and Mexican-American children. *Journal of School Psychology* 1984; 22: 299–303.

Armengol CC. Stroop test in Spanish: Children's norms. *The Clinical Neuropsychologist* 2002; 16: 67–80.

Aylward EH, Schmidt S. An examination of three tests of visuomotor integration. *Journal of Learning Disabilities* 1986; 19: 328–330.

Azrin RL, Mercury MG, Millsaps C, Goldstein D, Trejo T. Pliskin NH. Cautionary note on the Boston Naming Test: Cultural considerations. *Archives of Clinical Neuropsychology* 1996 11: 365–366 (abstract).

Barnes GW and Lucas GJ. Cerebral dysfunction vs. psychogenesis in Halstead-Reitan tests. *The Journal of Nervous and Mental Disease* 1974; 158: 50–60.

Baron IS. *Neuropsychological evaluation of the child*. 2004. New York: Oxford University Press.

Barr WB. Neuropsychological testing of high school athletes: Preliminary norms and test-retest indices. *Archives of Clinical Neuropsychology* 2003; 18: 91–101.

Bornstein RA. Normative data on intermanual differences on three tests of motor performance. *Journal of Clinical and Experimental Neuropsychology* 1986; 8: 12–20.

Brown RG, Jahanshahai M, Marsden DC. The execution of bimanual movements in patients with Parkinson's, Huntington's, and cerebellar disease. *Journal of Neurology, Neurosurgery, and Psychiatry* 1993; 56: 295–297.

Campbell JM. Review of the Peabody Picture Vocabulary Test, Third Edition. *Journal of Psychoeducational Assessment* 1998; 16: 334–338.

Chase-Carmichael CA, Ris MD, Weber AM, Schefft BK. The neurological validity of the Wisconsin Card Sorting Test with a pediatric population. *The Clinical Neuropsychologist* 1999; 13: 405–413.

Cohen M, Town P, Buff A. Neurodevelopmental differences in confrontational naming in children. *Developmental Neuropsychology* 1988; 4: 75–81.

Conners CK. *Conners' Continuous Performance Test* (2nd ed.). 1995. Multi-Health Systems, Canada.

Chelune GJ and Baer RA. Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology* 1986; 8: 219–228.

Cooper ME, Rosen WG. Utility of the Boston Naming Test as a screen for language disorders in children. *Archives of Clinical Neuropsychology* 1977; 12: 303 (abstract).

Costa LD, Scarola LM, Rapin I. Purdue pegboard scores for normal grammar school children. *Perceptual and Motor Skills* 1983; 18: 748.

Crossen JR and Wiens AN. Residual neuropsychological deficits following head injury on the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist* 1988; 2: 393–399.

Demsky Y, Carone DAJ, Burns WJ, Sellers A. Assesment of visual-motor coordination in 6- to 11-year-olds. *Perceptual and Motor Skills* 2000; 91: 311–321.

Dodrill, CB and Troupin, AS. Effects of repeated administrations of a comprehensive neuropsychological battery among chronic epileptics. *Journal of Nervous and Mental Disease* 1975; 161: 185–190.

Feinstein A, Brown R, Ron M. Effects of practice of serial tests of attention in healthy subjects. *Journal of Clinical and Experimental Neuropsychology* 1994; 16: 436–447.

Finlayson MA and Reitan RM. Handedness in relation to measures of motor and tactile-perceptual function in normal children. *Perceptual and Motor Skills*; 43: 475–481.

Fromm-Auch D and Yeudall LT. Normative data for the Halstead Reitan neuropsychological tests. *Journal of Clinical Neuropsychology* 1983; 5: 221–238.

Gardner RA, Broman M. The Purdue pegboard: normative data on 1334 school children. *Journal of Clinical Child Psychology* 1979; 8: 156–162.

Gioia GA, Isquith PK, Kenworthy L. The Boston Naming Test with children: is it time to re-engineer? *Journal of the International Neuropsychological Society* 1996; 2: 29 (abstract).

Goldschmidt L, Richardson GA, Cornelius MD, Day NL. Prenatal marijuana and alcohol exposure and academic achievement at age 10. *Neurotoxicology and Teratology* 2004; 26: 521–532.

Haaland KY and Delaney HD. Motor deficits after left or right hemisphere damage due to stroke or tumor. *Neuropsychologia* 1981; 19: 17–27.

Heaton RK, Grant I, Matthews CG. Differences in neuropsychological test performance associated with age, education, and sex. In Grant I and KM Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders*. 1986. New York: Oxford University Press.

Homack S and Riccio CA. A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Archives of Clinical Neuropsychology*. In Press.

Johnstone B and Wilhelm KL. The construct validity of the Hooper Visual Organization Test. *Assessment* 1997; 4: 243–48.

Kaemingk KL, Mulvaney S, Halverson PT. Learning following prenatal alcohol exposure: Performance on verbal and visual tasks. *Archives of Clinical Neuropsychology* 2003; 18: 33–47.

Kelly T. The development of executive function in school-aged children. *Clinical Neuropsychological Assessment* 2000; 1: 38–55.

Kennedy KJ. Age effects on Trail Making Test performance. *Perceptual and Motor Skills* 1981; 52: 671–75.

Knights RM and Moule AD. Normative data on the Motor Steadiness Battery for Children. *Perceptual and Motor Skills* 1968; 26: 643–50.

Knights RM, Ivan LP, Ventureyra ECG, Bentivoglio C, Stoddart C, Winogron W, Bawden HN. The effects of head injury in children on neuropsychological and behavioral functioning. *Brain Injury* 1991; 5: 339–351.

Kramer J, Delis D, Kaplay E, O'Donnell L, Prifitera A. Developmental sex differences in verbal learning. *Neuropsychology* 1997; 11: 577–584.

Lande MB, Kaczorowski JM, Auinger P, Schwartz GJ, Weitzman M. Elevated blood pressure and decreased cognitive function among school-age children and adolescents in the United States. *Journal of Pediatrics* 2003; 143: 720–724.

Lass N and Golden S. A correlational study of children's performance on three tests for receptive language abilities. *Journal of Auditory Research* 1975; 15: 177–182.

Lee TM, Cheung CCY, Chan JKP, Chan CCH. Trail making across languages. *Journal of Clinical and Experimental Neuropsychology* 2000; 22: 772–778.

Massman PJ and Doody RS. Hemispheric asymmetry in Alzheimer's disease is apparent in motor functioning. *Journal of Clinical and Experimental Neuropsychology* 1996; 18: 110–121.

Matthews CG and Haaland KY. The effect of symptom duration on cognitive and motor performance in Parkinsonism. *Neurology* 1979; 29: 951–956.

Ment LR, Vohr BR, Allan W, Katz K, Schneider K, Westerveld M, et al. Change in cognitive function over time in very low-birth weight infants. *Journal of the American Medical Association* 2003; 289: 705–711.

Miller EN, Selnes OA, McArthur JC, et al. Neuropsychological performance in HIV-1-infected homosexual men: The Multicenter AIDS Cohort Study (MACS). *Neurology* 1990; 40: 197–203.

Moe V and Smith L. The relation of prenatal substance exposure and infant recognition memory to later cognitive competence. *Infant Behavior & Development* 2003; 26: 87–99.

Morrison MW, Gregory RJ, Paul JJ. Reliability of the finger tapping test and a note of sex differences. *Perceptual and Motor Skills* 1979; 48: 139–142.

Murata K, Weihe P, Renzoni A, Debes F, Vasconcelos R, Zino F, Araki S, Jorgensen PJ, et al. Evoked potentials in children exposed to methylmercury from seafood. *Neurotoxicology and Teratology* 1999; 21: 343–348.

Paniak CE, Murphy D, Miller H, Lee M. Wechsler Memory Scale-Revised Logical Memory and Visual Reproduction norms for 9 to 15 year-olds. *Developmental Neuropsychology* 1998; 14: 555–562.

Peters M, Servos P, Day R. Marked sex differences on a fine motor skill task disappear when finger size is used as a covariate. *Journal of Applied Psychology* 1990; 75: 87–90.

Pind J, Gunnarsdottir EK, Johannnesson HS. Raven’s Standard Progressive Matrices: New school age norms and a study of the test’s validity. *Personality and Individual Differences* 2003; 34: 375–386.

Prigatano GP, Altman IM, O’Brien KP. Behavioral limitations that traumatic brain-injured patients tend to underestimate. *The Clinical Neuropsychologist* 1990; 4: 163–176.

Raven J, Raven JC, Court JH. *Manual for Raven’s progressive matrices and vocabulary scales: standard progressive matrices.* 1998. Oxford: Oxford Psychologists Press.

Reddon JR, Gill DM, Gauk Se, Maerz MD. Purdue pegboard: test-retest estimates. *Perceptual and Motor Skills* 1988; 66: 503–506.

Reitan RM, Wolfson D. Relationships between specific and general tests of cerebral functioning. *The Clinical Neuropsychologist* 1996; 10: 37–42.

Reitan RM, Wolfson D. Dissociation of motor impairment and higher-level brain deficits in strokes and cerebral neoplasms. *The Clinical Neuropychologist* 1994; 8: 193–208.

Robinson AL, Heaton RK, Lehman RA, Stilson DW. The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology* 1980; 48: 605–614.

Rollinson DC, Rathleve NK, Moss M, Killiany R, Sassower KC, Auerbach S, Fish SS. The effects of consecutive night shifts on neuropsychological performance in the emergency department: a pilot study. *Annals of Emergency Medicine* 2003; 41: 400–406.

Romine CB, Lee D, Wolfe ME, Homack S, George C, Riccio CA. Wisconsin Card Sorting Test with children: a meta-analytic study of sensitivity and specificity. *Archives of Clinical Neuropsychology* 2004; in press.

Rosselli M and Ardila A. Developmental norms for the Wisconsin Card Sorting Test in 5- to 12-year-old children. *The Clinical Neuropsychologist* 1993; 7: 145–154.

Ruff RM and Parker SB. Gender and age specific change in motor speed and eye-hand coordination in adults: Normative values for the finger tapping and grooved pegboard tests. *Perceptual and Motor Skills* 1993; 76: 1219–1230.

Searight HR, Dunn EJ, Grisso T, Margolis RB. The relation of the Halstead-Reitan neuropsychological battery to ratings of everyday functioning in a geriatric sample. *Neuropsychology* 1989; 3: 135–145.

Seidel WT. Applicability of the Hooper Visual Organization Test to pediatric populations: preliminary findings. *The Clinical Neuropsychologist* 1994; 8: 59–68.

Shimoyama I, Ninchoji T, Uemura K. The finger-tapping test: A quantitative analysis. *Archives of Neurology* 1990; 47: 681–684.

Spreen O and Gaddes WH. Developmental norms for 15 neuropsychological tests age 6 to 15. *Cortex* 1969; 5: 171–191.

Spreen O and Strauss E. *A Compendium of Neuropsychological Tests* (2nd ed.). 1998. New York: Oxford University Press.

Streissguth AP, Barr H, Sampson PD. Moderate prenatal alcohol exposure: effects on child IQ and learning problems at age 7½ years. *Alcoholism: Clinical Experimental Research* 1990; 14: 662–669.

Trites R. *Neuropsychological Test Manual*. 1977. Lafayette Instrument Company. Ottawa, Canada.

