

2008-NH-R0521.002

BAXTER v. CHARLES TEMPLE

SHELBY BAXTER, BY HER MOTHER AND NEXT FRIEND, PATRICIA BAXTER

v.

CHARLES TEMPLE & a.

No. 2007-102.

Supreme Court of New Hampshire, Merrimack

Argued: January 31, 2008

May 20, 2008

Seufert Professional Association, of Franklin (Christopher J. Seufert on the brief), and Thornton & Naumes, LLP, of Boston, Massachusetts (Neil T. Leifer and Andrew S. Wainwright on the brief, and Mr. Leifer orally), for the plaintiff.

Wiggin & Nourie, P.A., of Manchester (Gary M. Burt and Doreen F. Connor on the brief, and Mr. Burt orally), for the defendants.

Hall, Stewart & Murphy, P.A., of Manchester (Francis G. Murphy on the brief and orally), for the American Academy of Clinical Neuropsychology, as *amicus curiae*.

DUGGAN, J.

The minor plaintiff, Shelby Baxter, by and through her mother and next friend, Patricia Baxter, appeals the exclusion by the Trial Court (Hollman, J.) of two expert witnesses in her negligence action against the

2

defendants, Charles and Kelly Temple. The exclusion of these witnesses resulted in dismissal of the plaintiff's case. We reverse in part, vacate in part, and remand.

I. Factual and Procedural Background

The record supports the following relevant facts. Between May 11, 1995, and May 11, 1996, the plaintiff and her parents resided in an apartment in Concord that they rented from the defendants. In early September 1995, the plaintiff, who was almost fourteen months old at the time, was tested for lead paint poisoning. The test results revealed an elevated blood lead level of thirty-six micrograms per deciliter. On September 26, 1995, the New Hampshire Department of Health and Human Services investigated the premises and found substantial evidence of lead paint contamination.

The plaintiff subsequently filed this action, alleging, among other things, that the defendants failed to warn her of the presence and dangers of the lead paint. She contended that her exposure to and ingestion of the high levels of lead paint present in the apartment caused her to suffer from "lead paint poisoning and the effects thereof including but not limited to: reduced life expectancy, brain damage, past and future pain and suffering, and loss of expected earnings capacity. . . ."

To prove her case, the plaintiff designated three expert witnesses: (1) Barbara Bruno-Golden, Ed.D., a neuropsychologist who evaluated the plaintiff by administering a series of neuropsychological tests to determine her cognitive and behavioral status; (2) William Bithoney, M.D., a pediatrician who concluded that the plaintiff suffers from organic brain syndrome caused by lead poisoning; and (3) Arthur Kaufman, M.Ed., a vocational rehabilitation specialist. On the day of trial, the defendants moved *in limine* to exclude the testimony of Dr. Bruno-Golden as unreliable under New Hampshire Rule of Evidence 702, RSA 516:29-a (2007), and Daubert v. Merrell Dow Pharmaceuticals, Inc., ~~509 U.S.~~ 579 (1993).

Thereafter, the trial court held a six-day Daubert hearing. The trial court heard testimony from: (1) Dr. Bruno-Golden; (2) Sandra J. Shaheen, Ph.D., a pediatric neuropsychologist introduced by the plaintiff to support Dr. Bruno-Golden's testimony; and (3) David Faust, Ph.D., a psychologist presented by the defendants to criticize Dr. Bruno-Golden's testimony. The trial court, in a lengthy order, subsequently ruled that Dr. Bruno-Golden's testimony was inadmissible.

The plaintiff moved for reconsideration, requesting that the trial court allow Dr. Bruno-Golden to testify to her administration and scoring of three

specific tests -- two tests measuring IQ and one test measuring attention -- that the plaintiff alleged were "not subject to the methodological objections raised by the defendants." The defendants objected, arguing that these three tests could not properly and reliably be extracted from the comprehensive battery of tests that Dr. Bruno-Golden administered and, further, that such limited testimony would "confuse the jury, not assist it." The trial court agreed with the defendants and denied the plaintiff's motion.

