For The Defense™

dri™
The Voice of the Defense Bar
The magazine for defense, insurance and corporate counsel

June 2017

TOXIC TORTS AND ENVIRONMENTAL LAW page 48

Including

- **Contaminated Drinking Water: Avoiding and Defending Against Lead-Poisoning Claims**

And More

Also in This Issue

GOVERNMENTAL LIABILITY page 12

Raising the Bar: **Steps Associates Should Take to Handle Their First Hearing Successfully** page 81

Defense Ethics: **The Role of Presidential Legal Counsel in the Trump Era** page 82

And More

"Guesswork" in the Absence of Norms

Cross-Cultural Neuropsychological Assessment in the Forensic Setting

By Gary M. Burt, Adam R. Mordecai, and Brendan D. O'Brien

As immigrants from different cultures enter the United States, some will invariably be exposed to a variety of toxic substances, which will make understanding cross-cultural neuropsychological assessments crucial for practitioners.



Throughout much of its history, the United States has welcomed people from a variety of cultures throughout the world. This most recent wave of immigration has included refugees from Africa, Syria, the Far East, and other countries. These immigrants have brought with them a correspondingly wide variety of cultures. Most heavily represented within this latest influx are the Somali Bantu, Syrian refugees, and people from Burundi, the Philippines, China, Mexico, Venezuela, the Dominican Republic, and Central America. *See generally* Steven A. Camarota & Karen Zeigler, *U.S. Immigrant Population Record 41.3 Million in 2013*, Center for Immigration Studies (Sept. 2014); Nadwa Mossaad, *Refugees and Asylees: 2015*, Office of Immigration Statistics, Department of Homeland Security (Nov. 2016). Each of these groups has brought with it to the United States its own culture, language, and customs. Neuropsychologists providing mental health services to these various groups are well aware that language, culture, and custom tremendously influence performance on the test instruments used to assess brain behavior relationships. In fact, those very same experts, practicing in countries outside of the United States, have commented on the relative weakness of the test instruments developed for a largely white, Western European population to assess brain behavior relationships reliably and validly in differing populations. Whether the test instruments applied in a cross-cultural context (*i.e.*, administered to a person belonging to a cultural group for which the test was not developed) are truly measuring brain behavior relationships, or simply the influence of language, culture, or cus-

■ Gary M. Burt and Adam R. Mordecai are shareholders, and Brendan D. O'Brien is an associate, of Primmer Piper Eggleston & Cramer, practicing in the firm's Manchester, New Hampshire, office. Mr. Burt specializes in trial advocacy on behalf of individuals, insurance carriers, and businesses, and he has taken over 200 trials to verdict. He has been a member of DRI for 30 years. Mr. Mordecai is a past president of the Tri-State Defense Lawyers Association. His practice is focused on defending personal injury claims, handling insurance coverage issues, and litigating construction and employment disputes. Mr. O'Brien's practice focuses on premises liability, product liability, professional liability, and complex tort matters. He is a member of DRI and the Tri-State Defense Lawyers Association.

tom, cannot be determined without critical scientific inquiry and years of study. To the extent that whether neuropsychological tests can be successfully administered cross culturally has been studied, the authorities are overwhelmingly critical of applying a test instrument to a group or culture for which the test instrument has not been normed.

Understanding the test instrument, its strengths and limitations, its psychometric properties, including reliability, validity, and construct validity, will be very important in analyzing and addressing test interpretation.

Attempting to address these issues, various methods have been suggested, including using non-verbal testing to reduce the influence of language on test performance and adapting tests to address the influence of language, culture, and custom. Some psychologists have insisted that the only true method to develop good test instruments is to develop good local norms for the specific culture to be assessed. While the profession has generally criticized using non-verbal tests as a measure of performance, and similarly criticized assessing subjects in a language other than their own native language, the equally expensive and daunting task of creating local norms for the wide variety of cultures that exist has led some to conclude that using assessment instruments for groups for which they were not normed is the least harmful way in which to assess a specific culture.