Washington JA and Craig HA. Performances of low-income, African-American preschool and kindergarten children on the Peabody Picture Vocabulary Test-Revised. *Language, Speech and Hearing Services in the Schools* 1992; 23: 329–333.

Williams J, Rickert V, Hogan J, Zolten AJ, Satz P, D'Elia L, Asarnow RF, et al. Children's Color Trails. *Archives of Clinical Neuropsychology* 1995; 10: 211–223.

Williams KT. *Expressive Vocabulary Test: Manual*. 1997. Circle Pines, MN: American Guidance Service.

Wilson BC, Iacoviello JM, Wilson JJ, Risucci D. Purdue Pegboard performance of normal preschool children. *Journal of Clinical Neuropsychology* 1982; 4(1): 19–26.

Yeates KO, Blumenstein E, Patterson CM, Delis D. Verbal learning and memory following pediatric closed-head injury. *Journal of the International Neuropsychological Society* 1995; 1: 78–97.

Yeudall LT, Fromm D, Reddon JR, Stefanyk WO. Normative data stratified by age and sex for 12 neuropsychological tests. *Journal of Clinical Psychology* 1986; 42: 918–946.

Table 1. Comprehensive List of Neuropsychological Tests by Developmental Stage

***Domain*: OMNIBUS TESTS OF INTELLIGENCE AND ABILITIES**					
Test	**Infancy 0–2.5 yrs**	**Preschool 2.5–5 yrs**	**Grade School 6–11 yrs**	**Adolescent 12–17 yrs**	**Adult 18 + yrs**
Bayley Scales of Infant Development–II	X	x			
Mullen Scales of Early Learning	X	x			
Tests of Nonverbal Intelligence-III			X		
McCarthy Scales of Children's Abilities		x	X		
Universal Nonverbal Intelligence Test			X	x	
California Test of Mental Maturity		x	X	x	X
Kaufman Assessment Battery for Children-2 (KABC-2)		x			
Kaufman Brief Intelligence Test-2 (KBIT-2)		x	X	x	X
Kaufman Adolescent and Adult Intelligence Test			x (11)	x	X
Leiter International Performance Scale-Revised		x	X	x	X
Reynolds Intellectual Assessment Scales		x	X	x	X
Slosson Full-Range Intelligence Test		x	X	x	X
Structure of Intellect Learning Abilities Test		x	X	x	X
Raven's Coloured Progressive Matrices		x	X	x	X
Raven's Standard Progressive Matrices		x	X	x	X
Cognitive Abilities Test, Form 5			X	x	X
Comprehensive Test of Verbal Intelligence			X	x	X
Woodcock-Johnson III, Tests of Cognitive Abilities			X	x	X
Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-III)		x	X		
Wechsler Intelligence Test for Children-IV (WISC-IV)		x	X	x	X
Wechsler Adult Intelligence Scale-III (WAIS-III)				x	X
Wechsler Abbreviated Scale of Intelligence (WASI)			X	x	X

Stanford-Binet Intelligence Scale-IV, V		x	X	x	X
Differential Ability Scales		x	X	x	
NEPSY		x	X	x (12)	

***Domain*: ACHIEVEMENT TESTS**

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
Woodcock-Johnson Achievement Test-III (W-J-III)			X	x	X
Kaufman Test of Education and Achievement-II (KTEA – II)		x (4.5 yrs)	X	x	X
Wechsler Individual Achievement Test-II (WIAT - II)		x (4 yrs)	X	x	X
Wide Range Achievement Test 3 (WRAT 3)		x (5 yrs)	X	x	X

***Domain:* ATTENTION/CONCENTRATION**

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
Conners' Rating Scales-Revised		x	X	x	
Conners' Continuous Performance Test			X	x	X
Auditory Continuous Performance Test				x	
Neurobehavioral Evaluation System (NES) Animal, Letter CPT		x	x	x	X
Test of Everyday Attention for Children			x	x	
Brief Test of Attention					X
Test of Everyday Attention					X
WPPSI-III Subtests		x	x		
WISC-IV Subtests			x	x	
WAIS-III Subtests				x	X

***Domain:* EXECUTIVE FUNCTION/WORKING MEMORY**

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
Children's Category Test		x (5)	x	x (12–16)	

Cognitive Abilities Test, Form 5		x	x	x	
NEPSY Subtests		x	x		
Ross Information Processing Assessment-Primary		x	x		
Porteus Mazes Test			x	x	
Controlled Oral Word Association			x	x	X
Delis-Kaplan Executive Function System			x	x	X
Tower of London-Drexel University			x	x	X
Trail-making Test (TMT)			x	x	X
Booklet Category Test-II				x	X
Ruff Figural Fluency Test				x	X
Wisconsin Card Sorting Test-R (WCST)			x	x	X
Ross Information Processing-2nd Edition				x	X
WPPSI-III Subtests		x	x		
WISC-IV Subtests			x	x	
WAIS-III Subtests				x	X
WASI-Similarities			x	x	X
Children's Color Trails Test			x	x	
Paced Auditory Serial Arithmetic Test (PASAT)				x	X

***Domain*: LANGUAGE AND VERBAL FUNCTION**

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
Child Language Ability Measures		x	x		
WPPSI-III, Vocabulary, VC Index		x	x		
NEPSY, Language Subtests		x	x		
Test for Auditory Comprehension of Language		x	x		
Test for Reception of Grammar		x	x		
Token Test for Children		x	x		
Test of Word Knowledge		x	x	x	

Test of Language Competence-Expanded		x	x	x	x
Boston Naming Test		x	x	x	X
Comprehensive Receptive and Expressive Vocabulary Test-II		x	x	x	X
Oral and Written Language Scales		x	x	x	X
Peabody Picture Vocabulary Test-III (PPVT-III)		x	x	x	X
Receptive One-Word Picture Vocabulary Test		x	x	x	X
Stanford-Binet Intelligence Scale-IV, V (Vocabulary)		x	x	x	X
Woodcock Language Proficiency-Revised		x	x	x	X
Woodcock-Munoz Language Survey		x	x	x	X
Multilingual Aphasia Examination-III		x	x	x	X
Luria Nebraska Neuropsychological Battery-Children's Version Scales: Receptive Speech, Expressive Speech			x		
WISC-IV, Vocabulary, VC Index			x	x	
Clinical Evaluation of Language Fundamentals-R (CELF-R)			x	X	X
Clinical Evaluation of Language Fundamentals-3, Screening Test (CELF-3)			x	X	X
Controlled Oral Word Association			x	X	X
Delis-Kaplan Word Fluency Test			x	X	X
Detroit Tests of Learning Aptitude			x	X	X
Fullerton Language Test for Adolescents			x	X	X
Test of Adolescent/Adult Word Finding			x	X	X
WASI Vocabulary			x	X	X
Luria Nebraska Neuropsychological Battery-Forms I and II, *Scales*: Receptive Speech, Expressive Speech				X	X
WAIS-III Vocabulary, VC Index				X	X
Expressive Vocabulary Test (EVT)	x (2)	x	x	X	X

Domain: VISUOSPATIAL

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
WPPSI-III Perceptual Organization Index, Matrix Reasoning and supplemental subtests		x			
Spatial Orientation Memory Test		x	x		
Kaufman Assessment Battery for Children - 2 (KABC - 2),Visuospatial *Subtests*		x	x		
Test of Visual-Perceptual Skills-Non-Motor		x	x		
NEPSY, Visuospatial *Subtests*		x	x		
Test of Visual-Motor Skills		x	x	X	X
Bender Gestalt Test, Bender Gestalt - II		x	x	X	X
Beery-Buktenica Developmental Test of Visual-Motor Integration, 5th ed. (VMI 5)		x	x	X	X
Hooper Visual Organization Test (HVOT)		x	x	X	X
Porteus Mazes Test		x	x	X	X
Rey-Osterrieth Complex Figure (ROCF)		x	x	X	X
Stanford-Binet Intelligence Test-IV Copying Test, Bead Memory Test		x	x	X	X
Luria-Nebraska Neuropsychological Battery, Children's Version, Visual Function Scale			x		
WISC-IV Perceptual Organization Index, Matrix Reasoning and supplemental subtests			x	X	
WASI Block design, Matrix reasoning			x	X	X
Luria-Nebraska Neuropsychological Battery, Forms I and II, Visual Function Scale				X	X
Ruff-Light Trail Learning Test				X	X
WAIS-III Perceptual Organization Index, Matrix				X	X

Reasoning and supplemental subtests					

Domain: LEARNING AND MEMORY

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
NEPSY Memory and Learning Subscales		x	x		
Wepman's Auditory Memory Battery		x	x		
Ross Information Processing Assessment-Primary		x	x		
Test of Visual Perceptual Skills Subtests: Visual Memory, Visual Sequential Memory		x	x		
Rivermead Behavioral Memory Test for Children		x	x		
California Verbal Learning Test, Children's Version (CVLT-C)		x	x	X	
Children's Memory Scale		x	x	X	
Bender Gestalt Test-II, Delayed Memory		x	x	X	X
Delayed Recognition Span Test		x	x	X	X
Rey-Osterrieth Complex Figure (ROCF)		x	x	X	X
Stanford-Binet Intelligence Test-IV Bead Memory, Copying Test Adaptation		x	x	X	X
Test of Memory and Learning		x	x	X	X
Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML-2)		x	x	X	X
Woodcock-Johnson-III, Subtests: Auditory Working Memory, Short-term Memory		x	x	X	X
Luria-Nebraska Neuropsychological Battery-Children's Revision, Memory Scale			x		
Children's Auditory Verbal Learning Test-II			x	X	
Benton Visual Retention Test			x	X	X
Continuous Visual Memory Test-R			x	X	X
DCS-A			x	X	X
Rey Auditory Verbal Learning Test			x	X	X
California Verbal Learning Test-II (CVLT – II)				X	X
Luria-Nebraska Neuropsychological Battery, Form I: Memory ; Form II: Memory, Intermediate Memory				X	X
Ross Information Processing Assessment-2nd Edition				X	X
Test of Memory Malingering				X	X
Wechsler Memory Scale (WMS) (WMS–R), (WMS–III)				X	X
Rivermead Behavioral Memory Test-II				X	X

Domain: MANUAL MOTOR/ SPEED DEXTERITY

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
WPPSI-III Coding		x	x		
WISC-IV Coding			x	X	
WAIS-III Coding				X	X
Finger Tapping-Standard			x	X	X
Finger Tapping-NES			x	X	X
Grooved Pegboard Test		x	x	X	X
Purdue Pegboard Test		x	x	X	X
Revised Purdue Pegboard Test		x	x (6)		
Santa Ana Formboard		x	x	X	X
Hand-Eye Coordination-NES			x	X	X
NEPSY-finger tapping		x (3–5)	x	x (12)	

Domain: NEUROTOXICOLOGY TEST BATTERIES

Test	Infancy 0–2.5 yrs	Preschool 2.5–5 yrs	Grade school 6–11 yrs	Adolescent 12–17 yrs	Adult 18 + yrs
Neurobehavioral Evaluation System (NES), (NES2), (NES3)			x	X	X
Behavioral Assessment and Research System (BARS)			x	X	X
Pediatric Environmental Neurobehavioral Test Battery (PENTB)	X	x	x	X	X
Adult Environmental Neurobehavioral Test Battery (AENTB)					X
Neurobehavioral Core Test Battery (NCTB)				x (16+)	X