Subsequently, the defendants moved in limine to exclude the testimony of Dr. Bithoney and Mr. Kaufman, arguing that both experts' opinions were unreliable because they were "based almost exclusively on Dr. Bruno-Golden's unreliable findings." The plaintiff conceded that Mr. Kaufman was precluded from testifying at trial, but contended that Dr. Bithoney's testimony was admissible because it was based upon sufficient information independent from Dr. Bruno-Golden's reports. The trial court found that Dr. Bithoney's testimony was unreliable and inadmissible. Because the plaintiff no longer had an expert to prove her case, the trial court concluded that the plaintiff could not proceed and dismissed the case.

On appeal, the plaintiff contends that the trial court erred by: (1) excluding Dr. Bruno-Golden's testimony as unreliable; (2) not permitting Dr. Bruno-Golden to testify to the results of the two IQ tests and the attention test; and (3) excluding Dr. Bithoney's testimony. We address only the first argument because we agree that the trial court erred in excluding Dr. Bruno-Golden's testimony as unreliable.

II. Standards for Admissibility of Expert Testimony

Rule 702 states:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

N.H.R. Ev. 702. Thus, expert testimony must rise to a threshold level of reliability to be admissible. Baker Valley Lumber v. Ingersoll-Rand, ~~148 N.H. 609~~, 613 (2002).

In Baker Valley, we applied the Daubert framework for evaluating the reliability of expert testimony to Rule 702. Id. at 614. Subsequently, in 2004, the legislature enacted RSA 516:29-a, which provides:

4

I. A witness shall not be allowed to offer expert testimony unless the court finds:

- (a) Such testimony is based upon sufficient facts or data;
- (b) Such testimony is the product of reliable principles and methods; and
- (c) The witness has applied the principles and methods reliably to the facts of the case.

II. (a) In evaluating the basis for proffered expert testimony, the court shall consider, if appropriate to the circumstances, whether the expert's opinions were supported by theories or techniques that:

- (1) Have been or can be tested;
- (2) Have been subjected to peer review and publication;
- (3) Have a known or potential rate of error; and
- (4) Are generally accepted in the appropriate scientific literature.

(b) In making its findings, the court may consider other factors specific to the proffered testimony.

"Section II of RSA 516:29-a unambiguously codifies the four Daubert factors we applied in Baker Valley, and section I(b) codifies Daubert's requirement that the court preliminarily assess `whether the reasoning or methodology underlying the testimony is scientifically valid.'" State v. Langill, ___ N.H. ___, ___ (decided April 4, 2008) (quoting Daubert, 509 U.S. at 592-93; citation omitted). "The trial court functions only as a gatekeeper, ensuring a methodology's reliability before permitting the fact-finder to determine the weight and credibility to be afforded an expert's testimony." Baker Valley, 148 N.H. at 616 (citation omitted). The inquiry is a flexible one, and the focus "must be solely on the principles and methodology, not on the conclusions that they generate." State v. Dahood, ~~148 N.H. 723~~, 727 (2002) (quotation omitted). Moreover, the list of Daubert factors are "meant to be helpful, not definitive. Indeed, those factors do not all necessarily apply even in every instance in which the reliability of scientific testimony is challenged." Kumho Tire Co., Ltd.

v. Carmichael, 526 U.S. 137, 151 (1999). Thus, one or more of these factors is relevant only "if appropriate to the circumstances." RSA 516:29-a, II(a).

"Importantly, the Daubert test does not stand for the proposition that scientific knowledge must be absolute or irrefutable." Dahood, 148 N.H. at 727. To be sure, "it would be unreasonable to conclude that the subject of scientific testimony must be known to a certainty; arguably, there are no certainties in science." Id. (quotation omitted). Rather, "the proposed scientific testimony must be supported by appropriate validation -- i.e., good grounds, based on what is known." Id. (quotation omitted). "[A]s long as an expert's scientific testimony rests upon good grounds, . . . it should be tested by the adversary process -- competing expert testimony and active cross-examination -- rather than excluded from jurors' scrutiny for fear that they will not grasp its complexities or satisfactorily weigh its inadequacies." Langill, ___ N.H. at ___ (quotation omitted). Thus, "[i]f [the evidence] is of aid to a judge or jury, its deficiencies or weaknesses are a matter of defense, which affect the weight of the evidence but do not determine its admissibility." Dahood, 148 N.H. at 727 (citation omitted).

