The Assessment of Neurobehavioral Toxicity: SGOMSEC Joint Report

Nancy Fiedler,[1] Robert G. Feldman,[2] Joseph Jacobson,[3] Alice Rahill,[4] and Anthony Wetherell[5]

[1]UMDNJ-Robert Wood Johnson Medical School, Environmental and Occupational Health Sciences Institute, Piscataway, New Jersey; [2]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts; [3]Wayne State University, Psychology Department, Detroit, Michigan; [4]Department of Occupational Medicine, University of Rochester, Rochester, New York; [5]Human Factors Group, Chemical and Biological Defense Establishment, Porton Down Salisbury, United Kingdom

Exposure to neurobehavioral toxicants is a problem of international scope. Although many different procedures are available for the assessment of human behavioral function, performance tests are displacing traditional diagnostic tests for ascertaining the consequences of exposure to neurotoxic chemicals. Performance testing includes variables such as attention and concentration, sensory function, motor control, spatial relations, visuomotor coordination, memory, and affect. Special tests have also been devised for evaluating child development. One of the salient needs in these efforts is the construction of databases allowing access to normative data. — Environ Health Perspect 104(Suppl 2):179–191 (1996)

Key words: child development, computerized tests, intelligence, neurobehavioral assessment, neurological examinations, neuropsychological testing, performance testing, peripheral neuropathies, test batteries

Introduction

Exposure to neurobehavioral toxicants in the environment is an urgent international problem. The problems range from the catastrophic effects of industrial accidents such as Bhopal to ubiquitous background

environmental exposures to chemicals such as lead. While it is expected that high-level exposures occur only in developing nations, such exposures continue to occur in the most technologically advanced countries in the world. For example, despite widespread knowledge of the deleterious effects of inorganic lead on the nervous system, cases of lead poisoning (i.e., > 80 mg/dl in blood) have been documented within the construction trades in this decade (1). The neurobehavioral effects of acute high-level exposures are well known. For example, the overt clinical manifestations of acute pesticide or lead poisoning can be easily recognized. Patients presenting with acute delirium and convulsions accompanied by high blood lead levels do not require neurobehavioral testing to document the adverse health effects due to exposure. Further, neurobehavioral tests were not required to document the nervous system effects in Bhopal. However, for less catastrophic exposures, neurobehavioral assessment plays an important role in determining the functional impact of neurotoxic exposure.

Why is the assessment of the behavioral impact of neurotoxicants important? Unlike other chemicals such as carcinogens, for which no evidence of excessive exposure may be seen for years, exposure to neurotoxicants may impact an individual's functioning directly. Thus, entire subsets of a population may experience a reduced level of function in response to the effects of acute or chronic exposures. Neurobehavioral tests provide a systematic method for documenting behaviors that are essential for optimal functioning in a technologically complex society. Weiss (2) illustrated this point in the case of chronic low-level lead exposure. For example, suppose that background lead exposure at relatively low levels (10 mg/dl blood lead) reduces scores by 5 points (5 %) on a standard intelligence test. Translated into population terms, such a shift means that, in a population of 100 million, only 990 thousand individuals, rather than 2.3 million, will score above 130 (that is, in the upper ranges of intellectual function). There would be a corresponding inflation of the proportion of the population scoring below 70. The only way in which this impact could be known is through the use of standardized neurobehavioral tests.

On an acute basis, accident rates may increase on the job or in the community due to the transitory effects of exposure. Also, acute exposures may have a differential effect on subsets of the exposed population based on individual differences in susceptibility. For example, individuals reported to suffer from multiple chemical sensitivities (3) experience acute neurobehavioral symptoms in response to low-level environmental exposures such as solvents in perfumes and cleaning products. Neurobehavioral tests can document the extent of functional impairment due to acute exposures.

Recognition that a significant neurotoxic exposure has occurred can grow out of a suspicion that a hazard exists or from complaints and observations of changes in individuals. Suspicion that a hazard exists can arise from routine surveillance (monitoring of urine or other biological samples) or an industrial accident. Complaints may be initiated by individuals who perceive changes in their own cognitive, motor, or affective function. Over time, such individuals may come to be recognized as having been subjected to a common occupational

or residential exposure, often referred to as a cluster. Complaints may also consist of reports by parents that infants or children are failing to develop at the expected rate (developmental delay) or observations by health care workers of an increased incidence in particular functional deficits among the clients they serve.

In public health terms, neurobehavioral assessment provides a useful methodology for the regulator to detect and characterize health effects at primary, secondary, and tertiary levels of prevention. At the level of primary prevention, a neurotoxic exposure may occur at what are regarded as background levels and individuals may not appear symptomatic. Neurobehavioral testing can provide a systematic evaluation of subclinical effects. Childhood lead exposure provides a good example of the utility of these methods for detection of an effect at this level of prevention. No cases are evident; lead's neurotoxic impact is revealed by a population shift in IQ scores. For secondary prevention, exposure to a neurotoxic agent may be documented and individuals appear symptomatic. Neurobehavioral tests are important to systematically establish the objective impact on behavioral function. Documentation of behavioral dysfunction at this level can help prevent morbidity or permanent dysfunction. Finally, at the tertiary prevention level, acute high-level or chronic exposure to neurotoxic agents may occur, and individuals may show symptoms of behavioral dysfunction (e.g., poor concentration or memory). Neurobehavioral assessment will provide a standardized evaluation of the level of disability and can be used prospectively to track the effects of intervention, such as removal from exposure or the permanence of impairment. Examples of the utility of neurobehavioral assessment for tertiary prevention are clearly presented in the literature on organic solvents and lead (4).

The purpose of this paper is 4-fold: a) to discuss the situations or problems for which neurobehavioral assessment techniques are useful; b) to outline the functional aspects of behavior that should be included in an assessment; c) to provide guidelines for the use of the various methods available; and d) to discuss the advantages and disadvantages of these methods. Special consideration will be given to the sensitivity of these methods for detecting and characterizing effects at each level of prevention. Neurobehavioral methods for assessment of child development will also be reviewed.

Targets of Neurotoxic Effects

Behavior is the outcome of multiple mechanisms within the central nervous system (CNS); its expression may be internal (subjective state) or externally observable by others. In humans emotional responses to stress, learning processes, and innovative problem-solving techniques are expected activities of the intact nervous system. The CNS is vulnerable to the actions of environmental factors that include physical conditions (trauma, temperature), as well as chemical factors. Exposure to chemicals may result in neurobehavioral effects depending on the particular chemical, the circumstances of exposure, the duration and intensity of the exposure, and the susceptibility of the organism.

Because human behaviors change routinely in normal conditions to adapt to actual and perceived conditions, the definition of normal or baseline parameters of behavior is difficult. Consequently, it becomes even more important to have methods for detecting changes in neurobehavior when they are the outcome of chemical exposure or conditions other than ordinary life experiences.

Peripheral Nervous System

The central nervous system comprises the brain and spinal cord. The components of the peripheral nervous system (PNS) lie outside these structures and include the spinal and cranial nerves. Motor and sensory symptoms can arise from damage to peripheral nerves. Clinical manifestations of toxic peripheral neuropathies begin with complaints of numbness or tingling, usually in the feet before the fingers because the longer nerve fibers are affected first. Damage may progess to the less distal portions of a nerve as the exposure continues. Even after ending exposure to chemicals capable of inducing neuropathy, there will be further progression of neurologic impairment followed by a plateau and a very slow or gradual recovery of function in some instances. Nerve damage is often irreversible, however.

Central Nervous System

The central nervous system is vulnerable to neurotoxic effects at lower levels of exposure than the peripheral nervous system. In fact, depending upon the particular neurotoxicant, an individual may be unaware of any relationship between symptoms and exposure. Such a patient may exhibit behavioral changes recognized only by his family or co-workers. Neurobehavioral

effects of exposure to neurotoxicants usually precede other symptoms, including those of peripheral neuropathy. Symptoms such as poor attention, drowsiness, memory problems, mood changes, and impaired fine motor performance may interfere with job tasks resulting in costly injuries and lost productivity.

Neurotoxic effects on the central nervous system must be differentiated from effects induced by other neurological disorders. Behavioral effects in humans can be acute or insidious and chronic in their emergence. Attention has been focused on the need to identify neurotoxic effects of the brain as early as possible to avoid permanent damage by continuing exposure. Carefully selected neuropsychological testing provides standardized procedures for evaluating specific aspects of behavioral function arising from damage to various areas of the brain.