Table 2. Selected Neuropsychological Tests Listed by Domain

Omnibus IQ and Ability Tests

	Age Range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Bayley Scales of Infant Development - II	1 - 42 months	MDI, PDI, Behavior rating scale(%)	1992 US census, 1700 infants	reliability coeffiecients = .75 - .93	concurrent = .49 - .79	25 to 30 minutes	$1,025 - full kit	PCBs, Lead, Methylmercury
Kaufman Brief Intelligence Test - 2	4 - 90 years	3 composite scores	US census, 2022 individuals	split-half = .74 - .98, test-retest - .80 - .97	concurrent validity studies indicate correlation of K-BIT scores with established school achievement and intelligence tests	15 - 30 minutes	$162 - full kit	
Kaufman Assessment Battery for Children - 2	2.5 - 12.5 years	raw scores, scaled scores	1980 US census, 1981 children	test-retest = .77 - .97, odd-even correlations .80s - .90s	predictive validity for standard school achievement 12 in future ranged from .21 - .70. Correlation with Stanford-Binet .74	35 - 75 minutes	$470.00 - full kit	PCBs and Lead
McCarthy Scales of Children's Abilities	2.5 - 8.5 years	6 scaled scores - verbal, perceptual-performance, quantitative, memory, motor, and a composite score	1970 US census, 1,032 children	split-half = .79 - .93, test-retest = .75 - .90	r2 - S-B IQ = .81, WPPSI IQ = .62 - .71	45 - 75 (manual), 60-90 (mental measurements yearbook)	$630 - full kit	PCBs, Lead, Methylmercury, Manganese
NEPSY	3 - 12 years	5 domain-specific subtest scores, standard scores	1995 US census, 1000 children	stability coefficients = .67 - .76, internal consistency = .69 -.91	WISC-III r2 = .20 - .59; WPPSI-R r2 = .24 -.60.	60 minutes -ages 3 - 4, 120 minutes - ages 5 - 12	$599.00 -full kit	Organic solvents

Omnibus IQ and Ability Tests

	Age Range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Wechsler Preschool and Primary Scale for Children - III	2.6 - 7.3 years	VIQ, PIQ, FSIQ, processing speed index	1986 US census, 1,700 children	reliability coeffiecients - .83 - .95, test-retest - .84 - .93	.70 - .86 r2 WPPSI - R; .69 - .89 r2 WISC - III; .61 - .80 r2 Bayley.	30 - 50 minutes	$799 - full kit	WPPSI-R: Lead
Raven's Standard Progressive Matrices	6 - adult	Percentile scores	North American Samples	internal consistency = .60 .98, test-retest = .88	concurrent = - .54 - .88	45 minutes	$150 - full kit	Lead
Stanford-Binet Intelligence Scale IV	2 - 85 years	4 composite scores (t scores), subtest scores t scores	1980 US census, 5013 subjects	test-retest = .56 - .91	correlations with WISC-R = .60 - .83	60 minutes	NA	PCBs, Lead, Methylmercury
Stanford-Binet Intelligence Scale V	2 - 85 +	5 composite scaled scores, subtest age scaled scores, FSIQ, VIQ, nonverbal IQ	2000 US census, 4800 individuals	internal consistency = .95 - .98	Manual not yet available	50 minutes	$1825.50 - full kit	
WAIS - III	16- 89 years	VIQ, PIQ, FSIQ, VCI, POI, PSI, WMI	1995 US census, 2450 adults	subtest test-retest = .70 - .93	criterion validity = .53 - .88	60 -90 minutes	$799.00 - full kit	WAIS, WAIS-R: many
WASI	6 - 89 years	VIQ, PIQ, FSIQ	1997 census, 2245 individuals	test-retest = .72 - .95	WISC-III & WAIS III r2 = .72 - .92	30 minutes	$799.00 - full kit	
WISC - IV	6 - 16.11 years	raw, scaled, percentile, composites	2000 US census, 2200 children	internal consistency - .79 - .97	Manual not yet available	65 - 80 minutes	$950 - full kit	WISC-R, WISCIII: PCBs, Lead, Methylmercury, Arsenic

Achievement Tests

	Age range	Outcome scores	Normative Comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Kaufman Test of Educational Achievement - II	4.5 - 25 (comprehensive form) 4.5 - 90 (Brief form)	raw, age and grade standard scores, percentile, normal curve equivalents, and stanines	1994 US census, updated in 1995-96 mirroring 1994.	not updated for new normative sample. Internal consistencies .77 - .88. test-retest .90 or higher.	correlation with other tests (WRAT, SB, KABC) .83 - .88	comprehensive - 25 to 70 minutes. Brief form 20 -30 minutes	$299.99 - full kit	
Wechsler Individual Achievement Test - II	4 – 85 years	raw, standard, age and grade equivalents	1988 US census, 5586 individuals	internal consistency =.70 - .90	WIAT - above .80. WRAT3 - .68 - .77.	45 - 120 minutes - dependent on age	$375.00 - full kit	
Wide Range Achievement Test - 3	5 – 75 years	raw, scaled, absolute, grade scores	4473 subjects, 5 - 75	.82 - .95	correlation with WRAT-R .79 - .92 and correlation with WAIS - R .66 - .73	15 - 30 minutes	$150.00 - full kit	Inorganic mercury, alcohol, marijuana
Woodcock-Johnson - III	2 – 90+ years	raw, age and grade equivalents, percentiles and equivalency scores	US population, 8818 indivduals	test-retest = .69 - .96. median reliability for standard battery = .81 - .94	.50 - .80 r2 WIAT.	60 - 70 minutes	$466.50 for each of two forms	Methylmercury

Attention and Concentration Tests

	Age range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Conners' Continuous Performance Test - II	4 - adult	omission erros, average reaction time, commission errors, standard error of reaction	2686 clinical and non- clinical subjects.	test-retest - .55 - .84		14 minutes	$595.00 - full kit	Lead
Conners' Rating Scale - Revised	3 - 17 years	raw, T scores	US &Canadian populations, 8000	internal reliability - .73 - .94 test- retest .47 - .92	r2 - conners cpt = .44; parent to teacher form = .28 - .50	20 minutes or less	$203.00 - full kit	Lead
Neurobehavioral Evaluation System - CPT	7+ years	reaction time, error scores, and standard deviation	limited	test-retest - .51 .92		50 minutes for entire battery (5 10 min per subtest)	$100.00	PCBs, Methylmercury

Executive Function Tests

	Age range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Children's Category Test	5 – 16 years	raw, standard, t scores	1988 US census, 920 individuals	internal consistency .81 - .91. test-retest .75	.14 - .27	15 - 20 minutes	$375.00 - full kit	Methylmercury
Trail-making Test	9 – 14 years, 15+ years	time in seconds, percentiles	2000 US census, 1664 individuals	Reliability coefficients in .80s and .90s. Alternate forms reliablility .78 - .92	construct validity = .36 - .93	5 - 10 minutes	$90.00 -full kit	
Children's Color Trails Test	8 – 16 years	time	678 children	alternate form reliability .85 - .90. test-retest reliability .90 - .99	81% variance with children with mild head injuries.	5 - 7 minutes	$135.00 - full kit	
Color Trails Test	18 + years	based upon time in seconds for completions of test parts. Also includes a qualitative scoring component based on # of errors, near misses, correction, and prompts	1,528 subjects 18 - 89, including African American, Hispanic subsamples	test-retest reliability .64 - .79	.41 - . 50 with trail-making test	3 - 9 minutes	$104.00 -full kit	

Executive Function Tests

	Age range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Wisconsin Card Sorting Test	6.5 - 89 years	scores for # of categories completed, number of trials to complete 1st category, percent of perseverations, failure to maintain set, percent of consecutive efficiency in successive categories, raw scores for 9 variables, T scores, standard scores, and percentile ranks	1995 US census, 899 individuals	interscorer reliability .83 - 1.0	validation studies done with children/adolescents with focal brain damage, ADHD &LDs, but no numbers given	20 - 30 minutes	$305.00 -full kit	PCBs, Lead, Solvents
Stroop Color and Word Test	7 + years	raw, T scores	Heterogeneous body of normative data	.69 - .89	.29 - .37 with WAIS-R; .31 with ROCF	5 minutes	$99.50 - full kit	

Language and Verbal Function Tests

	Age range	Outcome scores	Normative comparison groups	Published in Spanish	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Boston Naming Test	5 to adult	raw scores		Yes	Correlation btw test version .92	concurrent validity = .74 - .87	10 - 20 minutes	$219.00 - full kit	Methylmercury, PCBs, Lead, Solvents
Peabody Picture Vocabulary Test - III	2.5 to 90 years	raw, standard scores, percentiles, age equivalents, & stanines	1994 US census, 2725 individuals	Yes	Alpha reliability .90 - .98. split half reliability .83 - .97. test-retest reliabiltiy .77 - .90. internal consistency .95. temporal stability .92	correlation with WISC .82 - 92. correlation with KBIT .62 - .82.	11 - 12 minutes	$154.95 - full kit	Lead
Token Test for Children	3 - 12.5 years	raw, standard scores	304 native speakers of American English.	No	not available	correlation of .71 with PPVT	10 minutes	$125.00 - full kit	Lead
Revised Token Test	6 - Adult	raw scores, standard scores			test-retest .90. interscorer reliability .99	.71 correlation with PPVT.	30 - 75 minutes	$175.00 - full kit	
CELF 3	6 – 21 years 2 5	subtests age scaled scores, expressive and receptive scaled scores	US population, 2400 No individuals	Yes	internal consistency = .55 - .95. test-retest = .50 - .93.	.42 - .79 with CELF-R	30 - 45 minutes for entire test	$348.00 - full kit	
EVT	naming, 5 - 90 synonyms	scaled scores, percentiles, age equivalents	2725 individuals	No	test-retest = .77 - .90; split-half = .83 - .97; alphas = .90 - .98.	r2 with the PPVT 2 = .62 - .88	10 - 20 minutes	$159.99 - full kit	

Visuospatial Tests

	Age range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Beery-Buktenica Developmental Test of Visuomotor Integration - V	3 – 17 years	raw score, scaled score	1990 US census, 2,614 individuals	alpha reliability - .82. split half reliability .88. interrater reliability .94 - .98.	construct validity .55 - .66. concurrent validity .60 - .75	10 - 15 minutes	$99.50 - full kit	Lead
Bender Gestalt Test II	4 to adult	raw score, scaled score	2000 US census, 4,000 individuals	interscorer reliability .83 - .94. test-retest .80 - .88. internal consistency .91	.80 r2 Bender; .65 r2 VMI.	10 - 15 minutes	$130.00 - full kit	Bender 1 : Lead, Methylmercury
Hooper Visual Organization Test	5+	raw and t scores	500 children, 5 - 13. 207 children 5 - 11 (independent study)	split half .82.		10 - 15 minutes	$185.00 - full kit	