In Langill, we interpreted RSA 516:29-a, I(c) as requiring the trial court to also "examine whether a witness has in actuality reliably applied the methodology to the facts of the case." Langill, ___ N.H. at ___. However, for the testimony to be inadmissible, the flaws in application must so infect the procedure as to skew the methodology itself. Id. Otherwise, "the adversary process is available to highlight the errors and permit the fact-finder to assess the weight and credibility of the expert's conclusions." Id. (citation omitted).

III. Admissibility of Dr. Bruno-Golden's Testimony

The trial court found that Dr. Bruno-Golden used the Boston Process Approach (BPA) in evaluating the plaintiff. It found that "the Boston Process Approach as used by Dr. Bruno-Golden is generally accepted in the appropriate scientific literature as a sound clinical approach to evaluating individuals for brain injury." However, the court concluded that the evidence failed to show that it is "generally accepted . . . in the making of a forensic assessment." Therefore, the trial court found that the plaintiff had not shown that Dr. Bruno-Golden's methodology "is generally accepted in the appropriate scientific literature as reliable in a legal proceeding." In reaching this conclusion, the trial court focused upon the plaintiff's failure to demonstrate that the specific battery -- the entire series of tests viewed as a whole -- employed by Dr. Bruno-Golden in this case was, or could be, tested, was subject to peer review and publication, or has a known or potential rate of error. Additionally, the trial court concluded that Dr. Bruno-Golden's methodology as administered to the plaintiff was not reliable.

We generally review a trial court's determination of expert reliability under Rule 702 for an unsustainable exercise of discretion. State v. Pelletier, 149 N.H. 243, 251 (2003). "When the reliability or general acceptance of novel scientific evidence is not likely to vary according to the circumstances of a particular case, however, we review that evidence independently." Dahood, 148 N.H. at 726 (citation omitted).

Here, because the reliability of the methodology used by Dr. Bruno-Golden, the BPA, is not likely to vary according to the circumstances of each case, we review its scientific reliability independently and make our own determination without regard to the trial court's findings. Id. However, we review the trial court's finding that Dr. Bruno-Golden did not reliably apply that methodology to the facts of this case for an unsustainable exercise of discretion. See Langill, ___ N.H. at ___.

A. Overview of the BPA and Dr. Bruno-Golden's Testimony

We derive the following facts from our review of the evidence in the record and pertinent legal and scientific sources. Neuropsychology is "a specialty of psychology concerned with the study of the relationships between the brain and behavior, including the use of psychological tests and assessment techniques to diagnose specific cognitive and behavioral deficits and to prescribe rehabilitation strategies for their remediation." Hoskins v. State, 702 So. 2d 202, 209 n. 5 (Fla. 1997) (quoting Stedman's Medical Dictionary 1049 (25th ed. 1969)). The field of neuropsychology has developed two major approaches to the selection of neuropsychological tests: the fixed (or standardized) battery approach and the flexible battery approach. See Stern, Admissibility of Neuropsychological Testimony After Daubert and Kumho, 16 NeuroRehabilitation 93, 95-96 (2001)(fn1); see also Minner v. American Mortg. & Guar. Co., 791 A.2d 826, 869 (Del. Super. Ct. 2000); 2 L. Russ et al., Attorneys Medical Advisor § 23:17, at 23-20 (2005) (citing Levin, A Guide to Clinical Neuropsychological Testing, 51 Archives of Neurology 854 (1994)). But see

Rabin et al., Assessment Practices of Clinical Neuropsychologists in the United States and Canada: A Survey of INS,

NAN, and APA Division 40 Members, 20 Archives of Clinical Neuropsychology 33, 48 (2004) (presenting results of a survey requesting that neuropsychologists self-identify as using one of three approaches: a fixed battery approach, a flexible battery approach, or a flexible approach). A battery is a group of tests used to evaluate neurological domains of functioning.

Pursuant to the fixed battery approach, the neuropsychologist administers a uniform series of tests, such as the Halstead-Reitan Neuropsychological Test Battery or the Luria-Nebraska Neuropsychological Test Battery, to all patients, regardless of their complaints or the referral question. L. Russ et al., supra §§ 23:17, :18, at 23-20, 23-21. This approach allows a neuropsychologist to identify the presence or absence of brain damage or impairment, identify the area of the brain involved, and assess whether the injury is recent or has had an opportunity to stabilize. Stern, supra at 96.