Application of Neurobehavioral Methods

Historically, four approaches have been used to evaluate neurobehavioral function: the clinical neurological examination, self-report checklists, performance tests, and neuropsychological tests. Primary prevention is the concern when an exposure has occurred, but individuals are generally asymptomatic and the nature and degree of neurobehavioral impairment are unknown. When this occurs, computerized performance tests and self-report checklists may be most appropriate because they are the most sensitive in detecting relatively subtle effects. Where secondary prevention is the concern because there is some overt evidence of neurobehavioral dysfunction, such as health complaints from individuals or exposed groups, both performance tests and traditional neuropsychological assessments will be useful. When clear evidence of behavioral dysfunction due to exposure is available, administration of a neurological examination together with a neuropsychological test battery can estimate the nature and degree of impairment.

In the case of primary and secondary prevention in which the degree of impairment is more subtle, it is usually not possible to link dysfunction to exposure in the individual case. In some instances when individuals are routinely exposed to hazardous chemicals, administration of performance tests and self-report checklists at the beginning of the exposure (e.g., prior to employment) and at regular intervals thereafter can provide a useful baseline

against which neurobehavioral changes can be evaluated. In the absence of such a baseline, neurobehavioral effects can be ascertained only by comparing level of function to established norms or by comparing a group of exposed individuals with nonexposed controls. Unfortunately, published norms do not exist for performance tests. Therefore, most primary prevention studies, designed to detect subtle effects of exposure, must rely on comparisons with an appropriate control group or on changes from baseline relative to controls to detect such effects. Many of the neuropsychological tests, on the other hand, have established norms. However, established norms will not necessarily be relevant for evaluating effects in the exposed population, which may differ from the normative sample in ethnic and cultural background, educational level, age, etc. Also, even in the case in which symptomatic individuals are being evaluated (i.e., secondary prevention), effects may be relatively subtle and may not fall within the impaired range on standard neuropsychological tests. Therefore, comparison with an appropriate control group will be necessary.

Ideally, if neurobehavioral testing were to become a standard method for monitoring exposure-related effects in the workplace, it would be introduced before exposure and provide baselines for subsequent evaluations. Selecting appropriate controls in the absence of historical data is difficult but essential. In general, controls should be as similar as possible to exposed individuals in ethnic background, socio-economic status, educational attainment, occupation, age, and gender. A matching procedure can be used to ensure that for each exposed subject a control similar in these background characterisitics is also evaluated. To enhance sensitivity, a 2:1 or 3:1 ratio of subjects to controls is desirable. It is never feasible to match on all the relevant background characteristics, however. Any characteristic in which the exposed and control subjects differ other than the exposure itself is considered a potential confounding variable. For example, if more exposed subjects than controls are chronic alcoholics, any observed differences in neurobehavioral function between the two groups may well be due to differences in incidence of alcoholism rather than to exposure. Therefore, even if the exposed and control subjects are matched on certain critical background characterisitics, it is important to assess other potential confounding variables as well.

When between-group differences are found on such variables, they can be controlled statistically in all analyses comparing the performance of exposed subjects with that of controls. A detailed discussion about the selection of and statistical control for potential confounders can be found in Jacobson and Jacobson (5).

Finally, the evaluation, in which tertiary prevention is the assessment level, most closely resembles a traditional clinical evaluation of brain injury. A neurological examination will be most useful in this situation in which overt symptoms of frank dysfunction can guide the inquiry. Also, one would expect significant discrepancies in performance on standardized neuropsychological tests. Thus, comparison with normative values for the individual case as well as for groups of individuals should reveal significant discrepancies (e.g., 2 standard deviations below the mean) from expected values for individuals of similar demographic composition. As for the other examples, appropriate controls must be selected for group comparisons.

Test Methods

Neurological Assessment

The neurotoxicant-exposed individual is at risk for developing changes in locomotor, peripheral, sensory, and neurobehavioral functions involving perceptual, cognitive, and communications skills. Levels of toxic effect determine severity of impairment or disability. Although possibly detectable as subclinical manifestations on sensitive neurologic tests, earliest effects following exposure may be unrecognized by the subject. For example, a metabolic change could be measured in a blood or urine test or with electronically determined nerve conduction velocity studies while the patient has no symptoms and an examiner cannot detect abnormalities on clinical examination even with specially designed tests.

In the case of secondary or tertiary prevention situations, the patient will report neurologic complaints such as an unusual sensation in the extremities (e.g., neuropathy) or a change in mood. At this level, formal clinical neurologic tests may be able to identify indications of a physiologic effect. Diagnosis of toxic peripheral neuropathy is made by testing perception of sensations of pain, temperature, vibration, and joint position. Each of these modalities, if intact, will indicate the preservation of function in the various sizes of nerve fibers.

As outlined by Feldman and White (6), the examination would begin with a clinical interview, which would include a detailed description of the symptoms and functional changes that have been noted and their time of onset, duration, and intensity. To aid in determining whether the presenting complaint is related to a chemical exposure, detailed information should also be obtained regarding any exposure to chemicals in the workplace, home or hobbies, as well as any genetic and/or congenital factors that might provide alternative explanations for the complaint [see Feldman and White (6); Table 3].

Normal motor function can be defined as the ability of a person to initiate, sustain, and effectively perform desired movement of a part or all of the body with speed, accuracy, and strength. Automatic motor behaviors are completed without conscious awareness, using reflex pathways and adaptive mechanisms that adjust to variables in resistance to the intended motor activity from gravity or other obstacles. Weakness is perceived by an affected person as the need to exert more than usual effort to accomplish an action previously done without additional effort. Qualitative assessment of strength by grip tests, weight lifts, or other measures that require overcoming known resistance can provide objective data about one aspect of motor functioning. Other techniques are necessary to record tremor and coordination and to evaluate postural control (7). Walking pattern, a common clinically observable motor response, can be quantified by measuring the distance between foot positioning at rest and between foot placement while walking. The axial posture, flexed or erect, is an important indication of the ability of spinal-central reflexes to maintain the posture. This function uses spinal cord posterior columns as pathways of conduction potential and vibration sensation to inform the brain where the patient's feet are, and to send instructions from cerebellum and cerebral motor systems (extrapyramidal) to regulate a balance between the agonist and antagonist musculature.

Reduced sensation, demonstrated bilaterally in a symmetrical stocking and glove pattern, and hypoactive tendon reflexes in the ankles and knees indicate peripheral neuropathy on neurological exam. Electrophysiological methods are available to document the ability of a peripheral nerve to conduct an evoked nerve impulse and to measure its amplitude and speed of conduction. Therefore, neurotoxic effects on

the peripheral nervous system can be ascertained in individual cases or in groups of exposed persons.

Sensory systems contribute to the ability to move with coordination and finesse. Thus, impairment in tactile, visual, and auditory systems, as well as the vibratory and point position sensations will produce unsteadiness of gait (ataxia) and poor coordination. Motor functions cannot be expressed independently of the sensory systems with which they function.

Feldman and White (6) summarized the basic neurologic examinations used to detect nervous system function. In addition, they describe the common electrophysiologic techniques used to obtain evidence of disturbances in brain functions (i.e., electroencephalogram, evoked potentials, and imaging) and peripheral nerves. It emphasized that electrophysiological tests may be applied differentially, depending on techniques and recording conditions, instrumentation, and methods of data collection and interpretation. Conventional methods (8–10) have been well described. The results of all these techniques are not specific to any particular neurotoxicant. The changes reflect only the physiologic process and whether or not they are affected.

Performance Testing

Performance tests are designed to assess whether an individual can do a designated job. They can be sensitive, reliable, cheap, and quick and easy to administer and have their main application when considering groups of people at the primary prevention level of public health. Performance tests may also prove useful at the secondary level, but they are not normally considered at the tertiary level unless some particular type of performance is at issue. However, it is important to remember that while performance tests may show acceptable content and construct validity (that is, they are internally consistent and can be placed in the context of accepted theories or models of performance), they can be deficient in criterion validity [the degree to which they actually reflect real-life situations (11)].

Performance tests are not normally diagnostic, although some can differentiate, for example, between colds and influenza (12). Thus, they will not normally aid in identifying chemicals. Performance tests make no claim to reveal effects on any brain areas, transmitter systems, or even on the nervous system itself. Performance is usually affected directly by an agent's effects on the CNS, but it can also be affected indirectly through a subject's awareness of peripheral effects, real or imagined.

One of the main advantages of performance tests is that they can measure the basic factors that compose real-life performance. Thus, they have breadth of application, i.e., they can be applied in various combinations to build a picture of any real-life task. The main disadvantage is that they cannot be applied to any one real-life task in any depth. If this is what is needed, then it would perhaps be better to use simulators or to measure performance on the job itself. However, simulators or adequate performance measures are not always available.