As more and more immigrants from different cultures enter the United States, some will invariably be exposed within the United States to a variety of toxic substances that can affect the brain. Adults and children moving into low-income housing, for example, may be exposed to lead in paint, piping, or other similar sources. Those exposures, along with exposure to mercury, pesticides, manganese, arsenic, and even tobacco smoke, undoubtedly will lead to claims of brain injury caused as a consequence of the exposure to the toxin. To prove these claims, the plaintiff will necessarily have to establish that there was, in fact, brain injury that was caused by the exposure to the toxin. That proof generally comes from neuropsychological test instruments and the conclusion drawn from test performance. In the forensic setting, however, the plaintiff's expert will need to address appropriately the problems inherent in performing a cross-cultural assessment using test instruments that are not normed for the subject of the evaluation. *Daubert* and its progeny should stand in the way of the admission of any resulting cross-cultural assessments, unless the plaintiff proves that the assessment methodology is sufficiently reliable in accordance with the tenets of a *Daubert* analysis.

This article will focus on the numerous problems with cross-cultural neuropsychological assessments involving neuropsychological test instruments for which standard norms for the subject's particular culture do not exist. The authors maintain that without good science, when a test instrument is administered in a language other than the individual's primary language, or when standard, culture-appropriate norms have not been developed for the test, no reliable neuropsychological assessment inference or conclusion can be drawn based on the individual's test performance. The degree to which language, culture, and custom can influence test performance is largely unknown, but to the extent that the scientific studies exist, they argue persuasively and strongly against using such instruments or methods to assess individuals of a culture for which the tests have not been normed. This is especially true in the forensic context.

Toxic Exposure and Its Effects

The world is full of substances that can harm humans in sufficiently high doses. One substance that has generated significant litigation as a result of human exposure is lead. Lead is a naturally occurring toxic metal that has been widely used in household products such as paints, toys, and batteries. In children, elevated blood lead levels may have a significant effect on a child's intelligence and other neurobehavioral functions, such as memory, processing speed, attention, and spatial perception. In adults, exposure to high levels of lead can cause cognitive deficits.

Similar to lead, mercury has been shown to harm humans' neurodevelopment when they are exposed to the toxin at sufficiently high levels. Children exposed to high levels of mercury have been shown to have lower IQs, worse memories, and poorer verbal skills than children who have not been exposed to it. Adults can experience similar cognitive impairments when exposed to mercury. Because mercury can contaminate drinking water and certain seafood, humans are often exposed to it through their diets.

Pesticides have also been shown to harm human neurobehavioral function. Exposure to certain pesticides have been associated with lower verbal and memory scores on standardized tests among children, while exposure to other pesticides have been associated with deficits in children's motor and mental functioning. In adults, exposures to certain pesticides have been associated with memory deficiencies and difficulty with motor function. As pesticides are often sprayed on plants, fruits, and vegetables, humans can be exposed to them through the air, their food, or their drinking water.

Exposure to manganese, an essential mineral, has also been shown to impair children's neurobehavioral function when they are exposed to sufficiently high levels of the mineral. Specifically, children exposed to high levels of manganese have been found to have lower intelligence and memory scores. Because manganese is one of the most abundant minerals in the Earth's crust, humans can be exposed to it through contaminated drinking water.

Although more known for being poisonous, arsenic can have an effect on the neurobehavioral development of children similar to manganese. Chronic, low-level exposure to arsenic during childhood has

been associated with deficits in intelligence and memory. Typically, arsenic is consumed through contaminated drinking water, but it can also be ingested through tainted food or polluted air.

Exposure to environmental tobacco smoke has also been associated with neurodevelopmental deficits in children. Several studies suggest that children exposed to environmental tobacco smoke score worse on standardized intelligence tests and perform worse academically than those children who did not have such an exposure. These studies indicate that in addition to its well-known effect on respiratory function, exposure to environmental tobacco smoke may be associated with an adverse effect on neurodevelopment.

Neuropsychological Assessment Test Instruments

When evaluating a neuropsychological assessment, one of the primary concerns should be the history of the use of the instruments in the scientific literature for the toxin of concern. Thus, one will need to know if a test instrument alone, or the instrument in a battery, has been used to study the effect of a given toxin on either a group or an individual. Using a test that has not been put to that use previously may have limited value unless there is scientific support for using the test to assess function regardless of the cause of impairment.