Under the flexible battery approach, the neuropsychologist administers a group of core tests, uses the results in conjunction with the patient's history to formulate a hypothesis concerning the patient's cognitive functioning, and then administers additional specially selected tests to further explore cognitive deficits identified by the core tests and test the hypothesis. L. Russ et al., supra §§ 23:17, :19, at 23-20 to -21, 23-22; see also Stern, supra at 95-96. In contrast to the fixed battery approach, the flexible battery approach allows the examiner to not only identify brain damage, but also better define the specific nature of the impairments resulting from that damage, Stern, supra at 96, thus providing more "in-depth testing in areas of specific deficits," Minner, 791 A.2d at 86 (quotation omitted).

The BPA is a variation of the flexible battery approach that adds a qualitative element to evaluating brain function. Consistent with the flexible battery approach, the BPA, in most instances, uses a collection of core and satellite tests to assess various domains of cognitive functioning, such as verbal memory, visual memory, planning, attention span, language, visual perception, academic performance, and self-control. W.P. Milberg et al., The Boston Process Approach to Neuropsychological Assessment, in Neuropsychological Assessment of Neuropsychiatric Disorders 65, 67 (Igor Grant & Kenneth M. Adams eds., 1986). These core or satellite tests, however, are not predetermined; the approach is to look at domains of behavior and an examiner may choose from a number of published tests that will each provide a good sampling of that behavior. Like the flexible battery approach, the BPA also uses the concept of hypothesis testing, pursuant to which the neuropsychologist evaluates the results of an initial cognitive battery test, in conjunction with information from parents or teachers, to determine whether

8

and to what extent additional areas of brain function require further exploration.

The BPA was developed in response to the perceived disadvantages associated with a purely test-score approach to psychological assessment, like the fixed battery approach, including the distortions in the interpretations, conclusions, and recommendations that result from the one-sided database associated with the test-score approach, Stern, supra at 96, and lack of precision and sensitivity in observing and assessing particular domains of behavior, cf. Minner, 791 A.2d at 869. See also W.P. Milberg et al., supra at 65-66. Thus, in an effort to provide a more precise delineation of functions, the BPA rejects a completely quantitative approach to neuropsychological testing. Instead, because qualitative data, such as direct observation of a patient's test-taking behavior, purportedly allows a neuropsychologist to obtain significant information concerning the patient's life situation, Stern, supra at 96, a neuropsychologist assesses not only the quantitative scores a patient achieves on a particular test, but also the qualitative nature and effectiveness of the behavior the patient demonstrates in attempting to solve the problems presented in the test. W.P. Milberg et al., supra at 65-67; Stern, supra at 96.

Dr. Bruno-Golden is a neuropsychologist who has supervised, reviewed or tested well over 200 children with lead paint exposure. As of 2005, her clinical practice was approximately seventy-percent pediatric and thirty-percent adult. When conducting neuropsychological assessments of children, Dr. Bruno-Golden follows the same general procedure for each child; i.e., the BPA.

Before beginning any testing, Dr. Bruno-Golden conducts a clinical diagnostic interview, reviews available medical and school records, and speaks with the person referring the child. At the beginning of the testing day, she gives the child's parents a behavioral screening questionnaire for them to complete while she is administering tests to the child. Consistent with the flexible battery approach, she then administers a comprehensive neuropsychological battery, which she prefers to administer in one full day because it provides her with an opportunity to observe the child's attention and concentration as it would be during a normal school day.

Dr. Bruno-Golden begins the morning session of testing by asking the child to draw a picture of herself for the purpose of observing the child's approach to the testing situation. She then administers a standardized core test, normally an intelligence test, that allows her to screen the child's abilities relative to others in the child's age group. Generally, Dr. Bruno-Golden uses the Wechsler Intelligence Scale for Children (WISC), which yields a full scale IQ. The

WISC test serves as both an intellectual measure and as a screening tool

9

that identifies potential neurological issues. The WISC test has several versions and has been updated through the years.

Next, Dr. Bruno-Golden administers the Rey-Osterrieth Complex Figure (ROCF) test, which assesses visual construction skills, non-verbal memory, and language-based issues. The ROCF test consists of three steps: copy, immediate recall, and delayed recall. First, the administrator shows the child a figure and asks the child to copy that figure. The materials are then taken away from the child, and after a few minutes, the administrator instructs the child to draw what she remembers. The materials are again taken away, and, thirty minutes after the copy, the administrator instructs the child to draw the figure for a third time.