There is actually a small number of basic performance test designs, but there are so many variations of each that it is not possible to describe them all. Some tests have been devised as parts of test batteries or collected together into batteries, but there are very few standardized batteries such as are found in neuropsychological testing. This is because the very process of standardization tends to reduce the tests' sensitivity. The lack of standardization means that it is difficult to compare results between laboratories, but this is not normally an issue at the primary public health level. What is important here is to determine whether any changes have occurred that might be caused by exposure to a chemical, and this can be achieved by proper use of performance tests as part of a controlled design.

In very general terms, sensitivity can be increased by making the test more difficult. With some tests, this can easily be done, e.g., mathematical processing tests can contain more difficult problems to solve. With other tests, increasing difficulty can only be done by introducing other aspects of performance. For example, in vigilance or attention tests, the subject searches for a named stimulus such as a specified digit among sequences of random digits. This can be made more difficult by asking the subject to search for certain strings of digits, but this then begins to involve short-term memory.

Regarding standardization, some attempts have been made, but the resulting batteries have usually failed to achieve general acceptance, mostly on the grounds of reduced sensitivity. One notable exception is the AGARD STRES Battery (13,14) that, after a somewhat slow start, now seems to be gaining acceptance. The basic form of the battery takes 25 to 30 min to administer and could be considered a first step in the investigation of any performance effects of suspected exposure (15). Before describing performance tests, however, it is appropriate to mention some of the principles of their use.

General Procedural Aspects. Performance tests are usually of short duration, such as 3 to 5 min, although they can be as long as required. For example, some people run vigilance tests for several hours to duplicate the kind of context, such as air traffic control, in which vigilance behavior is paramount. The scores obtained are variations on, or derivations from, measures of speed and accuracy. There are two ways of testing: let the subjects complete the test and see how long they take, or impose some time limit and see how far they get. The former is the more sensitive, but is not normally possible since time is almost always limited.

The test may be administered using paper-and-pencil methods or purpose-built equipment, although most are now administered using personal computers. This increases the precision, reliability, and consistency of the test but can introduce problems of validity. The performance psychologist and the computer programmer should be aware of this. Performance tests, because of their sheer number and variety, do not have norms and must be compared with some standard or control. This can be achieved in two main ways. One method is to compare scores with those obtained earlier from the same subjects. The advantage is that the subjects serve as their own controls; the disadvantage is that they might have changed in various ways, other than having been exposed, between the two tests. Thus, any difference in test results could be due to a variety of causes other than exposure to chemicals. Because exposure is not normally anticipated, preexposure control results are not normally available.

The other method is to compare test subjects' scores with those of another group of people identical in all respects except that they have not been exposed. This sounds easy, but it is actually very difficult to achieve. For example, if all the workers in an area or factory might have been exposed, then no control group from the same area or factory might be available. Here, the investigator must find the best comparison possible from other areas, and often the best is not very good. The control group may differ from the test group in many ways, including their inherent abilities and levels of knowledge, experience, and skill. For example, if the test group naturally happens to be worse than the control group at one or more of the tests, then its

scores will be worse and could be interpreted erroneously as an effect of chemical exposure. Despite this great disadvantage, this form of control is the one that most often has to be used, and it is important that very great care is taken to match the test and control groups as carefully as possible.

When such control groups are used in experiments, it is vital that neither the subjects nor the experimenters know which group has been given which treatment. This is called a double blind procedure and ensures that there is no bias. Double blind procedures are not possible in cases of suspected exposure to chemicals, but it is important that the investigators remain unaware of the groups' identities as far as possible so that they remain free from any bias.

Learning is a considerable problem with performance tests. It will occur when any test has to be completed more than once, e.g., to check the progress of an illness, or when subjects' test scores are compared with previously obtained scores. Performance will improve with practice, which could mask any performance impairment, and lead an investigator to the erroneous conclusion that all is well. Learning cannot be overcome, but it can be minimized by proper test design, prior training, and proper study design.

With regard to test design, subjects can and do learn the items on the test. Thus, when they have to repeat a test, their results will reflect less of the function that the test is supposed to measure and more their memory of the items on previous tests. With psychomotor tests such as reaction time, tracking, or manual dexterity, this must be tolerated. With cognitive tests such as mathematical, verbal, or spatial processing, the items are often randomized or pseudorandomized to produce the same, or at least similar, degrees of difficulty.

Memory tests suffer from particular learning problems, which is not surprising since they are designed to measure learning. An example of the sort of problem that can arise is with word lists. Subjects commonly offer words they learned on one test when asked to recall words presented in later tests. This can be a serious problem when tests are repeated frequently, e.g., to monitor the time–course of effect.

With regard to training, subjects should be given instructions on how to do the test and at least allowed to practice until they are sure of what to do. Repeated training is advisable, up to even as many as four trials, until a performance plateau has

been established. When not enough time is available for extended training, the effects of practice should be analyzed.

With regard to design, the performance of test subjects must be compared with that of control subjects after the same amounts of practice, i.e., first test with first test, *n*th test with *n*th test. A drop in performance can be interpreted as evidence of impairment, given that the design criteria have been satisfied. The lack of an expected improvement in performance is sometimes interpreted as impaired learning, but cautiously, since absence of evidence is not evidence of absence.

Classification. Performance tests are based on four main theories or models of performance: factor analysis, general information processing, multiple resource/resource strategy, and stage processing. In practice, tests based on factor analysis and resource models are very similar in appearance. They tend to be phenomenological in that they measure various skills that humans exhibit, e.g., reaction time, verbal ability, and tracking. Processing stage tests take a more functional approach—that of dissecting the processing stages that occur in all types of performance, e.g., detection, discrimination, recognition, identification, decision, response selection and response execution. Phenomenological tests can indicate what types of performance are affected; processing stage tests can show which stages are affected. Most performance psychologists use both types of tests, but phenomenological ones predominate.

Several taxonomies have been proposed for phenomenological tests; one of the simplest is into sensory, cognitive, and motor functions. One of the most practical taxonomies is in terms of seven functional areas: *a)* attention (detection of rapidly or frequently occurring events), *b)* vigilance (detection of infrequent or uncertain events), *c)* simple information processing such as coding, *d)* complex information processing such as logical reasoning and spatial reasoning, *e)* memory, *f)* simple psychomotor skills such as tapping, aiming, or simple reaction time, and *g)* complex psychomotor skills such as manual dexterity or tracking. Sometimes sensation is considered a separate area as are psychophysical tests (e.g., flicker fusion, EEGs), although these might be considered to lie more in the province of neurology.

Regarding test nomenclature, a point to remember is that tests may not always measure what they purport to measure. The name of a test reflects what the designer

or user thinks is the function measured, but all tests involve all functions to some degree. For example, a memory test involves not only memory but also perception and motor functions, and all tests involve working or short-term memory. These contaminating functions may be minimal but they are still there, and some tests may be better measures of the contaminating function than of the function they claim to measure.

The main types of performance tests, with some of their main variations, are described below. There is some overlap with neuropsychological tests since they cover similar functional domains and, in some cases, the same tests are used in both neuropsychology and performance fields. All of the tests have been used successfully to study the effects of various stressors, mainly drugs.

Attention/Vigilance. These are considered together since, in practical terms, they differ only with respect to the frequency of stimulus presentation. All of the tests present signals embedded in noise for the subject to detect. The number, frequency, and complexity of the signals can be varied with the amount, type, and degree of similarity of the noise. These parameters are often manipulated to vary the difficulty or sensitivity or to match more closely a particular real life skill that might be at issue.

Attention tests with frequent stimuli tend to last 3 to 5 min. Vigilance tests can last longer if the stimuli are so infrequent that more time is needed to present enough target stimuli to provide a realistic assessment.

One of the simplest tests is letter cancellation in which subjects are presented with sheets of paper full of random letters and they have to strike out certain letters. The difficulty can be varied, for example, by altering the size, font, or number of letters or the number of targets versus the total number.

There are several versions presented by computer. Most present sequences of letters or digits, and subjects have to detect given targets. Sensitivity may be varied by changing the rate of presentation and the complexity of the targets. For example, rates of presentation can vary from one every few seconds to two or more per second; targets may be single alphanumeric characters or groups of characters. One of the most difficult and sensitive versions of the test presents digits at a rate of 100 per min, and the targets are triads of odd or even digits.

Mathematical Processing. The ability to perform simple arithmetic has been identified as a discrete factor in factor analytical studies (*16*), and mathematical processing tests have been used to study the effects of several drugs and effects of exposure to methyl chloride (*17*).