Learning and Memory Tests

	Age range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
California Verbal Learning Test, Children's version	5 – 16 years	raw scores, SD based scores	1988 US census, 920 children	split half .88, alpha .85, test-retest .31 - .90		30 minutes	$159.00 full kit	Lead, Methylmercury
California Verbal Learning Test, II	16 – 89 years	raw and scaled scores	1999 US census, 1087 adults	test-retest = .30 - .88; alternate forms = .72 - .79.	.63 - .86 r2 CVLT	30 minutes	$199.00 full kit	
Rey-Osterreith Complex Figure	6 – 89 years	raw scores & percentiles	children and adults	interrater reliability of .94; temporal reliability .76 - .88.	.49 with WISC-R performance scale	10 - 15 minutes	No cost	Lead
Wide Range Assessment of Memory and Learning 2	5 – 90 years	raw, standard, scaled, percentiles	1980 US census and 1988 Rand McNally Commercial Atlas and Marketing Guide, 2,363 children	Alpha - .86 - .93	internal validity = .98 -1.00; intercorrelations of indexes = .24 - .83.	less than 60 minutes	$519.00 full kit	Inorganic mercury
Wechsler Memory Scale - III	16 - 89 years	subtests: ASS, percentiles. Composite: standard scores	1250 individuals, 16 - 89	internal consistency .81 - .87. stability coefficients .62 - .82.	Correlations with WMS-R .34 - .73	60 minutes	$499.00 full kit	Inorganic mercury, Methylmercury

Manual Motor/Speed Dexterity Tests

	Age range	Outcome scores	Normative comparison groups	Reliability	Validity	Administration time	Cost	Toxicants studied (children)
Grooved Pegboard Test	5 - adult	time to completion	2,500 adults and children	test-retest .80 - .81; reliability coefficients for # of errors .20 - .21		5 minutes	$129.00	Methylmercury
Revised Purdue Pegboard	2.5 - 6 yrs	time to completion, peg placement raw score	206 RH children			5 minutes	specially made	
Purdue Pegboard Test	5 - adult	raw scores	1334 US children aged 5-15, 225 Canadians.	test-retest = .63 - .82.		trial discontinued after 5 minutes	$109.00	
Santa Ana Form Board	5 - adult	total correct				less than 5 minutes	~$200.00 - needs to be specially made	mercury
Finger Tapping Test (maual)	6 - adult	mean raw scores (5 trials/ condition)	120 right-handed children	test-retest = .58 - .93		5 minutes	$95.00	Lead
Finger Tapping Test (NEPSY)	3 - 12 years		1995 US census, 1000 children			5 minutes	$599.00 - full NEPSY kit	Organic solvents
Finger Tapping Test (NES)	7+ years	reaction time, error scores, and standard deviation	limited	test-retest - .51 .92		5-10 min	$100.00	PCBs, Methylmercury
Hand Eye Coordination (NES)	7+ years	reaction time and error scores and standard deviation	limited	test-retest - .51 .92		5-10 min	$100.00	PCBs, Methylmercury

Table 3. Exposure-Effect Results, Toxicants (Children)

Bayley Scales of Infant Development	
No statistically significant association	**Results indicate toxicant related impairment**
PCBs/DDE In cohort of infants no association between transplacental exposure to DDE and Bayley MDI scores at 12 months of age and no association between post-natal PCB or DDE exposure through breast feeding and Bayley PDI or MDI scores (Gladen et al., 1988). PCB concentrations in early maternal milk not associated at statistically significant levels with Bayley mental and motor scales in German infants at 7 and 18 months of age. Cord blood PCB had small and slightly positive association with Bayley mental score (Walkowiak et al., 2001).	*PCBs* Higher transplacental exposure to PCBs associated with lower Bayley PDI scores at 6 months and 12 months of age (Gladen et al, 1988). Lower scores on Bayley mental scale and motor scale in Taiwanese infants exposed in utero to PCBs from contaminated cooking oil relative to non-exposed infants in control group (Rogan et al., 1988). Increases in PCB levels in early maternal milk associated with lower Bayley mental and motor scores in German infants at 30 and 42 months. Cord blood PCB had inverse association with motor score (Walkowiak et al., 2001). Prenatal PCB exposure associated with slightly lower Bayley PDI scores in cohort of Dutch infants at 3 months of age. When corrected for confounders, PCB and dioxin exposures in highest-exposed breast-fed infants at 7 months had lower scores than formula-fed infants (Koopman-Esseboom et al., 1996).
Lead At 24 months, no statistically significant association between Bayley psychomotor index and antenatal, birth, and postnatal blood lead levels. An inverse association between blood lead levels and Bayley Mental (Wigg et al. , 1988). No association between children's postnatal blood lead levels and Bayley-II (Spanish version) MDI and PDI scores at 12 and 24 months (Gomaa et al., 2002). In cohort of infants, no significant association between cord blood lead levels and the PDI score (Bellinger et al., 1984). No statistically significant association between Bayley PDI scores at 3, 6, and 12 months and cord PbB levels after adjustment for confounders in cohort of Chinese infants (Shen et al., 1998). No statistically significant association between prenatal and neonatal PbB levels and scores on Bayley PDI at 3 and 6 months of age (Dietrich et al., 1987) No statistically significant relationship between Bayley MDI or PDI scores and prenatal exposures to lead in cohort of Australian children at 6, 12, and 24 months (Cooney et al., 1989).	*Lead* At 24 months, a statistically significant inverse association between antenatal and postnatal PbB levels and Bayley MDI, with elevation in PbB at 6 months having greatest negative effect (Wigg et al., 1988). Infants in with highest umbilical cord PbB levels scored lower on Bayley MDI at 6, 12, 18 and 24 months of age than infants in lower exposure groups (Bellinger et al., 1987). At low lead levels, scores on Bayley MDI inversely associated with prenatal and umbilical cord PbB levels in infants at 3 months of age. An inverse association also noted at 6 months of age, but just outside statistical significance using 2-tail test. Score decrements partially mediated by lead-related reductions in birth weight and gestation (Dietrich et al., 1987). Increases in blood lead concentrations from 10 to 30 micrograms/dl in infants at 24 months associated with estimated 2.5 decrement in Bayley MDI. Decrements at birth, 6, 12, and 18 months also recorded but not at conventional levels of statistical significance (Wasserman et al., 1992). Inverse association between maternal blood lead concentration and Bayley scores at 6 months of age (Ernhart et al., 1987). After adjustment for confounders, maternal bone lead levels inversely associated with Bayley-II (Spanish version) MDI scores in Mexican infants at 24 months of age (Gomaa et al., 2002). At 3, 6, and 12 months of age Bayley MDI scores were inversely associated with cord PbB levels after adjustment for confounders in cohort of Chinese infants (Shen et al., 1998).
Methylmercury No association found between prenatal MeHg exposure and testing at age 19 and 29 months on the PDI index of the Bayley (Davidson et al., 1999). No association found between prenatal MeHg exposure and testing at age 29 months on the MDI index of the Bayley (Davidson et al., 1999). At 19 and 29 months of age, no association between fetal MeHg exposure and the MDI or PDI of the Bayley was found (Myers et al., 1995)	*Methylmercury* Association found between prenatal exposure to MeHg and testing at age 19 months on the MDI index of the Bayley (Davidson et al., 1999).

Beery Developmental Tests of Visual-Motor Integration	
Lead No association between current blood lead levels or blood lead levels at birth, and at 6, 12, 18, 24, 57 months and VMI scores in 10-year old children (Stiles and Bellinger, 1993). *Methylmercury* No association with VMI scores in Seychelles Island children at 9 years of age due to prenatal methylmercury exposure (Myers et al, 2003). *PCBs* No association with VMI scores in the children tested in Michigan who were exposed in-utero to PCBs (Jacobson et al., 1990).	*Lead* At 7 years of age, children in Port Pirie, Australia cohort showed an inverse relation between blood lead concentration and score on Beery Test of Visual-Motor Integration (Baghurst et al., 1995).
Bender Gestalt Tests	
PCB No association with copying scores in study of 7-year-old Faroese Island children prenatally exposed to PCB (Grandjean et al., 2001). *Methylmercury* No association with copying scores in study of 7-year-old Faroese Island children prenatally exposed to methylmercury (Grandjean et al., 1997).	*Methylmercury* Inverse association with recall scores in study of 7-year-old Faroese children prenatally exposed to methylmercury (Grandjean et al., 1997). *Lead* A multi-country study in Europe found a significant inverse association between blood lead levels and performance on "GFT" version Bender Gestalt test (German scoring system that uses raw error rather than age-adjusted standard scores and requires reproductions to be drawn on wavy lined paper) in schoolchildren between 6 and 11 years of age (Winneke et al., 1990). Impaired function on Bender Visual Motor Gestalt Test in high lead exposure group (measured by lead levels in shed teeth) of Danish first grade schoolchildren compared to low lead group (Hansen et al., 1989).
Boston Naming Test	
Lead No association found between the BNT scores of young adults and high lead content of teeth shed at the ages of 6 and 7 (Needleman et al., 1988). *Methylmercury* No association with BNT scores in Seychelles Island children at 9 years of age due to prenatal methylmercury exposure (Myers et al, 2003).	*Methylmercury* Inverse association with BNT scores in 7-year-old Faroese Island children prenatally exposed to methylmercury (Grandjean et al., 1997). Case report of 19-year-old exposed to mercury between 4 to 9 years finds BNT confrontation naming of pictured objects below average (Diamond et al., 1995).
California Verbal Learning Test (Children)	
Methylmercury No association with CVLT learning scores or recognition scores in 7-year-old Faroe Island children prenatally exposed to methylmercury (Grandjean et al., 1997). *Lead* No association between current blood lead levels or blood lead levels at birth, and at 6, 12, 18, 24, 57 months and CVLT-C List A, List B, Entire Test, Free Recall, Cued Recall, and Intrusions scores in 10-year-old children (Stiles and Bellinger, 1993). No association found between the CVLT scores of young adults and high lead content of teeth shed at the ages of six and seven (Needleman et al., 1988).	*Methylmercury* Methylmercury levels inversely associated with CVLT short-term recall and long-term recall scores in 7-year-old Faroe Island children prenatally exposed to methylmercury (Grandjean et al., 1997). *Lead* Inverse association between CVLT-C Perseverative scores and blood lead levels at 24 months in 10-year-old children (Stiles and Bellinger, 1993).