During the thirty-minute interval between the copy trial and the delayed recall trial, the examiner may not administer tasks involving "visuospatial stimuli," but may administer a verbal test. J.E. Meyers & K.R. Meyers, Rey Complex Figure Test and Recognition Trial: Professional Manual 8 (Psych. Assess. Resources, Inc. 1995). Dr. Bruno-Golden generally obtains a written language sample and administers the Connors Continuous Performance Test (CCPT) during this thirty-minute interval. The CCPT is a fifteen-minute computerized test that measures a child's attention. The test instructs the child to press a button every time she sees any letter that is not an "X" on the screen. The computer scores the test and interprets the child's performance. After the CCPT, Dr. Bruno-Golden completes the delay recall portion of the ROCF test and breaks for lunch. At this time, the parents return the completed behavioral screening questionnaire.

During lunch, Dr. Bruno-Golden formulates a hypothesis by considering the responses to the screening questionnaire and evaluating the results of the morning tests, which together indicate whether the child exhibits cognitive or behavioral issues that require further exploration. She then selects specific tests for the afternoon session that will measure these potential deficits and provide a more comprehensive neuropsychological assessment. These specific tests allow Dr. Bruno-Golden to obtain information concerning various domains of functioning, including sensory-motor, language, visual-spatial judgment, academic performance, memory and learning, and problem solving.

In this case, Dr. Bruno-Golden evaluated the plaintiff for neurological issues that might be associated with her lead paint exposure. She conducted two evaluations -- in 2002, when the plaintiff was approximately seven years old, and in 2004, when she was approximately nine years old. During each session, Dr. Bruno-Golden followed the methodology outlined above, using core and satellite tests, each of which has been tested and peer reviewed, and has a known error rate. Dr. Bruno-Golden also reviewed the plaintiff's medical and psychosocial history as provided by her parents as well as the plaintiff's

10

medical and school records. On both testing days, she instructed the plaintiff to draw a picture of herself, administered a version of the WISC test, administered the ROCF test, obtained a language sample, and administered the CCPT. She then evaluated the results and selected additional tests for the afternoon sessions, including the Finger Tapping (or Finger Oscillation) Test, the Grip Strength Test, the Peabody Picture Vocabulary Test III, the Wide Range Achievement Test-Third Revision, the Wisconsin Card Sorting Test, the California Verbal Learning Test-Children's Version, and the Children Memory Scales.

In 2002, after evaluating the results of the morning tests, Dr. Bruno-Golden determined that the areas of attention, problem solving, memory, and learning, required more testing. For example, she selected a test that measures verbal working memory for structured linguistic material to ensure that this part of the plaintiff's working memory was intact because, during the morning session, the plaintiff had frequently asked the doctor to repeat herself and the directions. Dr. Bruno-Golden also decided to conduct additional core tests in language and motor skills because they were important to provide a comprehensive assessment. Consistent with hypothesis testing, Dr. Bruno-Golden ruled out potential problems based upon the results of these additional tests. Ultimately, her conclusions were grounded in the results of the WISC III, ROCF, and CCPT tests, which she found to be critical to her opinion.

In 2004, the plaintiff's IQ was approximately twenty points lower than her IQ in 2002. As a result of this decline and reported concerns from the plaintiff's parents and teachers, Dr. Bruno-Golden gave the plaintiff more extensive memory testing than she had administered in 2002. Otherwise, she selected tests that evaluated the same cognitive or behavioral constructs, but with slight variations to account for the difference in age. Ultimately, Dr. Bruno-Golden found the results of the WISC IV, WISC III--Digital Span, Wide Range Achievement Test (WRAT)--Decoding, ROCF, Children's Memory Scale, and the CCPT to be critical in reaching her conclusions.

Based upon the results of these assessments, Dr. Bruno-Golden concluded that the plaintiff had "impairments in

the areas of visual dysgraphia and kinetic apraxia, and persistent problems with attention/executive skills, associated with memory and retrieval problems with both verbal and nonverbal lengthy and more complex information." She noted that the plaintiff's neurobehavioral presentation was consistent with her known history of lead paint exposure, and opined that the plaintiff "continue[d] to be at risk for developing language based learning problems, particularly in light of her overall general intellectual decline since her 2002 assessment, as indicated in her current WISC-IV scores with respect to her 2002 WISC-III scores." Thus, Dr. Bruno-Golden diagnosed the plaintiff with, among other things, an unspecified organic mental disorder.