Mathematical processing tests vary in their complexity and sensitivity. One of the first of these tests was the paper and pencil Number Facility (NF) test (*16*). It consisted of 90 questions, each consisting of three one- or two-digit numbers. Subjects had to complete as many questions as possible in 3 min and write the answers in boxes. The test was standardized on U.S. servicemen and was widely used throughout the 1960s as a sensitive, reliable, and valid measure of mathematical processing. Twenty equivalent forms were produced for repetitive testing, although now the test may be administered using computers, which can produce as many equivalent forms as needed or generate items as required.

The NF test proved sensitive but was fairly difficult to perform, so some people were not able to complete many questions in the 3 min allowed. Thus, most subsequent mathematical processing tests have used addition or subtraction of single digits. The test used in the AGARD STRES battery, for example, presents three single digits with two operators and requires the subject to say whether the answer is greater than or less than 5.

Verbal Processing. Verbal processing is considered by some to measure the same area of function as mathematical processing; however, the two functions are intuitively discrete and they can be affected differentially. Several verbal processing tests have been reported in the field of experimental psychology, but the most widely adopted as a performance test is Baddeley's Grammatical Reasoning test (*18*). The test consists of several sentences, each followed by a pair of letters—AB or BA. The sentence describes which letter comes first, and the subject has to say whether the description is true or false. Examples include "A follows B—AB", "B does not follow A—BA", "A is preceded by B—BA".

The test is related to the ability to comprehend the structure and syntax of English and has proved sensitive to a variety of environmental stressors. However, its main disadvantage is that it is restricted to English. Attempts to translate to other languages have met with varied success, e.g., German rarely uses the passive voice. The AGARD STRES version of this test tried to

remedy this defect by increasing the number and complexity of the comparisons that had to be made, but the sensitivity of this version has yet to be fully assessed.

Spatial Processing. Spatial processing, like mathematical processing, has been identified as a discrete factor in factor analytic studies, and a variety of tests have been reported to assess various subfactors such as spatial relations, spatial orientation, and visualization. Some of these have found use as performance tests. They all use spatial or graphic items that are impossible, or at least very difficult, to verbalize.

One is the manikin test in which a stylized picture of a human figure is shown holding an object in one hand. The figure may be presented at any angle, forward or reversed, and the subject has to say which hand is holding the object.

Another is the Shephard and Metzler block test (*19*), consisting of pairs of two-dimensional representations of shapes produced by putting together eight cubes. In some pairs, the block shapes are the same, but one of the pair is rotated; in other pairs, one block shape is the mirror image of the other. Subjects have to say whether the shapes are the same.

A third is the Histogram test (*20*). Here, pairs of histograms are presented one histogram at a time. The first histogram is normally presented upright, and the second is rotated normally through 90 or 270 degrees. Subjects have to say whether the second histogram is the same as the first. This test has the advantage that the number of bars and their lengths can be varied to change the level of difficulty.

Memory. There are perhaps more memory tests than all other performance tests combined. This is because memory is a logical part of every aspect of human performance. Most memory tests assess short-term memory, and most are unsuited for repetitive testing because what the subject learns on one test interferes with recall on subsequent tests. Sometimes this phenomenon can be used to advantage, e.g., to study perseveration, but usually it seriously impairs the sensitivity of the test. Two tests that can be used repetitively are Wechsler's digit span test and Sternberg's memory search test.

Wechsler's digit span test is one of the most widely used short-term memory tests, perhaps because it is easy to use, it can be used repetitively, and it provides measures of two component memory skills labeled as rote recall and mental manipulation. Subjects are presented with sequences of

digits that they have to recall and report in the same order immediately, and the sequences include more and more digits until the subject fails. Subjects are usually allowed another attempt to minimize the effects of distractions, and the digit span is taken as the longest sequence of digits that can be recalled successfully. Typically, healthy subjects can recall 6 to 8 digits. The test is usually repeated with subjects recalling the digits in reverse order. This manipulation varies the cognitive workload while keeping the memory load constant. In the somewhat unusual circumstance in which subjects recall more digits backwards than forwards, a heightened arousal or motivation is postulated for the more complex material. Although the digit span test is widely used, the literature suggests that it is of variable sensitivity (*21*).

The Sternberg test presents subjects with a set of items (usually digits or letters) called the memory set followed by a single probe item, and their responses indicate whether the probe is a member of the memory set. The test can differentiate memory searching strategies, and the stages can be manipulated to vary their difficulty. For example, memory searching can be affected by varying the size of the memory set, detection by varying the figure-ground contrast of the probe items, and recognition by varying the clarity of the probe items. The test has proved to be quite sensitive and is being used increasingly, particularly in psychopharmacology. The main disadvantage is that the more stages that are covered the longer the test takes. For this reason, many researchers use it simply as a short-term memory test, which seems a waste of its potential.

Simple Psychomotor Skills. Examples of simple psychomotor tests include finger tapping, aiming, simple and choice reaction time, continuous reaction time (each response triggers the next stimulus (*22*), and unidimemsional tracking.

These tests are simple in principle, but they vary widely in their style of administration such that it is rare to find two laboratories with the same version. Despite this, the tests are generally sensitive and easy to administer.

Complex Psychomotor Skills. Complex psychomotor skills are measured by manual dexterity tests such as the O'Connor fine finger dexterity test. For this test, subjects pick up three small pins at a time from a tray, using only one hand, and place them in small holes. This skill has been identified as a discrete factor in factor analysis studies,

and the test has proved sensitive to a range of drug effects. Other tests under this heading include two-dimensional tracking, which is generally of three types: pursuit, in which the subject pursues a moving target; compensatory, where the target is stationary and the tracking device drifts; or combined compensatory/pursuit. Various refinements have been made, e.g., where the evasive movements of the target increase as the cursor gets closer. Generally, sensitivity increases as the tracking difficulty increases. One of the most sensitive is the unstable tracking test in which the subject has to control a cursor that tends to accelerate away from a target. This test originated from analyses of aircraft handling and is well-founded in human engineering theory.

Multitasking. Multitasking has proved useful in experimental psychology and performance work, and might also prove useful in the assessment of neurotoxic chemicals. Multitasking is simply performing two (or more) tasks concurrently. The tasks may be chosen on the basis of a particular real-life application (e.g., vigilance and tracking are often used) or to investigate or stretch reserve capacity, resource allocation, or time sharing functions. For this purpose, cognitive tests are sometimes combined with motor tests such as tapping a finger at a nominal rate of once per second. Variations in the rate of tapping can be used to reflect variations in performance load.

Neuropsychological Testing

At least 250 different tests have been used to evaluate the effects of neurotoxicants on behavior (21). Thus, when the researcher or clinician is confronted with the task of selecting tests to evaluate reported symptoms, no single test or battery of tests has been validated for characterization of dysfunction due to neurotoxicants. Before selecting a test battery for assaying a suspected exposure, the relevant literature on that particular agent or chemical class should be consulted. Specific guidance for tests to be used with the more frequent sources of exposure can be found in publications such as White et al. (23). Although potentially useful in the evaluation of gross impairment, traditional neuropsychological tests are less suited to characterize subtle cognitive dysfunction since they often provide summary scores that are insensitive to the nuances of performance on the test. For example, a total score is given for block design from the WAIS-R (24), a test frequently used to characterize the effects of exposure. Speed of performance, motor

coordination, and visuospatial skills are necessary to earn a high score. It is impossible from this test score, however, to quantify which aspect of performance is impaired. Likewise, performance-based tests from cognitive experimental psychology cannot offer a complete characterization of behavioral dysfunction arising from brain injury. Thus, selection of tests from each tradition depends on the purpose of the evaluation and the exposure situation.

Some studies have attempted to suggest patterns of performance associated with particular agents (e.g., lead vs solvents (4); however, these patterns have not been well established. Therefore, while it would be desirable to define specific batteries of tests suited for identified exposure situations, the knowledge base does not allow that level of specificity. What is clear from the existing literature is that, to adequately characterize neurobehavioral effects of a neurotoxic exposure, tests from each of the domains listed in Table 1 should be selected. This conclusion is consistent with the World Health Organization's proposed standardized screening battery of neurobehavioral tests (26). Specific tests wax and wane in their popularity, but the domains to be represented are relatively consistent across studies and among clinical laboratories. Fiedler (3) provides a description of the functional domains to be assessed and representative neuropsychological tests for each category.