Understanding the test instrument, its strengths and limitations, its psychometric properties, including reliability, validity, and construct validity, will be very important in analyzing and addressing test interpretation. For example, the NEPSY-II, a developmental neuropsychological assessment battery, is commonly used to assess children's various domains of function. The instrument, however, has undergone limited factor analysis, and serious questions exist whether the various subtests, in fact, measure the domain at issue. M.R. Reynolds *et al.*, Confirmatory Factor Analysis of the NEPSY-II for Preschool Age Children (Aug.2009) (paper presented to the annual meeting of the American Psychological Association, Toronto); M.R. Reynolds *et al.*, *What Does the NEPSY-II Measure in School Age Children*? (paper presented to the annual meeting of the American Psychological Association, Toronto).

An attorney seeking to gain basic knowledge about the various tests used in an assessment is encouraged to consult Strauss, Sherman, and Spreen, *A Compendium of Neuropsychological Tests* (3rd ed.), or Lezak, *Neuropsychological Assessment* (5th ed.). Both discuss at length the instruments and their strengths, weaknesses, and use. Google Scholar also provides access to studies of the instruments as the studies have been published in leading journals.

Though test instruments may have solid psychometric properties with regard to a relevant population, the *Standards for Educational and Psychological Testing* (Standards), developed jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, make clear that "[i]f no normative or validity studies are available for a relevant population, test interpretations should be qualified and presented as hypotheses rather than conclusions." *Standards* §10.5. Validity, of course, "refer[s] to a particular interpretation for specified uses. It is incorrect to use the unqualified phrase 'the validity of the test.'" *Id.* at 11 (Introduction). Thus, tests themselves do not have validity: it is the interpretation of test performance as the performance compares to a normative standard that determines validity, and only if the normative standard includes members of the relevant population. For that very reason, many of the tests mentioned below have been adapted and at times translated into various different languages. Once they are adapted and translated, however, the test instruments need to be standardized to ensure the reliability and the validity of test score interpretation.

Intelligence Tests

There are a number of intelligence tests. The two most popular for adults and children are the Wechsler Adult Intelligence Scale WAIS), Fourth Edition, and the Wechsler Intelligence Scale for Children (WISC), Fifth Edition. The predecessors to both tests have been well studied, and they are generally considered the "gold standard" for intelligence measurement. Studies of both tests have included evaluation of test performance after exposure to toxins.

Apart from the Wechsler tests, other intelligence tests include the Reynolds Intellectual Assessment Scales, for both children and adults, the Raven's Progressive Matrices, the Stanford-Binet Intelligence Scales, Fifth Edition, and the Kaplan Brief Intelligence Test, Second Edition.

Memory and Learning Tests

Commonly used memory and learning tests include the Wechsler Memory Scale, Third Edition, the Wide Range Assessment of Memory and Learning, Second Edition, the Rey Complex Figure Test, the California Verbal Learning Test-II, and the California Verbal Learning Test, the children's version.

Visual-Spatial Relationship Tests

Tests used to measure visual-spatial relationship include the clock-drawing test, the

Hooper Visual Organization Test, and the facial recognition test.

Motor Function Tests

Motor function, something that is believed to be impaired by a number of toxins, can be tested by the finger-tapping test, the grip-strength test, and the Purdue Grooved Pegboard Test.

Tests themselves do not have validity: it is the interpretation of test performance as the performance compares to a normative standard that determines validity, and only if the normative standard includes members of the relevant population.

Attention Tests

The number of tests of attention is ever-expanding, but the tests generally used in assessment practice include the Colors Trails Test, the Children's Color Trail Test, the Trail Making Test, the Connors Continuous Performance Test-II, and the Pace Auditory Serial Addition Test.

Language Tests

The commonly used language tests include the Boston Naming Test and the Peabody Picture Vocabulary Test.

Mood and Personality Tests

Tests used to assess mood and personality include the Beck Depression Inventory of Executive Function, the Personality Assessment Inventory, and the Minnesota Multiphasic Personality Inventory.

Effort Tests

Effort tests are a necessary prerequisite to any forensic assessment. A number of tests now are available for adult and pediatric assessment, including the Test of Memory Malingering, Green's Symptom Validity Test, the Rey 15-Item Test, the Dot Counting Test, and the 21-Item Test.