11

B. Reliability of the BPA and Dr. Bruno-Golden's Methodology

The defendants assert that Dr. Bruno-Golden used a "completely flexible approach" to neuropsychological assessment, not the "flexible battery approach." We disagree.

First, as Dr. Shaheen testified, the field of neuropsychology does not uniformly distinguish between these approaches to neuropsychological test selection. Compare Stern, supra at 95 (explaining that the two schools of thought regarding test selection are the fixed battery approach and a more flexible approach, and then comparing the fixed battery approach to the flexible battery approach), with L. Russ et al., supra §§ 23:17, :20, at 23-20 to -21, 23-22 to -23 (explaining that the two major approaches to clinical neuropsychological testing are the fixed battery approach and the flexible battery approach, but noting that the BPA is another school of thought, although not further describing the BPA), and Rabin et al., supra at 48 (presenting results of survey requesting that neuropsychologists self-identify as using one of three approaches: a fixed battery approach, a flexible battery approach, or a flexible approach). Thus, we question whether such a distinction exists, and, if it does, the nature of that distinction. To the extent that the defendants are correct that Dr. Shaheen or Dr. Bruno-Golden admitted that Dr. Bruno-Golden used a "completely flexible approach," we note that defense counsel elicited this testimony in the context of an article drawing a distinction between the two approaches. See Rabin et al., supra at 48.

Second, even assuming such a distinction exists, the defendants agree that the BPA is a flexible battery approach. Further, as the trial court found, "in administering the 2002 and the 2004 neuropsychological tests on the plaintiff, Dr. Bruno-Golden employed the Boston Process Approach." Indeed, she used the same methodology that she normally uses when evaluating children; that is, she administered a core set of tests in the morning, and then, based upon those results, developed a hypothesis and selected satellite tests for the afternoon session. Nothing in this procedure suggests that Dr. Bruno-Golden, as the defendants assert, "self-select[ed] tests according to the [plaintiff]'s presentation and age parameters without any standardization." (Emphasis added.) Rather, in line with the defendants' definition of the flexible battery approach, Dr. Bruno-Golden used the same set of core tests, which tested the same cognitive and behavioral constructs, that she always uses when testing for lead paint exposure. Accordingly, Dr. Bruno-Golden used the BPA in a manner consistent with the flexible battery approach.

We now address whether, if used in a manner consistent with the flexible battery approach, as Dr. Bruno-Golden did, the BPA is admissible under RSA 516:29-a. The defendants assert that for Dr. Bruno-Golden's testimony to be

12

admissible, the comprehensive neuropsychological batteries that she used in evaluating the plaintiff as a whole must have been tested, have been subject to peer review, and have a known or potential error rate. The trial court accepted this proposition as true in evaluating the admissibility of Dr. Bruno-Golden's testimony. Because the specific combination of tests Dr. Bruno-Golden used arguably did not meet these three requirements, the defendants contend, and the trial court found, that her testimony is inadmissible. Distilled to its essence, the defendants' position is that Dr. Bruno-Golden should have used a fixed battery, such as the Halstead-Reitan battery, rather than devising her own battery consistent with the BPA.

In support of this position, the defendants cite provisions 12.4 and 12.5 of the Standards for Educational and Psychological Testing, which the American Psychological Association (APA) has approved, and language from a section in those standards discussing "test interpretation." See American Educational Research Association et al., Standards for Educational and Psychological Testing 117, 120-22 (1999) (APA Standards). The defendants then argue: (1) "the APA Standards allow practitioners to combine tests only when there is identifiable literature that assesses the validity of the combination"; and (2) "[t]he Standards . . . expressly address the validity issues that arise with the administration of non-standardized batteries." Reviewing these provisions in context, we find that the defendants have misapplied them to this case.

Standard 12.4 states:

If a publisher suggests that tests are to be used in combination with one another, the professional should review the

evidence on which the procedures for combining tests is based and determine the rationale for the specific combination of tests and the justification of the interpretation based on the combined scores.

