
Assessment of Developmental Toxicity: Neuropsychological Batteries

Jane Holmes Bernstein

Department of Psychiatry, Children's Hospital, Boston, Massachusetts, and Harvard Medical School, Boston, Massachusetts

Assessment of change in behavioral functioning in children as a function of neurotoxicity is not a trivial undertaking. Psychological tests, widely (though erroneously) considered to be the "gold standard" for measurement of behavior in humans, are not adequate for the task; they tap the structure of cognition, not the behavioral repertoire, and cannot (alone) address developmental change. Comprehensive neurobehavioral assessment must be undertaken within a multidisciplinary assessment strategy incorporating knowledge of brain and brain development, cognitive processes and their development, brain-behavior relationships, and detailed knowledge of neurotoxicants, their action and the exposure thereto. Initial assessment batteries must be adequately broad ranging and must incorporate strategies and data for evaluating the impact of predictable non-brain variables; they must also be cost efficient to respond to the realities of funding and the exigencies of field testing. Measures of neuropsychological outcome are optimally characterized as they relate to behavioral domains specified in terms of the competencies of infants and children of different ages; relevant information is derived from demographic, socioeconomic, medical, developmental, and educational sources, as well as from detailed observational data and performance on psychological tests. Two levels of assessment are proposed. — *Environ Health Perspect* 102(Suppl 2):141–144 (1994).

Key words: developmental neurotoxicity, neuropsychological assessment, children, development, psychological tests, behavioral domains

Introduction

Toxins do not affect behavior—at least, not directly. Toxins have their impact on brain. Brain is the necessary, though not a sufficient, substrate for behavior (1). In the adult, with a (presumably) well-differentiated brain, the relationship of brain to behavior is essentially one of "where": different brain systems support relatively discrete aspects of behavior. The impact of toxins in the adult, then, may well be seen at the level of more or less clearly definable cognitive functions or processes. The assessment of toxic agent impact on adult functioning can thus focus, at least initially, on the detection of change in cognitive functions/processes.

In the child, however, toxins do not simply "hit" a given system, the "where"; they hit a "what" at both a "where" and a "when" (2). Thus, the hit might be on processes such as cell migration, proliferation, aggregation or differentiation, or synaptogenesis or myelination; it might involve system A or system B; it might be at 12 days postconception or at 3 months postuterine life—or at 6, or 15 years of age. Each or all of these hits may have its own particular signature in behavior. In the

child, whose behavioral repertoire is developing, toxins can have deleterious impact on both specific cell groups or brain structures, the "where," and on the building of systems over time, the "what" and the "when." The assessment of behavioral change secondary to toxic agent exposure in the child must take account these developmental interactions.

Once the system is built, the relationship between brain and (at least, subsets of) behavior appears to be specified, relatively speaking. Given neural substrates are associated with (more or less) specifiable cognitive functions or processes. These can be tapped by means of psychological tests (3).

The Limitation of Psychological Testing

The term "psychological tests" refers to one of the primary products of the empiricist tradition in psychology, that is, learning and psychometric theory, as manifest in IQ and other standardized tests. In the context of the present discussion the history of such tests, and their development, is of note. First, such tests were developed without reference to the underlying brain that mediates the behavior in question. Over the early and middle decades of this century, psychologists struggling to validate their emerging discipline as "scientific" and "objective" in the physics-defined *zeitgeist* of the day explicitly eschewed the brain in favor of objectively definable cognitive

functions. Tests were developed to demonstrate the structure of cognition, not brain.

The cognitive functions delineated in humans were typically based on the functioning of the usual experimental animal of cognitive psychologists of the time, the college student, who, for the purposes of this discussion, is presumed to have a mature adult brain. However, it is by no means obvious that what constitute the components of the adult cognitive architecture, fully developed, are—at any given epoch—the same, or in the same relationship to each other, in the developing child.