Use of a test battery that includes tests in each domain (e.g., the WHO battery) will provide an adequate initial characterization of behavioral effects when a neurotoxic exposure of sufficient magnitude has occurred. For example, if workers from a factory have used organic solvents routinely over several years and have symptomatic complaints, application of these batteries to characterize the behavioral effects is advised. The literature is replete with examples of the utility of these tests for characterizing neurobehavioral effects (27). Of course, if these tests are applied in different cultures or languages, then consideration must be given to the impact of these modifications on the test results. For example, normative values generated from one culture or country may not be applicable to a different culture.

The following is a discussion of the tests commonly used to evaluate each of the functional domains. Table 1 lists relevant strengths and weaknesses of frequently used tests from each functional category.

Evaluation of Sensory Function. While the neurologic examination involves

evaluation of sensory function, such an examination may not be possible in every situation. Therefore, before administering a battery of neurobehavioral tests, it is important to determine that basic sensory processes, particularly of vision and audition, are intact. Most of these tests require, at the very least, intact visual and auditory function, and tests of motor speed and visuomotor skills require intact somatosensory function. For example, blue collar workers in the construction trades or in farming may have hearing impairment due to noise or tactile sensory imperception due to injuries to the hands. Clearly, these mechanical problems would account for poor performance on some neurobehavioral tests. Therefore, it is important to know about these difficulties so that cognitive impairment is not inferred inappropriately. Basic tests of visual acuity and hearing as well as simple tests of tactile perception should be performed before the initiation of a neurobehavioral examination.

Overall Cognitive Ability. Unfortunately, in most exposure situations, a standardized indicator of preexposure cognitive function is not available. Performance on all neurobehavioral tests is influenced by the individual's overall intellectual ability. To interpret results from group studies or from an individual evaluation, an estimate of preexposure ability must be obtained. In some situations, achievement test scores may be available from military or school records and should be obtained. If different tests are used from one individual to the next in a group, these scores may be converted to the same scale (e.g., T score) to allow rough comparisons between groups. Educational level has also been used as a rough surrogate of overall cognitive ability. In the absence of an actual test score documenting preexposure ability, many investigators and clinicians use standardized tests of verbal skills to estimate ability. This strategy rests on the assumption that neurotoxicants do not reduce performance on indicators of well-learned information such as vocabulary or reading ability.

The vocabulary test from the WAIS-R (24) requires that the individual define words in a free recall situation. Other tests of vocabulary, such as the Shipley (28) are given in a multiple choice format, which reduces the verbal expressive demands on the individual. Another frequently used group of tests to assess ability are tests of reading such as the National Adult Reading Test-Revised (29) or the Wide Range Achievement Test-Revised, Reading Subtest

Table 1.

Tests	Strengths	Weaknesses
Premorbid estimators of intelligence		
Vocabulary subtest of WAIS-R	Normative values by age and education (25); sample representative of U.S. population	Approximately 15 min to administer; requires expressive language function
NART-R	Simple and quick to administer; minimal expressive language demands	Assumes moderate vocabulary development
Reading subtest of Wide Range Achievement Test-revised	Simple and quick to administer; easy to self-administer; minimal expressive language demands; starts at reading level of individual	Assumes lower vocabulary development
Attention and concentration		
Digit Span Subtest of WAIS-R	Normative values by age and education (25); sample representative of U.S. population; 5 min to administer; two aspects of attention assessed	Auditory acuity required
Bourdon-Wiersma Vigilance Test	Quick-to-administer paper and pencil test	Normative values not available for U.S. population
Grooved pegboard	Normative values by age and education (25); 5 min to administer; minimal equipment; suitable for repeated measures	Assumes intact fingers, fingertips, and vision
Finger tapping	Normative values by age and education (25); quick to administer; minimal equipment	Assumes intact fingers and fingertips
Santa Ana Dexterity Test	Quick to administer; minimal equipment	Norms not available for U.S. population; assumes intact fingers, fingertips, and vision
Visuospatial relations		
Block Design Subtest of WAIS-R	Nonverbal and considered less culture bound; normative values by age and education (25)	Timed test; assumes intact vision and intact fingers; administration time = 10 to 15 min; summary score incorporates several aspects of performance
Raven's Progressive Matrices	Nonverbal and considered less culture bound; untimed	Assumes intact vision; administration time = 25 min; score conversion to percentiles
Visuomotor coordination		
Digit Symbol Subtest of WAIS-R	Normative values by age and education (25); quick administration—2 min	Less sensitive for lower SES individuals; assumes intact fingers and motor coordination
Trails A and B	Normative values by age and education (25); administration time = 5 to 10 min	Simple task of immediate memory; requires 30 min to assess memory after delay; auditory acuity required
Memory		
California Verbal Learning Test (CVLT)	Comprehensive assessment of verbal memory parameters	Norms inapplicable to less than college-educated groups; administration time = 20 min for immediate and requires 30 min for delayed memory assessment; auditory acuity required
Paired Associates Subtest of WMS-R	Norms for representative sample of U.S. population; administration time = 10 min	Simple task of immediate memory; requires 30 min to assess memory after a delay; auditory acuity required
Logical Memory Subtest of WMS-R	Norms for representative sample of U.S. population; administration time = 10 min	Requires 30 min to assess memory after a delay; auditory acuity required
Rey-Osterrieth Complex Figure Test	Complex visual stimulus; nonverbal test	Scoring criteria and norms available from individual investigators; delayed recall requires 30 min
Benton Visual Retention	Administration time = 5 to 10 mins; nonverbal test	Simplistic figures; memory assessment confounded by motor movement; norms within context of age and IQ level
Visual Reproduction Subtest of WMS-R	Norms for representative sample of U.S. population; administration time = 10 min; nonverbal test	Simplistic figures; requires 30 min to assess memory after a delay; memory assessment confounded by motor movement
Affect/Personality		
Profile of Mood States (POMS)	Covers broad range of moods; brief to administer; easy to self-administer	Psychiatric outpatient and college student normative sample
MMPI-2	Used extensively with many populations; norms updated and improved from original sample; self-administered	Lengthy—up to 2 hr to complete; some items may be objectionable to nonclinical samples; requires 8th grade reading level

(30). These tests require that the individual pronounce a series of words of increasing difficulty. Individuals are given credit if the words are pronounced correctly. All of these tests are heavily dependent on language abilities. Therefore, their use in other cultures must be adapted accordingly.

Attention/Concentration. The ability to orient to a stimulus and sustain attention is the precursor to successful performance on most neurobehavioral tasks; therefore, tests to assess this function must be included. Otherwise, a deficit on another test such as a test of memory may be misinterpreted as a primary memory dysfunction when an inability to sustain attention is the primary deficit. Digit span from the WAIS-R (24) is a widely used test of auditory attention in which the individual is asked to repeat an increasing string of digits presented by an examiner.

Instructions are given to repeat the digits as they are presented and to reverse them. The Bourdon-Wiersma (*31*) is another test widely used in the Scandinavian literature to assess vigilance. In this paper and pencil test, the individual must cross out each series of four dots interspersed among a page full of other dot configurations.

Motor Skills. The standard neurological examination includes a clinical evaluation of motor skills. However, neurobehavioral tests provide a relatively more standardized assessment of these skills which can be related to normative values. Grooved pegboard (*32*) is a simple test of fine motor coordination in which the individual places grooved pegs in grooved holes with the dominant and nondominant hand while being timed. This task requires some fine manipulation of the pegs to fit correctly into the grooves. Finger tapping (*33*) is another simple task of motor speed in which the individual taps as quickly as possible with the index finger of the dominant and nondominant hands. The number of taps is recorded with a mechanical counter. Although a version of this test is available on computer (NES2) (*34*), the manual version requires minimal equipment and normative values are available for its administration (*25*). Finally, the Santa Ana (*35*) is a test applied frequently in Scandinavia that also involves placing pegs in holes on a board while being timed.

Visuomotor Coordination. Digit symbol from the WAIS-R (*24*) is probably the most widely used test of visuomotor coordination and speed within the literature on neurobehavioral assessment of neurotoxicants. In this task, the individual is asked to record symbols associated with digits from a key that is present throughout the task. The individual is given 90 sec and the number coded correctly during this time period is the score. A version requiring only an oral response is available for motor-impaired individuals. Trials A and B (*33*) are also widely used tasks involving visuomotor coordination and speed. The instructions are to connect numbers in sequence (Trails A) or shift between numbers and letters in sequence (Trails B) while being timed. Speed of performance is measured, and mistakes add to the time required to complete the task.

Visuospatial Relations. Tests of visuospatial ability may serve as indicators of overall ability but are not recommended when exposure is suspected because they have also been sensitive to the effects of neurotoxicants. For example, Raven's

Progressive Matrices (*36*) is a test of visuospatial problem solving for which minimal verbal skills are required. Block design from the WAIS-R has also been extensively used. This task involves putting blocks together to mimic a design while being timed. Additional points are given for quick performance, but a design must be completely correct and performed within the time limit to receive full credit. Thus, speed as well as visuospatial ability contribute to performance.