Finger Tapping Tests	
	Lead Inferior performance by adults with history of childhood lead exposure on finger tapping test relative to adults with no such exposure history (White et al., 1993).
Grooved Pegboard	
	Methylmercury Increased prenatal exposure to MeHg was associated with decreased performance on the Grooved Pegboard Test, using the nondominant hand (Myers et al., 2003).
Kaufman Assessment Batteries for Children	
Lead Following adjustment for home environment and maternal intelligence, KABC scores not associated at statistically significant levels with prenatal, neonatal, and postnatal blood lead levels in Cincinnati lead cohort at 5 years of age (Dietrich et al., 1992). At 4 years of age, no statistically significant association in Cincinnati lead cohort between: KABC subscale scores and prenatal lead exposures (maternal PbB); postnatal PbB levels and KABC subscale scores after full covariate adjustment; between KABC mental processing composite score and PbB levels at 1 year of age, 2 years of age, or lifetime average PbB; between KABC simultaneous processing score and PbB levels as 1 or 2 years of age; or between KABC sequential processing, nonverbal, and achievement scores and yearly or lifetime PbB levels through age 4 (Dietrich et al., 1991).	*PCBs* Increasing PCB concentration in maternal plasma associated with lower scores on the sequential processing scale, the simultaneous processing scale, and the overall cognitive scale on the Dutch version of the KABC in Dutch children at 42 months of age (Patandin et al., 1999). Increasing PCB concentrations in early maternal milk associated with lower KABC scores in German children at 42 months of age (Walkowiak et al., 2001). *Lead* Inverse association between all tests in KABC battery and neonatal PbB levels in Cincinnati cohort children from lower income families at 4 years of age. Statistically significant inverse associations between KABC mental processing composite score and PbB levels at ages 3 and 4 and between simultaneous processing score and PbB levels at ages 3 and 4 and lifetime PbB average (Dietrich et al., 1991).
McCarthy Scales of Children's Abilities	
PCBs No association between prenatal PCB exposures and MSCA GCI scores in children born to older mothers or to parents with higher verbal IQ (Vreugdenbil et al., 2002). *PCB/DDE* No statistically significant association between perinatal exposure to PCBs and MSCA scores at 3, 4, or 5 years of age in cohort of children followed from birth (Gladen and Rogan, 1991).	*PCBs* Verbal Scale and Memory Scale scores lower in 4-year-olds exposed to PCBs in utero. Strongest association with Verbal Memory, Numerical Memory, Forward Digit Span, and Backward Digit Span subtests. Also an inverse association with the Quantitative Scale fell just short of statistical significance (Jacobson, Jacobson, and Humphrey, 1990). Small but significant decline in scores on Perceptual Performance Scale, Quantitative Scale, and General Cognitive Index in children at 38 months of age with prenatal PCB exposures (Stewart et al., 2003). Higher prenatal exposure to PCBs in cohort of Dutch children associated at 6.5 years of age with lower McCarthy (Dutch version) GCI scores in children born to younger mothers and to parents with lower verbal IQ scores. Prenatal PCB exposure also associated with lower memory and motor scale scores in children born to younger mothers and to parents with lower verbal IQ scores (Vreugdenbil et al., 2002).
Lead Maternal hair methylmercury levels not associated with McCarthy memory and motor subscale scores at 66 months of age, and GCI and perceptual performance subscale scores lose statistically significant association after removal of outliers (Myers et al., 1995). No statistically significant association between McCarthy GCI scores	*Lead* At 4 years of age, scores on MSCA GCI and on Perceptual-Performance and Memory subscales were inversely associated with blood lead concentrations at 6, 15, 24, 36, and 48 months (McMichael et al., 1988). Postnatal blood lead levels associated with lower scores on Verbal,

and prenatal exposure to lead in cohort of Australian children at 36 months of age (Cooney et al., 1989). *Manganese* No significant relationships observed between Mn levels at birth and McCarthy GCI at 3 and 6 years. No significant relationships after adjustment for child's gender and mother's education in recombined McCarthy Attention, Non-Verbal Memory, and Hand Skill scores at 6 years of age or in Hand Skill scores for girls at 3 years (Takser et al., 2003. *Methylmercury* No association between MeHg exposure and performance on the General Cognitive Index was identified in a cohort of 66-month-old children (Palumbo et al., 1999).	Quantitative, Perceptual-Performance, Memory, and Motor subscales (strongest association with Perceptual Performance) in children from 2 towns in Yugoslavia (Wasserman et al., 1994). *Manganese* Mn levels in cord blood inversely associated after adjustment for child's gender and mother's education level with 3 recombined MSCA scores at 3 years of age: Attention (number questions, tapping sequence, numerical memory I), Nonverbal Memory (pictorial memory, tapping sequence), and hand skills (sum of principal component z-scores of coordination of arms, copying of geometric patterns, and drawing of a man subtests), though the hand skills variable had a negative association only in boys (Takser et al., 2003). *Methylmercury* Scores on McCarthy GCI and perceptual performance subscale were negatively associated with levels of prenatal exposure to methylmercury (maternal hair sample) in cohort of children from the Seychelles at 66 months of age (Myers et al., 1995).
NEPSY	
	Organic solvents Deficits in NEPSY Design Copy and Visuo-motor Precision tasks in 3 to 7-year-old children prenatally exposed to organic solvents. Lower composite scores in receptive language (which included NEPSY Phonological Processing and Comprehension of Instructions subtests) and expressive language (which included NEPSY Body Part Naming, Speeded Naming and Verbal Fluency subtests) were also associated at more moderate levels with maternal solvent exposure (Till et al., 2001).
Neurobehavioral Evaluation System (NES)	
PCB No association found NES2 finger tapping and hand-eye coordination in study of 7-year-old Faroese Island children prenatally exposed to PCB (Grandjean et al., 2001). *Lead* No association found between certain NES subtests scores (symbol-digit substitution, pattern memory, pattern comparison, serial digit learning, and switching attention) of young adults and high lead content of teeth shed at the ages of six and seven (Needleman et al., 1988). *Methylmercury* No association with NES2 Hand-Eye Coordination and Tactual Performance Test scores in study of 7-year-old Faroese Island children prenatally exposed to methylmercury (Grandjean et al., 1997). No association was seen between the NES2 finger tapping and hand-eye coordination subtests and maternal MeHg hair concentrations when studying the children while in the first grade (Murata et al., 1999).	*PCBs* Poorer performance on NES2 Continuous Performance Test associated with increased wet-weight PCB concentrations in 7-year-old Faroese children prenatally exposed to neurotoxicants in seafood (Grandjean et al., 2001). *Methylmercury* Negative association with NES2 Finger Tapping and Continuous Performance Test scores in study of 7-year-old Faroese Island children prenatally exposed to methylmercury (Grandjean et al., 1997). A weak negative association with NES2 CPT scores and maternal hair concentrations during pregnancy found when the children were tested in the first grade (Murata et al., 1999). Slight but significant decrements on NES with increased prenatal exposure to methylmercury in 7-year-old Faroese Island children (Dahl et al., 1996).
Raven's Standard Progressive Matrices	
	Lead Inferior performance by adults with history of childhood lead exposure relative to those with no such history (White et al., 1993).
Rey Osterrieth Complex Figure Test	
Lead No association between blood lead levels at 57 months and ROCF Delayed Recall and Immediate Recall scores in 10-.year-old children (Stiles and Bellinger, 1993).	*Lead* Inverse association between Copy scores and blood lead levels at 57 months in 10-year-old children (Stiles and Bellinger, 1993).

No association found between the Rey - O scores of young adults and high lead content of teeth shed at the ages of six and seven (Needleman et al., 1988)	
Stanford-Binet Intelligence Scale – IV	
Lead No statistically significant association found at 3 years of age between Stanford-Binet (4th edition) Composite Score and lifetime average blood lead level (measured at 6, 12, 18, 24, and 36 months) or blood lead level in infancy (6–24 months) and no association at 5 years between Composite score and average in infancy (Canfield et al., 2003). *Methylmercury* No association seen on Bead Memory subtest in Brazilian children (Grandjean et al., 1999).	*Lead* At 3 years of age, Stanford-Binet Intelligence Scale (4th edition) Composite Scores were inversely associated with peak (from levels measured at 6,12, 18, 24, and 36 months) and concurrent blood lead concentration. At 5 years of age, Composite Score inversely associated with lifetime average blood lead concentration (average of 6, 12, 18, 24, 26, 48, and 60 month levels), peak blood lead concentration, and concurrent blood lead concentration (Canfield et al., 2003). *PCBs* Preschool children neonatally exposed to PCBs scored 4 points lower on Stanford-Binet Intelligence Scale than unexposed controls (Rogan et al., 1988). *Methylmercury* An association seen in Brazilian children between 7 and 12 years of age currently exposed to MeHg on the copying subtest of the Stanford-Binet (Grandjean et al., 1999). A weak association was found between maternal MeHg hair levels during pregnancy and achievement on the bead memory subtest among Madeiran children (Murata et al., 1999). An association between the S-B copying test and prenatal MeHg exposure was found in children in French Guiana (Cordier et al., 2002).
Stroop Color-Word Test	
PCBs No effect seen between prenatal PCB exposure and performance on the Stroop Color-Word Test (Jacobson & Jacobson, 2003).	
The Token Test	
Lead No statistically significant differences on Token Test Block 1, Block 2, Block 3, or summary scores between children with high lead levels and those with low lead levels (Needleman et al., 1979).	*Lead* Children with high lead levels performed less well on Token Test Block 4 than children with low lead levels (Needleman et al., 1979).
Trail Making Test	
Lead No significant effects were found on the Trail Making Test in children with blood lead levels ranging from below 5 to about 60 micrograms/100 ml of blood (Winneke et al., 1990). No significant association between dentin lead level and Trail Making A and B in cohort of Danish 1st grade school children (Hansen et al., 1989).	
Wechsler Adult Intelligence Scale (WAIS)	
Lead No difference in WAIS-R digit span subtest between adults with history of childhood lead exposure relative to non-exposed adults (White et al., 1993).	*Lead* Inferior performance on WAIS-R similarities, vocabulary, picture completion, and block design subtests by adults with history of childhood lead exposure relative to non-exposed adults (White et al., 1993).

	Mercury Problems in WAIS-R forward span, block design, picture arrangement, object assembly, and arithmetic subtests and slow performance on picture completion and digit symbol subtests identified in case report of 19-year-old exposed to mercury in home between 4 and 9 years of age (Diamond et al., 1995).
Wechsler Memory Scales	
Methylmercury No difference in WMS information subtest between adults with history of childhood lead exposure relative to non-exposed adults (White et al., 1993). WMS-R delayed recall, learning of verbal associates, and visual paired associates within normal limits in case report of 19-year-old exposed to mercury in home between 4 and 9 years of age (Diamond et al., 1995).	*Methylmercury* Inferior performance on WMS mental control, orientation, logical memory, digit span, visual reproduction, and paired associate learning in adults with history of childhood lead exposure relative to non-exposed adults (White et al., 1993). Problems on WMS-R visual memory span, alphabet, counting backwards, digit symbol, picture completion, block design, and object assembly in case report of 19-year-old exposed to mercury in home between 4 and 9 years of age (Diamond et al., 1995).
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	
Lead Low prenatal and early preschool blood lead levels had no statistically significant association after adjustment for confounders with WPPSI Full-Scale, Verbal, and Performance IQs in cohort of children at 4 years, 10 months (Ernhart et al., 1989).	
Wechsler Intelligence Scales for Children	
PCBs No association due to exposure to PCBs in utero and WISC-R Performance IQ or perceptual organization (average of picture-completion, picture arrangement, block design, and object assembly subtests) (Jacobson and Jacobson, 1996). No association found on digit span, similarities, or block design subtests of the WISC-R in a study of 7-year-old Faroese Island children prenatally exposed to PCB (Grandjean et al., 2001).	*PCBs* Exposure to PCBs in utero associated with lower WISC-R full-scale IQ and verbal IQ scores at 11 years of age, with the most highly exposed indicating the largest deficits. IQ summary scales derived from factor analysis found association between PCB exposure and poorer verbal comprehension (vocabulary, information, and similarities subtests) and freedom from distractibility (coding and mazes subtests) scales (Jacobson and Jacobson, 1996). Taiwanese children exposed to PCBs in utero scored an average of 7 points lower on WISC-R Performance IQ than non-exposed controls (Rogan et al., 1988).
Lead No association at 10 years of age between blood lead level at 24 months WISC-R Performance IQ or with scores on information, vocabulary, digit span, picture arrangement object assembly, coding, and mazes subtests (Stiles and Bellinger, 1993). No statistically significant differences on WISC-R Performance IQ or arithmetic, comprehension, similarities, picture arrangement, block design, object assembly, coding, and mazes subtests between children with high lead levels and those with low lead levels on WISC-R (Needleman et al., 1979). At 10 years of age, there were no statistically significant associations between WISC-R Performance IQ and vocabulary, digit span, object assembly, information, picture arrangement, coding, and mazes subtest scores and blood lead levels at 24 months of age, after adjustment for confounders, and no association between WISC-R scores and blood lead levels at 6, 12, 18, and 57 months or current blood lead levels (Bellinger et al., 1992). No statistically significant association between high versus low lead (dentine) exposures and WISC Performance IQ (and subtests) and Arithmetic, Similarities, and Digit Span subtests (Hansen et al., 1989).	*Lead* Total and Verbal IQ scores on WISC-R were negatively correlated with lead levels in teeth in sample of 7- and 8-year-old Italian children (Bergomi et al., 1989). Continuing deficit on WISC-R Performance IQ at 6.5 years of age associated with post-natal blood lead concentrations in Cincinnati Lead Study Cohort (Dietrich et al., 1993). At 10 years of age, blood lead level at 24 months associated with significantly lower WISC-R Full-Scale and Verbal IQ and with lower scores on similarities, arithmetic, comprehension, picture completion, and block design subtests (Stiles and Bellinger, 1993). At 10 years of age, there was a statistically significant inverse association between WISC-R Verbal IQ, Full-Scale IQ, and arithmetic, comprehension, picture completion, similarities, and block design subtest scores and blood lead levels at 24 months of age, after adjustment for confounders (Bellinger et al., 1992). Mexican children with higher lead exposures had lower scores on WISC-RM (WISC-R for Mexico) factors associated with a sequential factor comprised of arithmetic, digit span, and coding subtests relative to low exposure counterparts (Calderon et al., 2001).