Comment: For example, if measures of developed abilities (e.g., achievement or specific or general abilities) or personality are packaged with interest measures to suggest a requisite combination of scores, or a neuropsychological battery is being applied, then supporting validity data for such combinations of scores should be available.

Id. at 131.

13

This standard requires that, when Dr. Bruno-Golden administered tests in combination with one another as suggested by the publisher of a particular battery, she needed to have a rationale for the specific combination of tests, and a justification for her interpretation of those tests based upon the combined scores. As the example indicates, this standard would have applied if Dr. Bruno-Golden had administered an entire published battery, such as the NEPSY battery discussed below, since the publisher of that battery packages a particular combination of measures for various domains and suggests that the battery results in a requisite combination of scores. The standard, however, does not mandate, as the defendants suggest, that Dr. Bruno-Golden's specific battery as a whole, that is, the combination of all the core and satellite tests, be tested in order for it to be used to make a neuropsychological assessment.

Standard 12.5 provides:

The selection of a combination of tests to address a complex diagnosis should be appropriate for the purposes of the assessment as determined by available evidence of validity. The professional's educational training and supervised experience also should be commensurate with the test user qualifications required to administer and interpret the selected tests.

Comment: For example, in a neuropsychological assessment for evidence of an injury to a particular area of the brain, it is necessary to select a combination of tests of known diagnostic sensitivity and specificity to impairments arising from trauma to various regions of the cerebral hemispheres.

APA Standards, supra at 132.

Contrary to the defendants' assertions, Dr. Bruno-Golden complied with this standard. As the comment indicates, this standard requires that the individual tests selected by a neuropsychologist in combination be valid and appropriate for diagnosing the particular issue. The defendants' expert, Dr. Faust, agreed that it was proper for Dr. Bruno-Golden to administer each of the individual tests she found to be critical to her conclusions for the purpose of evaluating children with lead poisoning. He agreed that a valuable study of the effects of lead in children used groupings of tests that examined various domains of cognitive functioning, and used many of the same tests that Dr. Bruno-Golden used in her assessment. See Ris et al., Early Exposure to Lead and Neuropsychological Outcome in Adolescence, 10 J. Int'l Neuropsychological Society 261, 264-66 (2004).

14

Dr. Faust also conceded that different neuropsychologists may reasonably and justifiably select different groups of tests for evaluating lead poisoning in children; there is no consensus in the field as to a particular battery of tests that is proper for this evaluation; and no particular battery of tests is more reliable than another for evaluating lead poisoning. When asked what group of tests he would give to evaluate a child with elevated lead levels, Dr. Faust testified that if he had to evaluate the child, and the circumstances were proper, he would use the Halstead-Reitan battery, a fixed battery that provides a broad-based definition of brain damage by yielding a brain damage impairment index. As discussed above, this battery is limited in usage because it does not sensitively examine specific domains of functioning such as everyday functional capacity. See Faust et al., Challenging Neuropsychological Evidence in Brain Damage Litigation, For the Defense, June 1994, at 8.

In fact, in a study he conducted of lead poisoning in children, Dr. Faust developed his own battery of tests. See Faust & Brown, Moderately Elevated Blood Lead Levels: Effects on Neuropsychologic Functioning in Children, 80 Pediatrics 623 (1987). In the study, Dr. Faust noted:

Neuropsychologic assessment methods are more sensitive to cognitive deficit than standard intelligence tests alone or limited cognitive batteries. Studies on increased lead levels in which less sensitive techniques were used and in which nonsignificant findings were obtained may thus represent false-negative errors. Although some investigators have used sections of neuropsychologic test batteries, . . . none have administered a comprehensive set of neuropsychologic tests, such as the Halstead-Reitan battery.

Id. at 623-24. Rather than administer the Halstead-Reitan battery in his study, Dr. Faust, like Dr. Bruno-Golden, adopted a battery from the work of a prominent neuropsychologist, and used "standardized psychometric tests" as

well as "a number of clinical tests," noting that "[e]xtensive normative data [was] available on all of the tests and items." Id. at 625. As Dr. Bruno-Golden did, Dr. Faust tested general areas of functioning, including "psychomotor, memory, visual-motor and spatial, language and associated functions . . . , attention and concentration, and reasoning." Id.