Memory. Many tests of memory are available. Among those most frequently applied in the field of neurobehavioral assessment are subtests from the Wechsler Memory Scale-Revised (WMS-R) (*37*). The Paired Associates test from the WMS-R involves the verbal presentation of four easy and four difficult word pairs over six separate trials. The correct answers are summed over the first three trials. The purpose of this task is to evaluate the ability of the individual to encode verbal information. Delayed recall is also evaluated after a 30-min delay. This task is relatively simple compared to the California Verbal Learning Test (CVLT) (*38*). The latter involves learning a list of 16 common shopping items presented verbally over five consecutive trials. Delayed recall is evaluated after a 20-min latency by providing a recognition trial. The precursor to the CVLT, i.e., the Rey Auditory Verbal Learning Test (*39*), has been used on several occasions to evaluate the effects of neurotoxicants. In addition to learning efficiency over the trials, the CVLT also evaluates the strategies used to encode the word list, the effects of an interfering list on subsequent recall of the original list, and the effects of a 30 min delay on recall. Also, recognition of the list is evaluated separately from free recall. Thus, in one test, many parameters of verbal learning are evaluated.

It is important in any evaluation of memory to include memory for visual as well as verbal material. Tests of visual memory often involve abstract figures as stimuli that cannot be easily encoded verbally. For example, abstract figures are presented for 10 sec in the visual reproduction test of the WMS-R. The individual is then asked to draw them from immediate memory. Standardized scoring procedures are used to evaluate the performance. Similarly, in the Benton Visual Retention Test (*40*), abstract figures are presented, reproduced from memory, and scored with standardized procedures. Individuals are also asked to draw the figures from memory

after a 30-min delay. The Rey-Osterreith Complex Figure Test (*41*) follows a similar procedure except that the the figure is much more complex.

Affect/Personality. Many checklists and symptom rating scales are available to evaluate subjective ratings of mood. As stated previously, alterations in mood are often one of the first indicators of the effects of neurotoxicants. In selecting a measure of mood, it is important to evaluate the range of affect including symptoms of anxiety, irritability, and depression. The Profile of Mood States (*42*) and the Minnesota Multiphasic Personality Inventory-2 (*43*) have been used frequently to evaluate the mood changes in response to neurotoxicants.

Computerized Test Batteries. More recently, tests from the neuropsychological tradition and cognitive experimental or performance testing have been computerized for use in the assessment of the effects of neurotoxicants. Batteries such as NES2 (*34*) and the Milan Automated Battery (*44*) grew out of the need to conduct studies that could be conducted efficiently in the field. Also, these batteries have been some of the first to bring performance testing together with traditional neuropsychological tests for these evaluations. For example, NES2 includes reaction time and continuous performance as a part of the battery, along with computerized versions of digit symbol and finger tapping. It is important to remember that adaptation of the neuropsychological tests for the computer makes the test fundamentally different so that the normative data collected on the original tests are not applicable. In keeping with the tradition of experimental psychologists, the test parameters within the software are adjustable so that the duration and difficulty of each component test can be altered by the administrator. Therefore, it is critical to publish the parameters of test administration along with the results obtained. Experimental psychologists may feel constrained by the clinical standardization and the limiting of response modalities to a keyboard, mouse, or joystick. However, for comparison of groups, these batteries offer tests from both traditions that are easily administered and accessible.

Some general advantages of all neuropsychologic tests listed below are that these tests have normative values. As for the use of any test, good practice demands that care be taken to insure demographic comparability of the individual(s) tested and the normative values provided for the test. When normative values are clearly not

representative of the general U.S. population, this is listed as a weakness for the specific test. Another advantage of the tests listed is that standard instructions for test administration are provided in the test manuals. This reduces the variability due to differences in examiners. A universal disadvantage of these tests is that they must be given by an examiner with some experience and training in test administration. This increases the time involved and cost of administering a battery of these tests. However, the equipment required to give these tests is inexpensive, readily available, and quite portable. Thus, the primary cost of testing is for personnel. Finally, while many of these tests may be administered by a well-trained technician, interpretation of the results from an individual or group requires the expertise of a professional trained in the use of these tests and their applications.

Developmental Assessment

Research on numerous substances, including lead, alcohol, methylmercury, and polychlorinated biphenyls (PCBs) indicates heightened susceptibility of infants and children to neurotoxicity, particularly when exposure occurs early in the course of development. By contrast to the effects of acute adult exposures, which are frequently transitory, the effects of exposure during development are often more persistent. Moreover, effects of *in utero* exposure on the CNS often do not become evident for several years, that is, when the affected cognitive or behavioral system matures. In the case of industrial accidents or other acute exposures, particular attention needs to be given to effects on offspring of women exposed during pregnancy.

Much of our knowledge of the effects of neurotoxic exposure on development comes from prospective, longitudinal studies in which exposed infants are recruited prenatally or immediately after birth and assessed over the course of development. Prenatally, vulnerability of a particular brain structure or region may be heightened, particularly when exposure occurs during a period of rapid cell division or cell migration. After delivery, the blood–brain barrier and a more highly developed drug-metabolizing capacity may provide protection not available *in utero*. Because brain development and mylenation continue for several months after delivery, there may also be heightened vulnerability during infancy. Given the unique vulnerability during early development, the cognitive and behavioral

deficits seen in adults exposed to a particular chemical may provide little indication of the types of neurobehavioral impairment that might be expected in infants and children exposed during development.

The test used most extensively to evaluate neurobehavioral function in the newborn is the Brazelton (45) Neonatal Behavioral Assessment Scale (NBAS). The NBAS, a 30-min examination procedure, assesses 17 reflexes and a range of behaviors including muscle tone, activity level, attention and orientation, and arousal (46). Although sensitive to many prenatal chemical exposures, including obstetrical medication, opiates, and PCBs (46,47), effects seen on the NBAS are often very transitory, and the test is not predictive of neurobehavioral function during later development. No norms are available.

The test used most frequently to evaluate neurobehavioral function during infancy is the Bayley Scales of Infant Development. The Bayley, which provides both a Mental Development Index and a Psychomotor Development Index, focuses primarily on the rate at which the infant attains age-appropriate developmental skills. Standardized norms are available for both the original Bayley and the recently released Bayley II. The Bayley has proven sensitive to a broad range of prenatal exposures, including alcohol (48,49), lead (50,51), methadone (52), and PCBs (53,54). Although predictive validity for school-age cognitive function is poor for children who perform within the normal range (55), most neurotoxic exposures detected by the Bayley are also associated with poorer cognitive performance at school age. Thus, the Bayley may be sufficiently sensitive to detect group differences associated with neurotoxic exposure even if it is not sufficiently reliable to predict for individual children.

The Bayley is an apical test; successful performance on a single item usually depends on the integrity of multiple elements of cognitive and fine motor function, as well as attention to the task and motivation to perform. The principal advantage of apical tests is sensitivity. Because the infant's performance on a given item can be affected by deficits in any of several domains, the Bayley is sensitive to a broad range of impairments. The principal weakness of an apical test is lack of specificity; little information is provided about which aspects of cognitive function have been compromised. The Bayley grew out of a maturationist tradition (56,57)

that regarded infant development as a series of milestones programmed to emerge over time. As a result, at each age only those domains in which new behaviors are emerging are assessed in detail. Nevertheless, given the Bayley's sensitivity, it is probably advisable to include it in any assessment of the effects of a previously unstudied exposure.

An alternative approach to infant neurobehavioral assessment is provided by certain newer tests of infant cognitive function exemplified by the Fagan visual recognition memory test (57). In the Fagan test, infant patterns of visual fixation to familiar and novel stimuli are used to assess recognition memory and visual discrimination, two processes that are fundamental to intellectual function during childhood and adulthood. By contrast to an apical test like the Bayley, the Fagan takes a "narrow band" approach and will provide an indication of neurobehavioral deficit only if there is a neurotoxic effect on one of the specific domains of function that it assesses. Nevertheless, the Fagan has been found to be sensitive to prenatal exposure to PCBs in human infants (58), to methylmercury in rhesus monkeys (59), and, when scored in terms of information processing speed, to prenatal exposure to alcohol as well (60). By contrast to the Bayley, the Fagan has been found to be moderately predictive of intellectual function during childhood (61).