	Danish first-grade schoolchildren with high lead exposures (dentine levels) had lower WISC Verbal IQ and Full-Scale IQ scores, and lower Information, Comprehension, and Vocabulary scores than those in a low lead-exposed group (Hansen et al., 1989). On Mexican version of WISC, Mexico City children (7 to 9 years of age) with higher concentrations of blood lead had lower Full-Scale, Performance, and Verbal IQ scores than those with lower blood lead levels (Munoz et al., 1993). Study of 7-year-old children in Yugoslavia found an increase in blood lead levels associated with decreases in WISC-III Full-Scale, Verbal, and Performance IQs and lower scores on Freedom from Distractibility, Perceptual Organization, and Verbal Comprehension factor scores (Wasserman, et al., 1997). Children with high lead levels had significantly lower WISC-R Full-Scale IQ, Verbal IQ, and information, vocabulary, digit span, and picture completion subtests than those with low lead levels on WISC-R (Needleman et al., 1979). *Arsenic* Mexican children with higher arsenic exposures had lower scores on WISC-R for Mexico Verbal IQ than those with lower exposures. Higher levels of arsenic were also inversely associated with WISC-RM factors assessing knowledge (vocabulary, information, and arithmetic subtests) and concepts (similarities, comprehension, and language subtests) (Calderon et al., 2001).
Methylmercury No association with WISC-R Similarities scores or Square Root Block Design scores in 7-year-old children prenatally exposed to methylmercury (Grandjean et al., 1997). No association with WICS-III full-scale IQ scores in Seychelles Island children at 9 years of age due to prenatal methylmercury exposure (Myers et al, 2003).	*Methylmercury* Negative association with WISC-R Digit Span scores in 7-year-old children prenatally exposed to methylmercury (Grandjean et al., 1997). An association between MeHg exposure and digit span performance was noted in Amazonian children tested between the ages of 7 and 12. (Grandjean et al, 1999) A weak association between maternal hair concentrations during pregnancy of MeHg and performance on digit span was noted in children entering the first grade in Madeira. (Murata et al., 1999)
Wide Range Achievement Tests	
PCBs No association on WRAT-R Spelling and Arithmetic scores due to exposure to PCBs in utero in children at 11 years of age (Jacobson and Jacobson, 1996).	*Inorganic Mercury* Case report of 19-year-old exposed to mercury between 4 to 9 years finds WRAT-R performance below expectation on all 3 subtests (Diamond et al., 1995).
Wide Range Assessments of Memory and Learning	
Methylmercury No association with WRAML design memory scores in Seychelles Island children at 9 years of age due to prenatal methylmercury exposure (Myers et al, 2003).	

Wisconsin Card Sorting Test	
Lead No association between blood lead levels at birth and 6, 12, 18, and 24 months and WCST total correct, total errors, categories achieved, number of trials, perseverative responses, nonperseverative responses, number of perseverative errors, percent perseverative errors, percent conceptual level responses, trials to first category, learning to learn, and failure to maintain set at 10 years of age. No association between all tests except those involving perseverance and current and 57 month blood lead levels (Stiles and Bellinger, 1993).	*Lead* In a 10-year follow-up study of children prospectively followed since birth with relatively low lifetime lead exposures, higher blood lead levels at 57 months associated with number of WCST perseverative responses, number of perseverative errors, and percent perseverative errors and current blood lead levels and number of perseverative errors in 10-year-old children (Stiles and Bellinger, 1993). *PCBs* Prenatal PCB exposure associated with more perseverative errors on Wisconsin Card Sorting Test at 11 years of age in cohort of Michigan children recruited shortly after birth, though authors caution results may be due to response inhibition (Jacobson and Jacobson, 2003).
Woodcock Johnson Tests of Achievement	
Methylmercury No association with letter-word recognition and applied problems subtest scores in Seychelles Island children at 9 years of age due to prenatal methylmercury exposure (Myers et al, 2003).	

References

Baghurst, PA, McMichael AJ, Tong S, Wigg NR, Vimpani GV, Robertson EF. Exposure to environmental lead and visual-motor integration at 7 years: The Port Pirie cohort study. *Epidemiology* 1995; 6: 104–109.

Bellinger D, Leviton A, Waternaux C, Needleman H, Rabinowitz M. Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. *New England Journal of Medicine* 1987; 316: 1037–1043.

Bellinger D, Needleman H, Leviton A, Waternaux C, Rabinowitz M, Nichols M. Early sensory-motor development and prenatal exposure to lead. *Neurobehavioral Toxicology and Teratology* 1984; 6: 387–402.

Bergomi M, Borella P, Fantuzzi G, Vivoli G, Sturloni N, Cavazzuti G, Tampieri A, Tartoni PL. Relationship between lead exposure indicators and neuropsychological performance in children. *Developmental Medicine and Child Neurology* 1989; 31: 181–190.

Calderon J, Navarro ME, Jimenez-Capdeville ME, Santos-Diaz MA, Golden A, Rodriguez-Leyva I, Borja-Aburto VB, Diaz-Barriga F. Exposure to arsenic and lead and neuropsychological development in Mexican children. *Environmental Research* 2001; 85: 69–76.

Canfield RL, Henderson Jr. CR, Cory-Slechta DA, Cox C, Jusko TA, Lanphear BP. Intellectual impairment in children with blood lead concentrations between 10 micrograms per deciliter. *New England Journal of Medicine* 2003; 348: 1517–1526.

Cooney GH, Bell A, McBride W, Carter C. Neurobehavioral consequences of prenatal low level exposures to lead. *Neurotoxicology and Teratology* 1989; 11: 95–104.

Cordier S, Garel M, Mandereau L, Morcel H, Doineau P, Gosme-Seguret S, Josse D, White RF, Amiel-Tison C. Neurodevelopmental investigations among Methylmercury-exposed children in French Guiana. *Environmental Research Section A* 2002; 89: 1–11.

Dahl R, White RF, Weihe P, Sorensen N, Letz R, Hudnell HK, Otto DA, Grandjean P. Feasibility and validity of three computer-assisted neurobehavioral tests in 7-year-old children. *Neurotoxicology and Teratology* 1996; 18: 413–419.

Davidson PW, Myers GJ, Shamlaye C, Cox C, Gao P, Axtell C, Morris D, Sloane-Reeves J, Cernichiari E, Choi A, Palumbo D, Clarkson T. Association between prenatal exposure to methylmercury and developmental outcomes in Seychellois children: effect modification by social and environmental factors. *Neurotoxicology* 1999; 20(5): 833–842.

Diamond R, White RF, Gerr F, Feldman RG. A case of developmental exposure to inorganic mercury. *Child Neuropsychology* 1995; 1: 140–149.

Dietrich KN, Krafft KM, Bornschein RL, Hammond PB, Berger O, Succop PA, Bier M. Low-level fetal lead exposure effect on neurobehavioral development in early infancy. *Pediatrics* 1987; 80: 721–730.

Dietrich KN, Succop PA, Berger OG, Keith RW. Lead exposure and the central auditory processing abilities and cognitive development of urban children: the Cincinnati lead study cohort at age 5 years. *Neurotoxicology and Teratology* 1992; 14; 51–56.

Dietrich KN, Succop PA, Berger OG Hammond PB, Bornschein RL. Lead exposure and the cognitive development of urban preschool children: The Cincinnati lead study cohort at age 4 years. *Neurotoxicology and Teratology* 1991; 13: 203–211.

Dietrich KN, Berger OG, Succop PA, Hammond PB, Bornschein RL. The developmental consequences of low to moderate prenatal and postnatal lead exposure: Intellectual attainment in the Cincinnati lead study cohort following school entry. *Neurotoxicology and Teratology* 1993; 15; 37–44.

Ernhart C, Morrow-Tlucak M, Marler M, Wolf A. Low level lead exposure in the prenatal and early preschool periods: Early preschool development. *Neurotoxicology and Teratology* 1987; 9: 259–270.

Ernhart C, Morrow-Tlucak M, Wolf A, Super D, Drotar D. Low level lead exposure in the prenatal and early preschool periods: Intelligence prior to school entry. *Neurotoxicology and Teratology* 1989; 11: 161–170.

Gladen BC and Rogan WJ. Effects of perinatal polychlorinated biphenyls and dichlorodiphenyl dichorethene on later development. *Journal of Pediatrics* 1991; 119: 58–63.

Gladen BC, Rogan WJ, Hardy P, Thullen J, Tingelstad J, Tully M. Development after exposure to polychlorinated biphenyls and dichlorodiphenyl dichlorethene transplacentally and through human milk. *Journal of Pediatrics* 1988; 113: 991–995.

Gomaa A, Hu H, Bellinger D, Schwartz J, Tsaih SW, Gonzalez-Cossio T, Schnaas L, et al. Maternal bone lead as an independent risk factor for fetal neurotoxicity: A prospective study. *Pediatrics* 2002; 110: 110–118.

Grandjean P, Weihe P, White R, Debes F, Araki S, Yokoyama K, Murata K et al. Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicology and Teratology* 1997; 19: 417–428.

Grandjean P, Weihe P, Burse VW, Needham LL, Storr-Hansen E, Heinzow B, Debes F, et al. Neurobehavioral deficits associated with PCB in 7-year-old children prenatally exposed to seafood neurotoxicants. *Neurotoxicology and Teratology* 2001; 23: 305–317.

Grandjean P, White RF, Nielsen A, Cleary D, de Oliveira Santos E. Methylmercury neurotoxicity in Amazonian children downstream from gold mining. *Environmental Health Perspectives* 1999; 107(7): 587–591.