Cross-Cultural Intelligence and Neuropsychological Testing: Theory and Criticism

To understand the theory behind, and criticisms of, cross-cultural neuropsychological assessments, a practitioner must understand how the following elements affect the scientific rigor on which the assessments must rest: (1) test development and appropriate norms; (2) test instrument reliability and validity concepts; and (3) test administration practice deviation or using language interpreters.

Test Development and the Importance of Norms

The importance of norms cannot be overstated. Norms serve as the fundamental building blocks for reliability and validity, and the absence of appropriate norms renders any neurological assessment unreliable. The development of reliable norms and tests for neuropsychological evaluation is a lengthy and intricate scientific process. Broadly speaking, the process is as follows: first, the test developer will start with the intention of measuring something with perceived testing value. The test developer will then develop test items and rigorously investigate those items to ensure that they effectively measure the thing that they are supposed to measure. Once the test developer achieves a satisfactory set of test items and develops an effective method for scoring those items, the test developer can then administer the test to a large number of people and tally the scores across the items to develop the data for the "norms" for the test instrument. The total points that a participant receives on a test are referred to as the participant's "raw score." Those raw scores by themselves are, however, of little value because tests vary in difficulty, and other variables can make them misleading. As a result, test developers convert the raw scores to standard scores that have uniform meaning to overcome the limitations of raw scores.

That standardization of raw scores can be achieved by testing a substantial number of individuals, lining up the scores received from the lowest to the highest, and determining for each individual the percentage of other test takers whose scores that person surpassed. Doing so could reveal that while a score of 80 might sound good, it is actually in the bottom 25th percentile on one test, but in another test, a raw score of 40 is actually in the top 25th percentile.

Results on tests in neuropsychology are often used to compare an individual's performance to that of other individuals. For the information about relative standing to be of value, however, the individual tested must be compared to a suitable group, and the data on which the comparison is based needs to be sound or trustworthy. In other words, for neuropsychological testing to have relative value, the subject's results need to be compared to a well-matched set of normative data. Without appropriate norms against which to compare results, an individual's test scores are of little use in judging the subject's cognitive function. Viewing results in isolation do not tell the test administrator anything about how the subject compares to the broader population or to the "normal" condition.

Selecting and using a suitable, normative sample is critical to ensure the reliability of the testing. Just as comparing a six-year-old's performance on a reading test to a normative sample comprised of high schoolers is unlikely to tell you much of value about whether that six-year-old has reached a reading level appropriate for his or her age, comparing neuropsychological test scores from a subject of one culture to a normative sample made up of only individuals from another culture will not provide reliable or valid results. It is, therefore, not only critical in neuropsychology to apply carefully developed tests and norms, but also to select norms that can be used to compare the testing subject's performance. Appropriate neuropsychological norms are essential. They serve as a standard against which a person's performance can be evaluated and then interpreted. In the context of cross-cultural neuropsychological testing, the importance of using appropriately matched norms is often overlooked.

Test Instrument Reliability and Validity Concepts

The *scores* generated by a test can have reliability *estimates* for certain populations, but a test standing alone is not "reliable." Reliability is a product of data gathering and statistical modeling, and it is not inherent in a particular test as some practitioners often mistakenly believe. Rather, the reliability of a given test *score* depends on the normative data to which the score is compared.

Likewise, validity is not a characteristic of a test. Validity is a characteristic of the inference that one draws from someone's performance on a test. And that is usually expressed as a score. Validity is, however, conditional and relates back to the context and purpose of the application of the test and to the norms used in drawing inferences from test performance. Validity considerations are at the heart of the very science of neuropsychology. The *Standards* define validity as follows: "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses....The process of validation involves accumulating relevant evidence to provide a scientific basis for the proposed score interpretations." Standard 9.4 cautions that "[w]hen a test is to be used for a purpose for which little or no validity evidence is available, the user is responsible for... obtaining evidence of reliability/precision of the test scores and the validity of the interpretations supporting the use of the scores for this purpose." The comment to Standard 9.4 indicates that *"[t]he provisions of this standard should not be construed as prohibiting the generation of hypotheses from test data. However, these hypotheses should clearly be labeled as tentative."* (emphasis added). Thus, practitioners conducting cross-cultural assessments are required, as a first step, to establish the "cross-cultural psychometric integrity for the measures used." Byrd et. al., *State of Multicultural Neuropsychological Assessment in Children: Current Research Issues*, 18 Neuropsychol. Rev. 214, 220 (2008).