Thus, the tools that psychology has developed to measure behavior—psychological tests—are limited in their application (at least, standing alone) to the problem of evaluating behavior change in the child. The cognitive architecture of the child is not a priori comparable to that of the adult; it differs from that of the adult, and differs differently, at different developmental periods. Furthermore, psychological tests tap behaviors "cross-sectionally," when applied to both adults and children. The theory governing them and their use does not, nor does it need to, incorporate concepts of development. But, in the developing child, the impact of toxic agents affect behavior across time.

To date, the preferred outcome measures for toxic agent impact studies in children have been IQ and academic achievement testing. The basic question

Address correspondence to Dr. Jane Holmes Bernstein, Neuropsychology Fegan 11, Children's Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115. Telephone (617) 735 6398. Fax (617) 730 0228/0457. E-mail bernstein@A1.tch.harvard.edu

has been framed in terms of IQ change or "learning disabilities." There are many problems with this as a strategy.

First, if there is no change in either IQ or achievement performance in the context of a suspected toxic agent exposure we know essentially nothing. Given the limitations of these tools as measures of "brain" (the level at which the toxins are presumed to act), a lack of change in IQ or academic achievement measures does not by any means rule out an effect of a given toxin. A toxic agent could, for example, derail the development of behavioral regulatory function, that is, the child's ability to independently organize his moment-to-moment behavior. This is all too frequently not evident under the well-structured conditions of formal testing. It could nonetheless give rise over the long term—under the "free field" conditions of real life—to school dropout, lack of employment, severe emotional distress, and possibly contact with the law. Potentially, all these effects may be very costly to society as well as to the individual.

A second problem is that it is by no means clear that learning disabilities can be diagnosed by IQ and achievement scores alone; they can be present even though scores and/or score profiles are not particularly discriminative. Under such conditions a broader based examination of cognitive skill profiles typically is needed.

A third problem is that IQ tests simply do not "cover the ballpark": they only tap part of the skills repertoire. Take the language domain: an adequate score on Verbal IQ measures, for example, is often taken to indicate adequate language competence, but Verbal IQ taps verbal knowledge, not necessarily language facility (4). These can be dissociated either positively or negatively. Even language tests only yield information about portions of the language domain; they do not determine how effectively language is used. Tests in general yield information about the structure of cognition, but they do not characterize the way in which cognitive skills or processes are deployed in actual "behaving."

An IQ and achievement test strategy cannot address the source of change when change is present. If there is change on IQ and achievement measures, we only know that there is change, we do not know what causes the change in psychological terms, nor do we know that the toxic agent under suspicion is responsible. Nor do we know if the source is "cognitive" or "toxic." Multiple social, cultural, economic, demographic, genetic, familial, or caretaking fac-

tors can influence behavioral outcome in humans (5). Further information will be needed about cognitive performance, about the action of and exposure to suspected toxic agents, and about psychosocial and demographic variables. An assessment that does not include the relevant information related to both brain and nonbrain variables which might impact on behavioral outcome is not cost efficient, especially considering the cost, in energy, time, money, and resources of merely mounting a research study with children.

Fifth, individual tests and test batteries also have their limitations. All tests are only as good as their construct validity in tapping the needed domains, and the quality of their standardization. And data based only, or primarily, on tests also has the problem of being time locked; normative standards are fixed in time while the organism is changing.

In summary, what are widely thought to be the "measurement tools" of psychology will not suffice to assess the possible impact of toxic agent exposure on behavior. In and of themselves, psychological tests do not "measure" in a formal fashion; they do not measure brain; they cannot address developmental change. However, they are perfectly good as "tools"; it is only when the tools become viewed as the be-all and end-all of the game that they fail. Unfortunately in psychology, and particularly where the educational psychology establishment has been concerned, the tools are dangerously close to becoming the end in themselves: the tail is wagging the dog. However, psychological testing is not psychological assessment (6). Psychological testing can only tap cognitive functions; it takes psychological assessment to evaluate a person, or a person's brain.