Visual acuity can be assessed during infancy by means of a new procedure developed by Teller et al. (62), which is based on infant visual fixation to vertical lines displayed at different horizontal distances on a screen. In principle, it is similar to the Fagan test because it measures differences between the time spent gazing at the vertical pattern compared to that spent gazing at a blank target. The density of the patterned target provides scores related to what vision scientists term spatial contrast sensitivity. It describes the ability of the visual system to distinguish darker from lighter arrays by specifying the width (visual angle) of the array and the depth of the dark-light difference, or contrast.

Many of the neuropsychological and performance tests originally designed for adults are also available for use with children. Several IQ tests are available, including the McCarthy Scales of Children's Abilities and the WPPSI (Wechsler Preschool and Primary Scale of Intelligence - Revised) for the preschool period; the Wechsler Intelligence Scale for Children (WISC)-III for school-age children;

and the Stanford-Binet and Kaufman Assessment Battery for Children, which cover both periods. The principal advantages of these tests include standardized norms and coverage of a broad range of domains of function. Examination of effects on subtest scores, which are frequently also normed, can provide information regarding effects on specific domains of function. Although there is evidence suggesting that IQ tests may be culturally biased, they are predictive of success in school and therefore provide a valid indicator of that important domain of intellectual performance during childhood. Childhood IQ tests have been found to be sensitive to prenatal or early childhood exposure to lead (51,63), alcohol (64,65), and PCBs (66). Effects on school achievement can also be assessed directly by administering standardized school achievement tests in such domains as reading, math, spelling, etc.

Neuropsychological tests available for use with children include the grooved pegboard and the Wisconsin card sorting tests. One of the most frequently used performance tests is the continuous performance test (CPT), which has repeatedly been found to be sensitive to the effects of prenatal exposure to alcohol (67–70). The Sternberg memory search test can easily be used with school-age children, and a mental rotation test designed by Kail (71) is also available, which measures RT in discriminating between mirror images of rotated letters to assess mental manipulation of visual images.

One very important potential confounding influence in neurobehavioral assessment during childhood is quality of intellectual stimulation and emotional support provided by parents. In some studies, chemical exposures have been found to be so confounded with such influences that it was not possible to evaluate their effects on intellectual development (5). One important advantage of performance tests over IQ and school achievement is that they are markedly less affected by socioenvironmental influences (5). Infant tests administered during the first year are also much less likely to be confounded with social environment (5,72).

One of the principal advantages of a prospective longitudinal approach for evaluating developmental toxicity is that it provides an opportunity to assess exposure during the potentially vulnerable prenatal and infant periods. Retrospective studies are feasible, however, where the chemical remains have been deposited in the child's tissue over an extended period of time. Needleman et al.'s (73) classic retrospective study demonstrating lead neurotoxicity based on levels of lead deposited in deciduous teeth provides a relevant example. In some cases medical records can be used to document early exposure, as in the 1979 PCB poisoning in Taiwan, where all exposed individuals were enrolled in a government registry (5). For some exposures, such as maternal smoking during pregnancy, parental recall can provide reliable information, although recall for many prenatal exposures will not be accurate (5).

Summary and Recommendations

As demonstrated in this chapter, a wide variety of procedures are available for the assessment of human neurobehavioral function. The validity of performance tests for evaluating the neurobehavioral effects of drugs is well established, but these types of tests have been used to only a limited degree for assessing the neurotoxic effects of chemical exposures in the clinic. As performance tests are incorporated into studies of neurotoxicity induced by chemical exposures, it would be helpful to establish a database. This database would provide information regarding their sensitivity in this context to help investigators select from the many tests when undertaking neurotoxicity studies. Although traditional neuropsychological tests have been used extensively in clinical assessments of chemical exposures, there is no centralized source of information about their validity in these studies. A database covering the use of both types of assessments in this context would therefore be very useful.

Over time, such a database would also be useful for investigators interested in conducting metaanalyses to integrate data from diverse studies using similar measures to investigate the neurotoxicity of a given chemical exposure. There are probably already sufficient data for such an analysis in the neurotoxicity of organic solvents. As indicated by several of the papers in this volume, more information is also needed about the validity of repeated assessments using these measures to track recovery or deterioration during the period after an exposure has occurred.

One issue that warrants increased attention is individual differences in vulnerability to exposures. We have already reviewed some of the evidence indicating the heightened vulnerability found when exposure occurs *in utero* or during infancy, and it has been suggested that vulnerability may also be increased by the process of aging or when individuals are under stress (74). Drug studies have demonstrated individual differences in vulnerability in women that are related to variations in hormone levels. Such differences might be expected for neurotoxic chemicals as well, particularly those like PCBs that are known to affect hormone levels. Evidence of multiple chemical sensitivity reviewed by Fiedler (3) also indicates the importance of increased attention to individual differences in susceptibility.

Finally, although assessment of neurobehavioral outcome has been the principal focus of this chapter, it is critical to recognize the importance of obtaining as concurrent and reliable an assessment of exposure as possible. For studies of substances such as lead and PCBs, which leave detectible biological residues, this would mean obtaining blood or other tissue samples as close to the time of exposure as possible. For others, detailed documentation of extent of exposure (e.g., proximity to the spill) should be recorded in an official government registry. Often government assistance can be made contingent on registration. Concurrent measures of exposure are critical for investigating long-term consequences, which may not become evident until several years after the fact. Wherever possible, it is important to document not only the fact of exposure but also its extent to investigate dose–response relationships and lowest-dose thresholds at which neurotoxic effects first become evident.

REFERENCES

1. Waller K, Osorio AM, Maizlish N, Royce S. Lead exposure in the construction industry: results from the California Occupational Lead Registry, 1987 through 1989. Am J Public Health 82:1669–1671 (1992).

2. Weiss B. Risk assessment: the insidious nature of neurotoxicity and the aging brain. Neurotoxicology 11:305–314 (1990).

3. Fiedler N. Neuropsychological approaches for the detection and evaluation of toxic symptoms. Environ Health Perspect 104(Suppl 2):239–245 (1996).

4. Hanninen H. The psychological performance profile on occupational intoxications. Neurotoxicol Teratol 10:485–488 (1988).

5. Jacobson JL, Jacobson SW. Prospective longitudinal assessment of developmental neurotoxicity. Environ Health Perspect 104(Suppl 2):275–283 (1996).

6. Feldman RG, White RF. Role of the neurologist in hazard identification and risk assessment. Environ Health Perspect 104(Suppl 2):227–237 (1996).

7. Newland MC. Quantification of motor function in toxicology. Toxicol Lett 43:295–319 (1988).

8. Kimura J. The blink reflex. In: Electrodiagnosis in Diseases of Nerve and Muscle—Principles and Practice (Kimura J, ed). Philadelphia:F.A. Davis, 1989;323–351.

9. Jabre JF, Hackett ER. EMG Manual. Springfield, IL:Charles C. Thomas Publisher, 1983;1–126.

10. Aminoff MJ. Electromyography in Clinical Practice. New York:Churchill Livingstone, 1987.

11. Wetherell A. Performance tests. Environ Health Perspect 104(Suppl 2):247–273 (1996).

12. Prasher D, Smith A, Findley L. Sensory and cognitive event-related potentials in myalgic encephalomyelitis. J Neurol Neurosurg Psychiatry 53:247–253 (1990).

13. AGARD. Human performance assessment methods. AGARDograph 308. Neuilly-sur-Seine, France:North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development, 1989.

14. AGARD. Human performance assessment methods. AGARDograph 308 Addendum. Neuilly-sur-Seine, France: North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development, 1991.

15. Wetherell A. The STRES battery: standardized tests for research into environmental stress. In: Contemporary Ergonomics 1990 (Lovesey EJ, ed). London:Taylor & Francis, 1990;270–275.

16. Moran LJ, Mefferd RB. Repetitive psychometric measures. Psychol Rep 5:269–275 (1959).

17. Repko JD, Jones PE, Garcia LS, Schneider EJ, Roseman E, Corum CR. Behavioral and neurological effects of methyl chloride. DHHS (NIOSH) Publ No 77–125. Washington:U.S. Government Printing Office, 1976.

18. Baddeley AD. A 3-minute reasoning test based on grammatical transformation. Psychon Sci 10:34 (1968).

19. Shepard RN, Metzler J. Mental rotation of three-dimensional objects. Science 171:701–703 (1971).

20. Shingledecker CA. A task battery for applied human performance assessment research. Rpt No AFAMRL-TR-84-071. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory, 1984.

21. Anger WK. Worksite behavioral research: results, sensitive test batteries, and the transition from laboratory data to human health. Neurotoxicology 11:629–720 (1990).

22. Leonard L. 5-choice serial reaction apparatus. Rpt No 326. Cambridge, UK: Medical Research Council Applied Psychology, Unit 1959.