Hansen O, Trillingaard A, Beese I, Lyngbyes T, Grandjean P. A neuropsychological study of children with elevated dentine lead level: Assessment of the effect of lead in different socio-economic groups. *Neurotoxicology and Teratology* 1989; 11: 205–213.

Jacobson JL, Jacobson SW, Humphrey HEB. Effects of in utero exposure to polychlorinated biphenyls and related contaminants on cognitive functioning in young children. *Journal of Pediatrics* 1990; 116: 38–45.

Jacobson JL and Jacobson SW. Intellectual impairment in children exposed to polychlorinated biphenyls in utero. *New England Journal of Medicine* 1996; 335: 783–789.

Jacobson JL and Jacobson SW. Prenatal exposure to polychlorinated biphenyls and attention at school age. *Journal of Pediatrics* 2003; 143: 780–788.

Koopman-Esseboom C, Weisglas-Kuperus N, de Ridder MA, Van der Paauw CG, Tuinstra LG, Sauer PJ. Effects of polychlorinated biphenyl/dioxin exposure and feeding type on infants' mental and psychomotor development. *Pediatrics* 1996; 700–706.

McMichael AJ, Baghurst PA, Wigg NR, Vipani GV, Robertson EF, Roberts RJ. Port Pirie cohort study: environmental exposure and children's abilities at the age of four years. *The New England Journal of Medicine* 1988; 319: 468–475.

Myers GJ, Davidson PW, Cox C, Shamlaye CF, Tanner MA, Choisy O, Sloane-Reeves J, Marsh DO, et al., Neurodevelopmental outcomes in Seychllois children sixty–six months after in utero exposure to methylmercury from a maternal fish diet: Pilot study. *Neurotoxicology* 1995; 16: 639–652.

Myers GJ, Davidson PW, Cox C, Shamlaye CF, Palumbo D, Cernichiari E, Sloane-Reeves J, et al. Prenatal methylmercury exposure from ocean fish consumption in Seychelles child development study. *Lancet* 2003; 361: 1686–1692.

Myers GJ, Davidson PW, Cox C, Shamlaye CJ, Tanner MA, Marsh DO, Cernichiari SA, Lapham LW, Berlin M, Clarkson TW. Summary of the Seychelles child development study on the relationship of fetal methylmercury exposure to neurodevelopment. *Neurotoxicology* 1995; 16(4): 711–716.

Munoz H, Romieu I, Palazuelos E, Mancill-Sanchez T, Meneses-Gonzalez F, Hernandez-Avila M. Blood lead level and neurobehavioral development among children living in Mexico City. *Archives of Environmental Health* 1993; 48: 132–139.

Murata K, Weihe P, Renzoni A, Debes F, Vasconcelos R, Zino F, Araki S, Jorgensen P, White RF, Grandjean P. Delayed evoked potentials in children exposed to methylmercury from seafood. *Neurotoxicology and Teratology* 1999; 21(4): 343–348.

Needleman HL, Gunnoe C, Leviton A, Reed R, Peresie H, Maher C, Barrett P. Deficits in psychologic and classroom performance of children with elevated dentine lead levels. *New England Journal of Medicine* 1979; 300: 689–695.

Needleman HL, Shell A, Bellinger D, Leviton A, Allred EN. The long-term effects of exposure to low doses of lead in childhood: an 11-year follow-up report. *New England Journal of Medicine* 1988; 322: 83–88.

Patandin S, Lanting CI, Mulder PGH, Boersma ER, Sauer PJJ, Weisglas-Kuperus N. Effects of environmental exposure to polychlorinated biphenyls and dioxins on cognitive abilities in Dutch children at 42 months of age. *Journal of Pediatrics* 1999 134; 33–41.

Palumbo D, Cox C, Davidson P, Myers G, Choi A, Shamlaye C, Sloane-Reeves J, Cernichiari E, Clarkson T. Association between prenatal exposure to Methylmercury and cognitive functioning in Seychellois children: A reanalysis of the McCarthy scales of children's ability from the main cohort study. *Environmental Research Section A* 84 2000; 81–88.

Rogan WJ, Bladen BC, Hung K, Koong S, Shih L, Taylor JS, Wy Y, et al. Congenital poisoning by polychlorinated biphenyls and their contaminants in Taiwan. *Science* 1988; 241: 334–636.

Shen X, Yan C, Guo D, Wu S, Li R, Huang H, Ao L, Zhou J et al. Low-level prenatal lead exposure and neurobehavioral development of children in the first year of live: A prospective study in Shanghai. *Environmental Research* 1998; 79: 1–8.

Stewart PW, Reihman J, Lonky EI, Darvill TJ, Pagano J. Cognitive development in preschool children prenatally exposed to PCBs and MeHG. *Neurotoxicology and Teratolgy* 2003; 25: 11–22.

Stiles KM and Bellinger DC. Neurological correlates of low-level lead exposure in school-age children: A prospective study. *Neurotoxicology and Teratology* 1993; 15: 27–35.

Takser L, Mergler D, Hellier G, Sahuquillo J, Huel G. Manganese, monoamine metabolite levels at birth, and child psychomotor development. *Neurotoxicology* 2003; 24: 667–674.

Till C, Koren G, Rovet JF. Prenatal exposure to organic solvents and child neurobehavioral performance. *Neurotoxicology and Teratology* 2001; 23: 235–245.

Vreugdenbil HJI, Lanting CI, Mulder PGH, Boersma ER, Weisglas-Kuperus N. Effects of prenatal PCB and dioxin background exposure on cognitive and motor abilities in Dutch children at school age. *Journal of Pediatrics* 2002; 140: 48–56.

Walkowiak J, Wiener J, Fastabend A, Heinzow B, Kramer U, Schmidt E, Steingruber H, Wundram S, Winnke G. Environmental exposure to polychlorinated biphenyls and quality of the home environment: effects on psychodevelopment in early childhood. *Lancet* 2001; 358: 1602–07.

Wasserman GA, Graziano JH, Factor-Litvak P, Popovac D, Morina N, Musabegovic A, Vrenezi N, et al. Independent effects of lead exposure and iron deficiency anemia on developmental outcome at age 2 years. *Journal of Pediatrics* 1992; 121: 695–703.

Wasserman GA, Graziano JH, Factor-Litvak P, Popovac D, Morina N, Musabegovic A, Vrenezi N, et al. Consequences of lead exposure and iron supplementation on childhood development at 4 years. *Neurotoxicology and Teratology* 1994; 16: 233–240.

Wasserman GA, Liu X, Lolacono NJ, Factor-Litvak P, Kline JK, Popovac D, Morina N, Musabegovic A, et al. Lead exposure and intelligence: The Yugoslavia prospective study. *Environmental Health Perspectives* 1997; 105: 956–962.

White RF, Diamond R, Proctor S, Morey C, Hu H. Residual cognitive deficits 50 years after lead poisoning during childhood. *British Journal of Industrial Medicine* 1993; 50: 613–622.

Wigg NR, Vimpani GV, McMichael AJ, Baghurst PA, Robertson EF. Port Pirie cohort study: childhood blood lead and neuropsychological development at age two years. *Journal of Epidemiology and Community Health* 1988; 42: 213–219.

Winneke G, Brockhaus A, Ewers U, Kramer U, Neuf M. Results from the European multicenter study on lead neurotoxicity in children: implications for risk assessment. *Neurotoxicology and Teratology* 1990; 12: 553–559.

Table 4. Recommended Test Matrix

	.5/ 1 yr	1.5 / 2 yr	3 yr	4 yr	5 yr	6 yr	7 yr	8 yr	9 yr	10 yr	11 yr	12 yr	13 yr	14 yr	15 yr	16 yr	17 yr	18 yr	19 yr	20 yr
Screen																				
Bayley - II	a/b	c/d																		
IQ																				
WPPSI-III (BD, MR, V)			a	b	c	d														
WASI (BD, MR, V, S)							a	b	c	D					a	b	c	d		
Academics																				
WRAT 3 (S, A, Re)							a	b	c	D					a	b	c	d		
Attention																				
Conners rating scale-R			a	b	c	d	a	b	c	D				d	a	b	c			
Conners CPT-II				b	c	d	a	b	c	D	a	b	c	d	a	b	c	d	a, b	c, d
Digits forward (Wech)							a	b	c	D	a	b	c	d	a	b	c	d		
Executive Function																				
Digits backwards (Wech)											a	b	c	d	a	b	c	d		
WCST							a	b	c	D					a	b	c	d		
Trail making test											a	b	c	d					a, b	c, d
Similarities (WASI)							x	x	x	X					x	x	x	x		
Language																				
BNT						d	a	b	c	D					a	b	c	d		
EVT (Naming)			a	b	c															
EVT (S)											a	b	c	d					a, b	c, d
PPVT - III			a	b	c	d	a	b	c	D					a	b	c	d		
CELF - 3 (sent. structure)											a	b	c	d					a, b	c, d
Vocabulary (Wechsler)			x	x	x	x	x	x	x	X					x	x	x	x		
Visual-Spatial																				
Bender Gestalt-II							a	b	c	D					a	b	c	d		
HVOT											a	b	c	d					a, b	c, d
ROCF											a	b	c	d					a, b	c, d
Beery VMI-5			a	b	c	d					a	b	c	d						
MR, BD (Wechsler)			x	x	x	x	x	x	x	X					x	x	x	x		
Learning & Memory																				
Bender-II recall							a	b	c	D					a	b	c	d		
ROCF recall											a	b	c	d					a, b	c, d
CVLT-C					c	d	a	b	c	D										
CVLT-II															a	b	c	d	a, b	c, d
WRAML 2-stories					c	d	a	b	c*	D	a	b	c	d	a	b				
WRAML 2-verbal learning											a	b	c	d						
WMS-III logical memory																	c	d	a, b	c, d
Coding recall (Wechsler)				b	c	d	a	b	c	D	a	b	c	d	a	b	c	d		
Motor																				
Revised Purdue pegboard			a	b	c	d														
Grooved pegboard							a	b	c	D	a	b	c	d	a	b	c	d	a, b	c, d
Finger tapping test							a	b	c	D	a	b	c	d	a	b	c	d	a, b	c, d
Coding (Wechsler)				b	c	d	a	b	c	D	a	b	c	d	a	b	c	d		
Novel/Pilot Tests			c	d	a	b	c	d	a	B	c	d	a	b	c	d	a	b	a, b	c, d

(* = stimuli change at age 9)

Table 5. Recommended Test Batteries by Developmental Stage

Group	*Age*	*Tests to be Administered*	*Estimated Time*
A	0.5 year	Bayley-II	20 minutes
	3 years	WPPSI–III (BD, MR, V) Conners rating scale revised EVT Naming PPVT-III Beery VMI-5 Revised Purdue Pegboard	65 minutes
	7 years	WASI (BD, MR, V, S) WRAT-3I (S, A, R) Conners Rating Scale Revised Conners CPT-II Digits forward (WISC–IV) WCST BNT PPVT-III Bender Gestalt-II (w/ recall condition) CVLT–C WRAML-2–stories Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	135 minutes
	11 years	Conners CPT-II Digits forwards (WISC–IV) Digits backwards (WISC–IV) Trail-making Test EVT (S) CELF 3-sentence structure Hooper VOT ROCF (w/ recall condition) Berry VMI-5 WRAML-2–stories WRAML-2–verbal learning Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	90 minutes
	15 years	WASI (BD, MR, V, S)	135 minutes