The Importance of Assessment

What then is psychological or neuropsychological assessment? A comprehensive assessment of a person requires consideration of all (or as many as possible) of the variables that might impinge on an individual and affect the way he or she behaves. Neuropsychologically, and where children are concerned, it requires that we evaluate potential changes in the whole repertoire of behavior that the human brain normally manifests across development. But by no means are all of these behaviors directly brain-related. The impact of experience and the environment on the development of behavioral competence is well documented in the neurosci-

entific, psychological and sociological literatures. I would argue that the primary goal of neuropsychological assessment, both in the clinical context and in the field—especially when one is working with children—is not, in some sense, to "rule in" brain, but to rule out all of the nonbrain possibilities that might account for a given behavioral outcome before suggesting that we might know what a particular brain-behavior relationship might be at any given age.

In assessment, then, psychological test performance is one, but only one, of a variety of outcome measures that will be needed; psychological tests are necessary, but not sufficient. To assess the potential impact of toxic agents on the developing brain and on behavior, an assessment strategy will be needed. This must be based on a (admittedly beginning) theory of neurobehavioral development in humans, incorporating what we know of brain, of the development of brain, of behavioral processes and their development, of brain-behavior relationships in both developing and adult human and nonhuman animals, and of nonbrain contributions to behavioral competence.

From this perspective, it will clearly take more, rather than less, to assess the possible impact of toxic agents on children. However, the assessment of possible environmental toxicity takes place in a sociopolitical context; there are some realities that cannot be ignored. There are notable limitations on what a small member of society will tolerate in terms of sustained assessment (even with breaks). There are by no means insignificant limitations on what bigger members of society (the research team) think is reasonable to do, or can realistically manage, "in the field." There are even greater limitations on what major societal institutions are willing to fund, even for the good of their littlest members and their future. Not only are there limited funds and many claimants, but in many instances society as a whole puts a high value on the products whose manufacture may lead to environmental toxicity. There are widely varying views of the cost-benefit ratios so entailed.

However, we already know that toxins can be damaging to the human animal. The costs of such damage to the individual and to society can be very high indeed, not only in terms of emotional stress and fiscal strain for an impaired youngster and his or her family, but also in terms of specialized medical care, early interventions, special educational services, loss of adult productivity, need for sheltered work opportuni-

ties, Social Security benefits, to name a few; such costs are borne by society as a whole. For example, \$150,000 is not an unusual cost for the initial medical care of a low birth weight premature infant, that is, before his or her first discharge from the hospital.

In terms of implementation, also, it is not cost efficient to do only the minimum, intervention possible especially when very detailed questions are already being asked and children need to be assessed to allow researchers to answer those questions coherently. As all developmental researchers can report, it takes so much organization, effort, funds and energy simply to mount an assessment of children at all, that it would be particularly foolhardy not to do enough to answer all the necessary questions. IQ and achievement tests are not enough theoretically, nor do they alone make practical sense. We can hardly afford not to implement a coherent assessment strategy if we are going to do it at all.

The Alternative Strategy

If performance on tests is not the optimal end point for developmental behavioral assessment, what is? Behavioral domains. A behavioral-domains approach has several significant advantages. Behavioral domains, adequately specified, can be tapped at all relevant ages, with tests appropriate to the developmental competencies of the target animal (7). Using the language domain as an example, it is impossible to determine mean length of utterance to tap the language competence of a 10-day-old infant, but it is feasible to evaluate whether the child is attempting to engage its caretakers in a reciprocal interaction—a critical foundation for communicative competence. The emphasis is on the use to be made of the skill, not on the skill itself.

The adequately specified behavioral domain is also a crucial pivot between animal and human research. In conjunction with detailed knowledge of the neural substrates for human and nonhuman behaviors, specific behavioral domains have been tapped with formally comparable tasks in developing and adult humans and nonhuman primates, with and without lesions (surgical and “natural”) (8,9), and with species-relevant tasks hypothesized to be functionally or behaviorally equivalent in nonprimate animals (10).

A behavioral-domains strategy also provides for principled data reduction in two regards. First, a major factor in assessing both the scientific validity and the cost effi-

ciency of a proposed study relates to statistical power: how many subjects are needed to permit valid assessment of how many data points? Set up costs, subject availability and attrition limit the size of cohorts for analysis. Using competency within behavioral domains as the unit of analysis facilitates examination of outcome with smaller subject groups. Second, the actual measures selected to characterize a given domain can be chosen for their optimal match with the competencies of the organism at the age in question. Analysis is conducted by behavioral domain, not by specific test or evaluation technique.