One can never simply *assume* that the measures used in an assessment have "good standardization and reliability." Those psychometric properties must be proved for each application of the selected tests. According to Byrd, the failure to take steps to demonstrate that test data interpretations would have reliability or validity, as applied to the test subject, and the failure to understand the psychometric properties of the tests administered thoroughly renders any results scientifically invalid. Applying selected neuropsychological tests to a subject or group of a certain culture for which no known normative data exists could not, without establishing validity, lead to any "valid" conclusions about whether the subject or the subjects had suffered brain injury.

Deviating from Test Administration Practice or Inserting Interpreters to Address Cross-Cultural Differences

Conducting a successful cross-cultural neuropsychological assessment is a challenging endeavor that requires considerable research and planning. One cannot simply apply neuropsychological tests and norms cross culturally and hope to yield results with any professionally acceptable diagnostic accuracy. The great weight of scientific authority on cross-cultural neuropsychological assessment agrees. Rosselli and Ardila, for example, have found that "interpretation of the performance of individuals from other cultures using US norms might result in significant errors in assessment, particularly if the examiner's intention is to evaluate the integrative of the brain-behavior relationship." Rosselli & Ardila, *The Impact of Culture and Education on Non-Verbal Neuropsychological Measurements: A Critical Review*, 52 Brain and Cognition 326, 331 (2003). Skuy cautions that "the use of appropriate norms is a *necessary* condition for accuracy of diagnosis." *Skuy et al., Suitability of Published Neuropsychological Test Norms for Urban African Secondary Schools Students in South Africa*, 30 Personality and Individual Differences 1413, 1424 (2001) (emphasis in original). Mulenga likewise notes that a "[l]ack of normative information means that one cannot effectively compare a child with others of the same age to determine the presence or absence of difficulties." Mulenga *et al., Performance of Zambian Children on the NEPSY: A Pilot Study*, 20 Developmental Neuropsychology 375, 376 (2001).

Wong similarly explained that "tests originally normed with one cultural group may not have equivalent construct validity when used with another cultural group, even when the test is accurately translated." T. Wong, *Ethical Controversies in Neuropsychological Test Selection, Administration, and Interpretation*, Applied Neuropsychology, Vol. 13, No. 2, at 71 (2006).

Faced with these obstacles, practitioners often deviate from acceptable test administration or interpretation practices during cross-cultural neuropsychological assessments. Some may assume that their experiences with entirely different cultures give them insight into a unique cultural group with which they are unfamiliar. There is, however, nothing within the scientific literature to suggest that experience with one ethnic group translates to expertise with another distinct group. It is a fallacy to assume that an expertise with a culturally distinct group grants license to assess a different group. Moreover, psychologists who believe that they have the ability to work competently with one ethnic group often fail to recognize that they do not have the competence necessary to work with others. Fortuny & Mullaney, *Assessing Patients Whose Language You Do Not Know*, The Clinical Neuropsychologist, Vol. 12, No. 1, at 119 (1998). Ethnic pop-

Applying selected neuropsychological tests to a subject or group of a certain culture for which no known normative data exists could not, without establishing validity, lead to any "valid" conclusions about whether the subject or the subjects had suffered brain injury.

ulations "unequivocally demand expertise and advanced native fluency in their respective languages" that many practitioners do not have. The relevant, applicable scientific literature recognizes that linguistic minorities differ in more ways than just language, "not only from mainstream English-speaking culture, but also from other English-speaking minorities."

Administering neuropsychological tests in English to subjects with a limited grasp of that language will not produce any diagnostically reliable results and is not supported by science.

Id. at 115. For this reason, among others, "[c]ultural expertise or competence at the individual level is essential for the clinician who is working with cross-cultural populations," Brickman & Cabo, *Ethical Issues in Cross-Cultural Neuropsychology*, Applied Neuropsychology, Vol. 13, No. 2, at 95 (2006).

Perhaps recognizing a lack of proficiency with the subject's language, a practitioner might enlist the aid of an interpreter to bridge the linguistic gap. Using interpreters in the neuropsychological evaluation process is, however, roundly criticized by the scientific authorities because "it is impossible for the monolingual clinician to assess the level of the fluency in the interpreter and verify the accuracy of the translated information." *Id.* at 96. Without true bilingualism, "errors that may be quite grave are unlikely to be detected by the clinician signing the report." Fortuny & Mullaney, *supra*, at 116.