23. White RF, Feldman RG, Proctor SP. Neurobehavioral effects of toxic exposures. In: Clinical Syndromes in Adult Neuropsychology: The Practitioner's Handbook (White RD, ed). Amsterdam:Elsevier, 1992;1–51.

24. Wechsler D. Wechsler Adult Intelligence Scale-Revised. San Antonio, TX:The Psychological Corporation, 1981.

25. Heaton RK, Grant I, Matthews CG. Comprehensive Norms for an Expanded Halstead-Reitan battery. Odessa, FL:Psychological Assessment Resources, 1992.

26. WHO. Chronic Effects of Organic Solvents on the Central Nervous System—Core Protocol for an International Collaborative Study. Environmental Health Ser No 36. World Health Organization, 1989.

27. Anger WK, Johnson BL. Human behavioral neurotoxicology workplace and community assessments. In: Environmental and Occupational Medicine (Rom WN, ed). Boston:Little Brown and Company, 1992;573–592.

28. Zachary RA. Shipley Institute of Living Scale Revised Manual. Los Angeles:Western Psychological Services, 1986.

29. Nelson HE. National Adult Reading Test (NART): Test Manual. Windsor, UK:Nelson Publishing, 1982.

30. Jastak S, Wilkinson GS. Wide Range Achievement Test Administration Manual. Los Angeles:Western Psychological Services, 1984.

31. Hanninen H, Lindstrom K. Behavioral test battery for toxicopsychological studies used at the Institute of Occupational Health. Helsinki:Institute of Occupational Health, 1979.

32. Trites RL. Neuropsychological test manual. Montreal: Technolab, 1981.

33. Reitan RM, Wolfson D. The Halstead-Reitan Neuropsychology Test Battery. Tucson:Neuropsychology Press, 1985.

34. Letz R, Baker EL. NES2 Neurobehavioral Evaluation System. Winchester, MA:Neurobehavioral Evaluation System, 1988.

35. Fleishman EA. Dimensional analysis of psychomotor abilities. J Exp Psychol 48:437–454 (1954).

36. Raven JD. Guide to Standard Progressive Matrices. London:HK Lewis, 1960.

37. Wechsler D. Wechsler Memory Scale-Revised Manual. San Antonio, TX:The Psychological Corporation, 1987.

38. Delis DC, Kramer JH, Kaplan E, Ober BA. California Verbal Learning Test. San Antonio, TX:The Psychological Corporation, 1987.

39. Lezak MD. Neuropsychological Assessment. New York:Oxford University Press, 1983.

40. Benton AL. Revised Visual Retention Test. 4th ed. San Antonio, TX:The Psychological Corporation, 1974.

41. Visser RSH. Manual of the Complex Figure Test CFT. Amsterdam:Swets & Zeitlinger, 1985.

42. McNair DM, Lorr M, Droppleman LF. Profile of Mood States. San Diego:Educational and Industrial Testing Service, 1981.

43. Hathaway SF, McKinley JC. Minnesota Multiphasic Personality Inventory 2. Minneapolis:University of Minnesota Press, 1989.

44. Fittro KP, Bolla KI, Heller JR, Meyd CJ. The Milan Automated Neurobehavioral System. J Occup Med 34:918–922 (1992).

45. Brazelton TB. Neonatal Behavioral Assessment Scale. Clinics in Developmental Medicine, No 50. Philadelphia:JB Lippincott, 1973.

46. Jacobson JL, Fein GG, Jacobson SW, Schwartz PM. Factors and clusters for the Brazelton Scale: an investigation of the dimensions of neonatal behavior. Dev Psychol 20:339–353 (1984).

47. Rogan WJ, Gladen BC, McKinney JD, Carreras N, Hardy P, Thullen J, Tinglestad J, Tully M. Neonatal effects of transplacental exposure to PCBs and DDE. J Pediatr 109:335–341 (1986).

48. Streissguth AP, Barr HM, Martin DC, Herman CS. Effects of maternal alcohol, nicotine, and caffeine use during pregnancy on infant mental and motor development at 8 months. Alcohol Clin Exp Res 4:152–164 (1980).

49. O'Connor MJ, Brill NJ, Sigman M. Alcohol use in primiparous women older than 30 years of age: relation to infant development. Pediatrics 78:444–450 (1986).

50. Bellinger DC, Needleman HL, Leviton A, Waternaux C, Rabinowitz MB, Nichols ML. Early sensory-motor development and prenatal exposure to lead. Neurobehav Toxicol Teratol 6:387–402 (1984).

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

52. Hans SL, Henson LG, Jeremy RJ. The development of infants exposed in utero to opioid drugs. In: Longitudinal Studies of Children at Psychological Risk: Cross-National Perspectives (Greenbaum CW, Auerback JG, eds). Norwood, NJ:Ablex, 1992.

53. Gladen BC, Rogan WJ, Hardy P, Thullen J, Tingelstad J, Tully M. Development after exposure to polychlorinated biphenyls and dichlorodiphenyl dichloroethene transplacentally

and through human milk. J Pediatr 113:991–995 (1988).

54. Rogan WJ, Gladen BC. PCBs, DDE, and child development at 18 and 24 months. Ann Epidemiol 1:409–413 (1991).

55. McCall RB, Hogarty PS, Hurlburt N. Transitions in infant sensorimotor development and the prediction of childhood IQ. American Psychol 27:728–748 (1972).

56. Gesell A. Infancy and Human Growth. New York:MacMillan, 1928.

57. Fagan JF, Singer LT. Infant recognition memory as a measure of intelligence. In: Advances in Infancy, Vol 2 (Lipsitt LP, ed). Norwood, NJ:Ablex, 1983.

58. Jacobson SW, Fein GG, Jacobson JL, Schwartz PM, Dowler JK. The effect of intrauterine PCB exposure on visual recognition memory. Child Dev 56:853–860 (1985).

59. Gunderson VM, Grant KS, Burbacker TM, Fagan JF III, Mottet NK. The effect of low-level prenatal methylmercury exposure on visual recognition memory in infant crab-eating macaques. Child Dev 57:1076–1083 (1986).

60. Jacobson SW, Jacobson JL, Sokol RJ. Prenatal alcohol exposure and infant information processing ability. Child Dev 64:1706–1721 (1993).

61. McCall RB, Carriger MS. A meta-analysis of infant habituation and recognition memory performance as predictor of later IQ. Child Dev 64:57–79 (1993).

62. Teller DY, McDonald M, Preston K, Sebris SL Dobson V. Assessment of visual acuity in infants and children: The acuity card procedure. Dev Med Child Neurol 28:779–789 (1986).

63. Bellinger D, Leviton A, Waternaux C, Needleman H, Rabinowitz M. Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. New Engl J Med 316:1037–1043 (1987).

64. Streissguth AP, Barr HM, Sampson PD. Moderate prenatal alcohol exposure: effects on child IQ and learning problems at age 7-1/2 years. Alcohol: Clin Exp Res 14:662–669 (1990).

65. Coles CD, Brown RT, Smith IE, Platzman KA, Erickson S, Falek A. Effects of prenatal alcohol exposure at school age. Physical and Cognitive Development. Neurotoxicol Teratol 13:357–367 (1991).

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

67. Streissguth AP, Martin DC, Barr HM, Sandman BM, Kirchner GL, Darby BL. Intrauterine alcohol and nicotine exposure: attention and reaction time in 4-year-old children. Dev Psychol 20:533–541 (1984).

68. Streissguth AP, Barr HM, Sampson PD, Parrish-Johnson JC, Kirchner GL, Martin DC. Attention, distraction and reaction time at age 7 years and prenatal alcohol exposure. Neurobehav Toxicol Teratol 8:717–725 (1986).

69. Streissguth AP, Bookstein FL, Sampson PD, Barr HM. Attention: prenatal alcohol and continuities of vigilance from 4 through 14 years. Dev Psychopathol 7:419–446 (1995).

70. Brown RT, Coles CD, Smith IE, Platzman KA, Silverstein J, Erickson S, Falek A. Effects of prenatal alcohol exposure at school age. II. Attention and behavior. Neurotoxicol Teratol 13:369–376 (1991).

71. Kail R. Sources of age differences in speed of processing. Child Dev 57:969–987 (1986).

72. McCall RB. Nature-nurture and the two models of development: A proposed integration with respect to mental development. Child Dev 52:1–12 (1981).

73. 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. N Engl J Med 300:689–695 (1979).

74. Riley EP. The long-term behavioral effects of prenatal alcohol exposure in rats. Alcohol Clin Exp Res 14:670–673 (1990).