Behavioral domains also have the advantage of continuing to be relevant. Psychological tests “wear out,” that is, their normative data become outdated. A behavioral-domains approach permits “old” data to be used, and to be integrated with the products of ongoing research in brain-behavior relationships and associated test or technique development.

Perhaps the most important contribution of the behavioral-domains approach is in its ability to provide comparable data over time. This allows the construction of an enduring database or registry to permit evaluation and tracking not only of the impact of toxicity but also of the efficacy of our response to it. In a funding climate that is not at present conducive to longitudinal research this is particularly important.

So, what will the assessment strategy require? Clearly, it cannot be based solely on neurobehavioral assessment. Neurobehavioral assessment cannot stand alone as an outcome measure: one cannot simply measure without reference to what one is measuring, when, and in whom. The overall enterprise will need to be multi-disciplinary with contributions from (at the very least) toxicology, developmental neuroscience, experimental psychology, clinical medicine, adult and developmental neuropsychology, and biostatistics.

What is the neurobehavioral component of this larger strategy then? As in toxic exposure research with adults, cost efficiency requires a two-step approach: the first, to ask the question: is there an effect? the second: what is its source psychologically or toxicologically? The following discussion is based on the assumption that, for any given proposed study, there is a sufficiently high “index of suspicion” to warrant an investigation of children themselves. Given the exigencies of field testing (with specific reference to the likelihood of cooperation and motivation on the part of both target individual and con-

trols), it makes no sense to mobilize the resources needed to get children scheduled for testing and then only perform a screening battery (assuming that the goal of “screening” is brevity in testing). Where the goal is indeed to screen, rather than to assess a risk that is more-or-less well-defined, a well designed developmental questionnaire and interview with the primary caretakers is both a better and more cost-efficient use of resources.

Proposed Test Batteries for Children

Beyond the initial screening level, the neurobehavioral assessment strategy for children is based on the following. First-level assessment batteries should have as their primary goals that they be sufficiently broad ranging, that they incorporate strategies and relevant data for evaluating the impact of predictable nonbrain variables, and they be cost-efficient. First-level batteries are only targeted with respect to the age groups (and language group, where applicable) of the target populations.

First-level assessment batteries address domains of functioning subsumed under executive control processes, skills/knowledge bases, and achievement (academic, social, societal). Relevant data are derived from a two-component evaluation procedure in which data are obtained separately from the child and from his or her caretakers (familial, medical, and educational). Information will be obtained from the child by means of diagnostic interview, administration of tests, and behavioral observations. The relative contributions of these elements will vary as a function of age. Information obtained from caretakers will be by means of structured interviews and the completion (oral or written) of questionnaires and inventories. These latter may be collected in face-to-face conversation or by telephone contact.

This type of assessment strategy can be optimally undertaken in the field with concurrent data collection by a two-person research team one to interview/test the child, one to interview the caretaker.

The data to be obtained from the child provide information relevant to the following domains: behavioral regulation, attentional capacities, learning skills, memory capacities, problem-solving skills, motor skill (gross, fine, graphomotor), sensory capacities, general cognitive ability (a full intelligence battery is neither needed, nor a justifiable use of time) communicative competence/language abilities, visuospatial processing, social cognition, socio-emo-

tional status/adjustment, and, academic achievement.

These domains are tapped by direct physical and neurological examination of the child, administration of formal tests in standard form, detailed observational data of both "free field" (caretakers), and psychological testing behaviors. Outcome measures at this level include achievement of developmental milestones, standardized test scores, performance relative to controls on specific cognitive measures, and scores on behavioral inventories.