Others, perhaps aware of the problems injected by using interpreters, may instead chose simply to administer test instruments to a subject in English, based on a belief that the subject is sufficiently proficient in English to proceed. There is as much danger in this approach as any other. While conversational fluency "may be adequate for day-to-day interactions, it falls short of the higher-order fluency required for cognitive processing in a 'context reduced' situation such as the neuropsychological evaluation." Mindt *et al.*, *Neuropsychological, Cognitive, and Theoretical Considerations for Evaluation of Bilingual Individuals*, 18 Neuropsychological Rev. 255, 260 (2008). Though individuals might acquire limited conversational proficiency within a few years, they take much longer to acquire the "higher-order language proficiency" required to participate in valid neuropsychological testing. *Id.* at 259–60. As a result, it is critical in cross-cultural assessments to test a patient's language proficiency when assessing a bilingual or an English language learner (ELL). Duarte, *Evaluating Bilingual Students for Learning Disabilities*, *in* Bakken et al., (eds.), *Learning Disabilities: Identification, Assessment, and Instruction of Students with LD (Advances in Special Education*, at 129, 132 (Emerald Group Publishing 2013). There are many tests that focus on language proficiency and can provide formal measures of a patient's English proficiency. Simply assuming that a subject is sufficiently proficient based on in-office conversations, length of time in school, or length of time living in the United States, is not enough for valid assessment of a bilingual or an ELL patient.

The "conversational proficiency" that may be mistaken as sufficient for assessment by some practitioners is known as "basic interpersonal communication skills," or "BICS," and refers to a person's ability to engage in the conversation typically found in informal settings. This type of language is usually referred to as "everyday language." "Cognitive academic language proficiency," or "CALP," on the other hand, is necessary for academic learning and is a prerequisite to meaningful neuropsychological testing results. *See* Mindt *et al.*, *supra*, at 259–60. CALP is critical for neurological assessment, and a lack of higher-order language fluency could, at least in part, help to explain why bilingual children in the United States and abroad may perform worse on standardized academic achievement tests. Without adequate language assessment measures, it is always uncertain whether an assessment represent a bilingual patient's true performance.

Administering neuropsychological tests in English to subjects with a limited grasp of that language will not produce any diagnostically reliable results and is not supported by science. As Fortuny recognized, "[t]here may be different reasons why clinicians… indulge in assessing patients whose language they do not know. These appear to include poor judgment and ignorance, and seem to be accompanied by a conspicuous absence of accurate self-appraisal with regard to the impact that these factors have on one's decision-making." Fortuny & Mullaney, *supra*, at 119. Fortuny noted that when "confronted with their actions, clinicians may offer various justifications… [and] cite the fact that the population is under-served." *Id.* Disregarding the linguistic and cultural issues results, however, only in "guess-work" and can have "far-reaching and devastating effects." *Id.*

Looking for supposed "patterns" of poor performance to determine impairment in the cross-cultural setting will not alleviate any of these concerns or overcome the problems created by embarking on a cross-cultural assessment without adequate planning or resources. Not knowing how an average, healthy person in the subject's culture would score, or how many times the average person within that culture would score in the low range, belies any effort to find a pattern of low scores. Without that critical foundational information (which would be contained within properly developed and appropriate norms), it is *impossible* to conclude that a "pattern" of low scores has any meaning whatsoever. Healthy, normal people within the normative sample will obtain low scores on neuropsychological tests. The number of abnormal scores depends on the number of scores generated. As the number of scores increases, the probability of even a normal, healthy person having an abnormal score increases as well. Brooks, *Healthy Children Get Low Scores Too: Prevalence of Low Scores on the NEPSY*, 25 Archives of Clinical Neuropsychology 182–190 (2010). This psychometric fact—that more scores will tend to lead to more *low* scores—suggests that any methodology looking for a

pattern of low scores will create a high risk of false positive error, even within an U.S. normative sample and *especially* in a cross-cultural assessment.

Can a Cross-Cultural Neuropsychological Assessment Ever Be Admissible Under *Daubert*?