Group	*Age*	*Tests to be Administered*	*Estimated Time*
		WRAT-3 (S, A, R) Conners CPT-II Conners Rating Scale Revised Digits forward (WISC–IV) Digits backwards (WISC–IV) WCST BNT PPVT-III Bender Gestalt-II (w/ recall condition) CVLT-II WRAML-2–verbal learning Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	
	19 years	Conners CPT-II Trail-making Test EVT (S) CELF 3–sentence structure Hooper VOT ROCF (w/ recall condition) CVLT-II WMS-III–logical memory Grooved Pegboard Finger tapping	95 minutes
B	1 year	Bayley-II	20–30 minutes
	4 years	WPPSI-III (BD, MR, V) Conners Rating Scale Revised Conners CPT-II EVT (Naming) PPVT-III Beery VMI-5 Coding (w/ recall) (WPPSI–III) Revised Purdue Pegboard	60 minutes
	8 years	WASI (BD, MR, V, S) WRAT-3 (S, A, R) Conners Rating Scale Revised Conners CPT-II	143 minutes

Group	*Age*	*Tests to be Administered*	*Estimated Time*
		Digits forward (WISC–IV) WCST BNT PPVT-III Bender Gestalt–II (w/ recall condition) CVLT–C WRAML-2–stories Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	
	12 years	Conners CPT-II Digits forwards (WISC–IV) Digits backwards (WISC–IV) Trail-making test EVT (S) CELF 3-sentence structure Hooper VOT ROCF (w/ recall condition) Berry VMI-5 WRAML-2–stories WRAML-2–verbal learning Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	95 minutes
	16 years	WASI (BD, MR, V, S) WRAT-III (S, A, R) Conners Rating Scale revised Conners CPT-II Digits forward (WISC–IV) Digits backwards (WISC–IV) WCST BNT PPVT-III Bender Gestalt–II (w/ recall condition) CVLT-II WRAML–2–verbal learning Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	140 minutes

Group	*Age*	*Tests to be Administered*	*Estimated Time*
	19 years	Conners CPT-II Trail-making Test EVT (S) CELF 3–sentence structure Hooper VOT ROCF (w/ recall condition) CVLT-II WMS-III–logical memory Grooved Pegboard Finger Tapping	95 minutes
C	1.5 years	Bayley-II	30 minutes
	5 years	WPPSI -III (BD, MR, V) Conners Rating Scale Revised Conners CPT-II EVT (Naming) PPVT-III Beery VMI-5 CVLT–C WRAML-2–stories Coding (w/ recall) (WWPSI–III) Revised Purdue Pegboard	100 minutes
	9 years	WASI (BD, MR, V, S) WRAT-3 (S, A, R) Conners Rating Scale Revised Conners CPT-II Digits forward (WISC–IV) WCST BNT PPVT -III Bender Gestalt-II (w/ recall condition) CVLT–C WRAML-2–stories Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	158 minutes
	13 years	Conners CPT-II Digits forwards (WISC–IV)	95 minutes

Group	*Age*	*Tests to be Administered*	*Estimated Time*
		Digits backwards (WISC–IV) Trail-making Test EVT (S) CELF 3-sentence structure Hooper VOT ROCF (w/ recall condition) Berry VMI-5 WRAML-2–stories WRAML-2–verbal learning Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	
	17 years	WASI (BD, MR, V, S) WRAT-3 (S, A, R) Conners CPT-II Conners Rating Scale Revised Digits forward (WISC–IV) Digits backwards (WISC–IV) WCST BNT PPVT-III Bender Gestalt-II (w/ recall condition) CVLT-II WMS-III–logical memory Coding (w/ recall) (WAIS-III) Grooved Pegboard Finger Tapping	145 minutes
	20 years	Conners CPT-II Trail-making Test EVT (S) CELF 3–sentence structure Hooper VOT ROCF (w/ recall condition) CVLT-II WMS–III–logical memory Grooved Pegboard Finger Tapping	95 minutes
D	2 years	Bayley-II	30 minutes

Group	*Age*	*Tests to be Administered*	*Estimated Time*
	6 years	WPPSI–III (BD, MR, V) Conners rating scale revised Conners CPT-II BNT PPVT-III Beery VMI-5 CVLT–C WRAML-2–stories Coding (w/ recall) (WISC–IV) Revised Purdue Pegboard	100 minutes
	10 years	WASI (BD, MR, V, S) WRAT-3 (S, A, R) Conners Rating Scale Revised Conners CPT-II Digits forward (WISC–IV) WCST BNT PPVT-III Bender Gestalt-II (w/ recall condition) CVLT–C WRAML-2–stories Coding (w/ recall) (WISC–IV) Grooved Pegboard Finger Tapping	158 minutes
	14 years	Conners CPT-II Conners Revised Rating Scale Digits forwards (WISC–IV) Digits backwards (WISC–IV) Trail-making Test EVT–(S) CELF 3-sentence structure Hooper VOT ROCF (w/ recall condition) Berry VMI-5 WRAML-2–stories WRAML-2–verbal learning Coding (w/ recall) (WISC–IV)	95 minutes

Group	*Age*	*Tests to be Administered*	*Estimated Time*
		Grooved Pegboard Finger Tapping	
	18 years	WASI (BD, MR, V, S) WRAT–3 (S, A, R) Conners CPT-II Digits forward (WAIS–III) Digits backwards (WAIS–III) WCST BNT PPVT-III Bender Gestalt-II (w/ recall condition) CVLT-II WMS-III–logical memory Coding (w/ recall) (WAIS–III) Grooved Pegboard Finger Tapping	150 minutes
	20 years	Conners CPT-II Trail-making Test EVT (S) CELF 3–sentence structure Hooper VOT ROCF (w/ recall condition) CVLT-II WMS-III–logical memory Grooved Pegboard Finger tapping	85 minutes

Table 6. Alternative Tests

Age Group	Domain	Recommended Tests	Alternative Tests
0–2 years	**Screening**	**Bayley-II**	Mullen
3–6 years	**IQ**	**WPPSI-III (BD, MR, V)**	Full WPPSI-III McCarthy KABC-2 KBIT-2 S-B-V
	Attention	**Conners Rating Scale-R** **Conners CPT-II (4–6 yrs)**	NES CPT, Animals (5–6 yrs) Digits forward (WISC) (6 yrs)
	Executive	**None**	None
	Language and Verbal Function	**EVT–naming (3–5 yrs)** **BNT (6 yrs)** **PPVT-III** **WPPSI–III (Vocab)**	WPPSI-III–naming EVT–synonyms (6 yrs)
	Visuospatial	**WPPSI–III (BD, MR)** **VMI-5**	Bender Gestalt-II (4–6 yrs) HVOT (5–6) S-B–IV–copying NEPSY–visuospatial
	Learning and Memory	**WISC–IV (coding recall) (4–6 yrs)** **CVLT–C (5–6 yrs)** **WRAML-2 (stories) (4–6 yrs)**	CVLT-C (5–6 yrs) WRAML-2–verbal learning (5–6 yrs) NEPSY (5–6 yrs) (Narrative)
	Motor	**Revised Purdue pegboard** **WPPSI-III–coding (4–6 yrs)**	NEPSY (5–6 yrs)
7–10 years	**IQ**	**WASI (BD, MR, V, S)**	WISC –IV S-B V KABC-2 KBIT-2
	Academic	**WRAT-3**	Woodcock-Johnson Kaufman WIAT
	Attention	**Conners Rating Scale** **Conners CPT-II**	NES CPT NEPSY

Age Group	Domain	Recommended Tests	Alternative Tests
		WISC–IV (digits forward)	
	Executive	**WCST** **WASI (S)**	Children's Categories Children's Color Trails (8–10) Stroop
	Language	**BNT** **PPVT-III** **WASI (Vocab)**	Token Test for Children Revised Token Test EVT CELF 3 IQ subtests
	Visuospatial	**Bender Gestalt-II** **WASI (BD, MR)**	HVOT ROCF IQ subtests VMI-5 S-B-IV copying
	Learning and Memory	**Bender Gestalt-II recall** **CVLT–C** **WRAML–2 (stories)** **WISC–IV (coding recall)**	NEPSY subtests WRAML 2 verbal learning ROCF recall
	Motor	**Finger Tapping** **Grooved Pegboard** **WISC–IV (coding)**	NES FTT NES Hand-eye coordination Santa Ana Dynamometer Purdue pegboard NEPSY FTT
11–14 years	**IQ**	**None**	WASI WISC–IV S-B–V KBIT-2 KABC-2
	Academics	**None**	WRAT-3 WIAT Woodcock-Johnson Kaufman
	Attention	**Conners CPT-II** **Conners rating scale -R (14 yrs)** **WISC –IV (digits forward)**	NES CPT
	Executive	**WISC –IV (digits backwards)** **Trail-making Test**	WCST Trail-making Test Children's Categories Stroop

Age Group	Domain	Recommended Tests	Alternative Tests
			Children's Color Trails
	Language	**BNT** **PPVT-III**	EVT CELF 3 (sentence structure)
	Visuospatial	**HVOT** **ROCF**	VMI-5 Bender Gestalt-II IQ subtests
	Learning and Memory	**ROCF recall** **WRAML–2 (stories)** **WISC–IV (coding recall)**	CVLT-C CMS WRAML–2 (verbal learning)
	Motor	**FTT** **GP** **WISC–IV (coding)**	NES-FTT Santa Ana Dynamometer
15–18 years	**IQ**	**WASI**	WISC-IV (15–16 yrs) WAIS-III (16–18 yrs) S-B-V KABC-2 KBIT-2 KAAIT
	Academic	**WRAT-3**	WIAT Woodcock-Johnson Kaufman
	Attention	**Conners rating scale-R (15–17 yrs)** **Conners CPT-II** **WISC–IV, WAIS–III (digits forward)**	NES-CPT IQ subtests
	Executive	**Digits backwards** **(Wechsler)** **WCST** **WASI (S)**	Trail-making Test Color Trails (18) Children's color trails (15–17) Stroop PASAT IQ subtests
	Language	**BNT** **PPVT-III** **WASI-V**	Revised Token IQ subtests EVT CELF 3
	Visuospatial	**Bender Gestalt -II** **WASI (MR, BD)**	HVOT ROCF IQ subtests VMI–5 (15–17 yrs)
	Learning and Memory	**Bender Gestalt-II recall** **CVLT–2** **WRAML–2 (verbal learning) (15–16 yrs)** **WMS–III (logical memory) (17–18 yrs)** **WISC–IV, WAIS–III**	WMS-III subtests (16–18 yrs) WRAML–2 (stories) ROCF recall

Age Group	Domain	Recommended Tests	Alternative Tests
		(coding recall)	
	Motor	**FTT** **Grooved Pegboard**	Dynamometer Santa Ana Purdue Pegboard NES FTT
19–20 years	**IQ**	**None**	
	Academics	**None**	
	Attention	**Conners CPT-II**	Any previous age-appropriate test
	Executive	**Trail making Test**	Any previous age-appropriate test
	Language	**BNT** **PPVT-III**	Any previous age-appropriate test
	Visuospatial	**ROCF**	Any previous age-appropriate test
	Learning and Memory	**ROCF recall** **CVLT-II**	Any previous age-appropriate test WMS-III
	Motor	FTT Grooved Pegboard	Any previous age-appropriate test