Information obtained from caretakers by questionnaire and interviews will include data about the child's family pertaining to family circumstances (cultural background, education, vocational and economic status) and to family history (medical, neurological, psychiatric condition); and data about the child, including birth history, achievement of developmental milestones (motor/sensory, cognitive, social, emotional), medical, psychological, and educational history. Additional data will be obtained from behavioral rating scales and social skills and personality inventories to address behavioral regulation competencies (state maintenance, motor and verbal activity levels, arousal, deployment of attention, etc., as appropriate for age), problem-solving skills, social skills/peer interaction, and emotional status, as well as vulnerability to psychiatric disorder. The potential impact

of different cultural expectations, home environments and caretaking styles is evaluated by appropriate measures in these domains.

A comprehensive assessment within this framework is estimated to take a maximum of 3 to 4 hrs of testing, with even the most competent child (who is likely to be able to do most). The primary caretaker is interviewed at the same time. Given the necessary information obtained, this compares favorably with the estimated 2 hr that is typically needed for the IQ-and-achievement-test strategy.

It is important to emphasize in this type of strategy that not all behaviors at all ages are most economically tapped by tests. In addition to questionnaires and inventories with caretakers, behavioral ratings, observational paradigms and reliability judgments of both behaviors and diagnostic categories can also be used in assessing the children's functioning. Psychology certainly does not lack for a wealth of well-standardized instruments of these types with appropriate controls for predictable biases in examiners, judges, or interviewees.

As in all behavioral research, specific procedures will need to be deployed for assessing and maintaining reliability in test administration, scoring and coding of data, data entry, etc., as well as for controlling bias in data collection.

In contrast to first-level assessment, second-level batteries are highly targeted. They build on first-level assessment data, but their primary goals are to address questions of specificity, sensitivity, and causation with respect to the relationship of toxic agent to a behavioral change identified by the first-level assessment. As such they explore one or more behavioral domains in depth. To answer specific questions, second-level batteries zero in on specific populations at risk; they are designed with knowledge of the suspected neurotoxicant, its mode of action and predilection for different brain systems and/or processes, and the timing, amount and duration of exposure to it. Second-level batteries should incorporate knowledge of the brain-behavior relationships of the neural structures or systems thus implicated; the development of both neural structures and the behaviors they subserve; and measurement techniques appropriate to tap the cognitive processes and functions involved.

Finale

As can be seen, there is no single battery for evaluating the potential impact of toxic agents on the developing child. I cannot recommend any specific tests in this endeavor; many are appropriate. Overall strategy, a principled theoretical framework, and adequately specified behavioral domains are what count, not tests.

REFERENCES

1. Greenough WT, Black JE, Wallace CS. Experience and brain development. *Child Dev* 58:539-559 (1987).
2. Nowakowski RS. Basic concepts of CNS development. *Child Dev* 58:568-595 (1987).
3. Lezak M. *Neuropsychological Assessment*, 2nd edition. New York:Oxford University Press, 1983.
4. Rudel RG, Holmes JM, Pardes J. *Assessment of Developmental Learning Disorders: A Neuropsychological Approach*. New York:Basic Books, 1988.
5. Matarazzo JD. Psychological assessment versus psychological testing. *Am Psycho* 45:999-1017 (1990).
6. Hutchinson LA, Amler RW, Lybarger JA, Chappell W. *Neurobehavioral Test Batteries for Use in Environmental Health Field Studies*. Atlanta:US Department of Health and Human Services, 1992.
7. Diamond A. Neuropsychological insights into the meaning of object concept development. In: *The Epigenesis of Mind* (Carey S, Gelman R, eds), Hillsdale, NJ:Erlbaum, 1991.
8. Freedman M, Oscar Berman M. Comparative anatomy of cortical and subcortical dementia. *Can J Neurol Sci* 13:410-414 (1986).
9. Stanton ME. Assessment of learning and memory in developmental neurotoxicity: animal models. *Environ Health Perspect* 102(Suppl 2):131-139 (1994).
10. Bernstein JH, Waber DP. Developmental neuropsychological assessment: the systemic approach. In: *Neuromethods*, Vol 17, *Neuropsychology*. (Boulton AA, Baker, GB, Hiscock, eds). Clifton, NJ:Humana Press, 1990; 311-371.