The scientific literature recognizes that test instruments developed in the Western context are not directly applicable to assessing children of other cultures. With respect to cross-cultural assessments, it therefore cannot be assumed that an instrument developed for patients in the United States would measure the same constructs or abilities in patients in other cultures. Isquith, *Approaches to Assessment of Very Young Children in Africa*, *in* Bovin & Girodani (eds.), *Neuropsychology of Children in Africa*, at 20 (Springer 2013). In the field of neuropsychology, in fact, it is "well recognized that the application of tests of cognitive ability from one ethnic group to another without appropriate standardization is highly problematic for both diagnostic and placement purposes." Shuttleworth *et al.*, *Cross Cultural Effects on IQ Test Performance*, 26 Journal of Clinical and Experimental Neuropsychology 903 (2004). The interpretative problem created by administering U.S.-normed tests to ethnic minorities is "daunting and presents formidable challenges." Manly & Echemendia, *Race Specific Norms*, 22 Archives of Clinical Neuropsychology 325, 323 (2007). Manly acknowledges that which norms to use is being asked by neuropsychologists "in settings as diverse as a community clinic for dementia and the stately environment of a federal court room." *Id.* But the answer, according to the scientific literature, is not simply to apply U.S. norms to a culturally dissimilar population and assume that a reliable diagnosis can thereby be obtained, but rather it depends on the question being asked:

> If our interest is to determine how [an] individual performs in relation to the majority of Americans, then we use US population reference norms. However, if we wish to know whether the particular score reflects clinically significant impairments in performance, then we must identify a reference group that most closely approximates the individual's demographic characteristics.... If we do not we may [] run the risk of misdiagnosis, with the resulting overpathologizing and overtreating.

Id. The same is true in the litigation arena, where the focus is on true impairments to performance rather than comparisons to the U.S. population. Given the scientific consensus that cross-cultural application of neuropsychological norms and tests is of limited (if any) validity, an expert proposing to conduct such an assessment in toxic tort litigation is sure to face a strong *Daubert* challenge. The expert may prevail and be allowed to testify, but that should be the case only if the expert has undertaken the work needed to validate his or her cross-cultural assessment independently.

In responding to *Daubert* motions, a plaintiff may simply point to the expert's clinical use of the challenged testing methodology, or if the plaintiff is a child, the expert may point to the use of a similar methodology in the school setting. Neither sufficiently excuses the reliability and validity problems introduced by applying U.S. norms to a culturally dissimilar individual. The use of U.S. norms in a clinical or school performance setting may be appropriate under certain circumstances, but using those same norms to diagnose brain injury in individuals of another culture is not.

In the school setting, neuropsychological measures can be used to determine if there are any areas in which a child is not performing at the level of the normative sample, *i.e.*, the child's school peers; those measures are not used to assess whether a child has brain impairment. Using tests that are normed to the United States for U.S. plaintiff children allows one to draw inferences about how they compare to other kids in their school, in terms of their progress relative to these kids, and that can be helpful in developing educational programs, but those are the only kinds of interferences that would be reasonable.

While children of minority cultures in U.S. schools are routinely given assessments based on U.S. norms, it is not for the purpose of discerning whether they have an impairment, to diagnose a brain injury, or to determine *why* a particular child is performing worse than his or her peers. Those assessments are for the purpose of determining how well they are performing as measured against their classmates, and whether they need educational support. Even if "many" school psychologists actually did use U.S.-normed measures, it would still not in and of itself equate to "general acceptance" in the field of neuropsychology.

Courts have generally found that practitioner-only acceptance cannot satisfy the general acceptance prong. *See, e.g., U.S. v. Alexander*, 526 F.2d 161, 164 n.6 (8th Cir.1975) ("Some commentators have posited the argument that the polygraph need only attain general acceptance among the polygraph operators themselves to satisfy the test for admissibility... This position must be rejected."). Indeed, if practitioner communities alone could define general acceptance it would "allow a group that advocates a technique or method to self-validate it simply by

declaring acceptance." Bert Black, *A Unified Theory of Scientific Evidence*, 56 Fordham L. Rev. 595, 633 (1988). Narrowing the relevant scientific community to those individuals whose "professional reputations and commercial interests.... depend on validation of the technique, [would make] general acceptance... a foregone conclusion." Adina Schwartz, *A "Dogma of Empiricism" Revisited: Daubert v. Merrell Dow Pharmaceuticals, Inc. and the Need to Resurrect the Philosophical Insight of Frye*, 10 Harv. J. Law & Tech. 149, 201 (1997). As the Federal Circuit Court of Appeals explained, "[t]he lesson of the Supreme Court's rejection of 'general acceptance' as the sole standard for expert testimony, in favor of the *Daubert–Kumho* reliability standard is that 'widespread use' or 'general acceptance' is an imperfect proxy for reliability." *Libas, Ltd. v. United States*, 193 F.3d 1361, 1368 (Fed. Cir. 1999). With respect to cross-cultural neuropsychological assessment, using Western or U.S. norms to evaluate an individual from another culture may be done in the classroom or clinical setting, but that does not translate into sufficient reliability or validity to overcome a *Daubert* challenge.

A more expedient and practical approach to validating a cross-cultural assessment methodology would be to collect enough data about other individuals to determine whether the methodology is working as expected.

If clinical application is not enough, and if a practitioner is not supposed to use a translator or Western norms to evaluate patients from other cultures, what is a psychologist to do? First, a practitioner should take stock of a patient's culture and language abilities. As noted above, this should include administering a formal test of the patient's language skills. It should also include researching the patient's culture and his or her background and health history and taking a survey of the profession to determine if there is someone better qualified (by background, language, or training) to conduct the cross-cultural assessment. If the practitioner determines that he or she can, and should, move forward with the assessment, consideration should be given to either adapting the proposed test instruments formally, or more realistically, to obtaining enough other data to judge the validity, reliability, and precision of the practitioner's methodology.

True test adaption is a complicated process, but authors attempting cross-cultural assessments have in the past adapted tests to fit the subject population. *See* Kitsao-Wekulo; Holding *et al.*, *Assessing Cognitive Outcomes in a Rural African Population*, 10 Journal of the International Neuropsychological Society 246 (2004). In her study of a sub-Saharan African group of children, Kitsao-Wekulo modified the tests' word lists and time limitations to suit the culture being studied better. In his development of the "Draw and Person," or "DAP," test, Cecil Reynolds found that while it worked well in the United States, in other countries, the scoring system had to be modified because the things that children emphasize in their drawings of humans are different in different cultures outside the United States. In her work with a rural Kenyan population, Holding found that test adaptation was necessary because the children in one part of Kenya would not touch the foam blocks used in one part of the tests that other children had accepted without hesitation. Holding *et al. supra*, at, 260.

A more expedient and practical approach to validating a cross-cultural assessment methodology would be to collect enough data about other individuals to determine whether the methodology is working as expected. The *Standards* provide that it is incumbent on the person administering a test to obtain evidence of reliability, validity, and precision. There are at least three ways in which that information can usually be gathered in the cross-cultural neuropsychological assessment context: (1) a psychologist can gather a cohort of healthy individuals from within the plaintiff's community to establish a roughly matched reference or control group against which to compare the plaintiff; (2) he or she can conduct a longitudinal study involving serial neuropsychological assessments of the plaintiff so that the plaintiff can serve as his or her own control over time, and his or her performance can be compared to previous performance levels; or (3) the practitioner could conduct a full epidemiological study, using a control group and developing a standardized sample of culturally alike people who had not been exposed to the toxin in question. Though the cost of this third approach would be significant, it would produce the most scientifically reliable and valid conclusions, and it would be considered the "gold standard" approach to the problems inherent in cross-cultural assessments. The costs associated with presenting scientifically sound opinions should not be grounds to default to unscientific guess work. Something must be done, adhering to the *Standards*, to ensure that the results of a proposed cross-cultural assessment are reliable, valid, and precise.

Conclusion

Without some independent check on the application of U.S. norms to a plaintiff of another culture, and without confirmation that the results from that comparison are a meaningful assessment of impairment, a cross-cultural neuropsychological assessment is nothing more than "guesswork." If the tests selected are not normed for the culture of the individual being assessed, and if nothing is done to gather validity evidence to support the proposed testing methodology, there is no way to know whether the results are indicative of a neurological impairment, or whether the results simply measure a difference in culture or language. Absent such independent validation, the results of a cross-cultural assessment should be rejected by any court as unreliable when faced with a *Daubert* challenge. FD