Notably, since Dr. Faust administered the foregoing tests as part of a research study, he standardized the results of his self-selected neuropsychological battery by administering the tests to fifteen children with moderately increased blood levels and fifteen "control" children. Id. at 624. However, nothing in the record indicates that a neuropsychologist who is clinically assessing whether a particular child's exposure to lead has resulted

15

in any cognitive deficits, even in the context of a lawsuit, must "standardize" the battery of tests he or she uses for the particular child. Nor does the record demonstrate that a neuropsychologist must follow a fixed battery that has never been proven to be suitable for sensitively detecting whether lead exposure resulted in any cognitive and behavioral deficits. Indeed, Dr. Faust conceded that the Halstead-Reitan battery was a compromised choice.

In light of the lack of a discrete combination of tests sensitive to lead poisoning, and Dr. Bruno-Golden's selection of tests that were individually suitable for evaluating lead poisoning in children, Dr. Bruno-Golden did not violate this standard. Underscoring this conclusion is the fact that the flexible battery approach is the generally accepted approach for neuropsychological testing, see, e.g., Sweet et al., The TCN/AACN 2005 "Salary Survey": Professional Practices, Beliefs, and Incomes of U.S. Neuropsychologists, 20 The Clinical Neuropsychologist 325, 333 (2006) (providing results of a 2005 survey of clinical neuropsychologists, and showing that seventy-six percent used a "flexible battery approach" toward test selection, eighteen percent used a "flexible" approach, and seven percent used a "standardized battery approach"); Rabin et al., supra at 48 (providing results of a similar 2004 survey showing that sixty-eight percent of clinical neuropsychologists favored a "flexible battery approach," twenty percent favored a "flexible approach," and eleven percent favored a "standardized battery"), and Dr. Bruno-Golden administered the BPA to the plaintiff in a manner consistent with the flexible battery approach. By definition, the flexible battery approach does not require the examiner to use a required set of tests for evaluating a question. Instead, it provides the trained examiner with some latitude to use her judgment and select tests that may properly address the problem presented by evaluating the relevant domains of functioning.

The concept of hypothesis testing itself requires the examiner to, based upon the individual and the initial general testing, e.g., the IQ test, form a hypothesis concerning which domains of behavior may be affected, and then select tests to evaluate those domains. Furthermore, the APA Standards encourage examiners to conduct individualized assessments. See APA Standards, supra at 121. Accordingly, when Standard 12.5 is viewed in its proper context, Dr. Bruno-Golden's specific battery was not required to be validated as a whole. Cf. Minner, 791 A.2d at 869 (Because the expert "based her initial results on a sufficient battery of tests," her "results are sufficiently reliable for admissibility.").

The defendants also rely upon certain language found in a section of the APA Standards entitled "Test Interpretation" in support of their position. See APA Standards, supra at 121-22. The paragraph cited begins: "For some purposes, including career counseling and neuropsychological assessment, test batteries are frequently used." Id. at 122. These batteries examine various

16

cognitive domains and "often include tests of verbal ability, numerical ability, nonverbal reasoning, mechanical reasoning, clerical speed and accuracy, spatial ability, and language usage." Id.

When psychological test batteries incorporate multiple methods and scores, patterns of test results frequently are interpreted to reflect a construct or even an interaction among constructs underlying test performances. Higher order interactions among the constructs underlying configurations of test outcomes may be postulated on the basis of test score patterns. The literature reporting evidence of reliability and validity that supports the proposed interpretations should be identifiable. If the literature is incomplete, the resulting inferences may be presented with the qualification that they are hypotheses for future verification rather than probabilistic statements that imply some known validity evidence.

Id.

The defendants argue that this language demonstrates that "validity issues . . . arise with the administration of non-standardized batteries." The language, however, does not distinguish between standardized and non-standardized batteries. Further, this section addresses the interpretation of test score patterns, and, if it applies at all, would likely be relevant to Dr. Bruno-Golden's conclusions, not her methodology.

Even so, if interpreted in the context of flexible batteries, the language seemingly requires that evidence of reliability and validity support a proposed interpretation of the interactions between the subtests of an individual test battery, not the interaction between every test of a comprehensive neuropsychological battery used for assessment. For example, Dr. Bruno-Golden noticed a scatter or variation in the plaintiff's scores on two subtests of the verbal portion of the WISC III. The literature indicated that this variation indicates a weakness in the plaintiff's verbal performance that the WISC III did not sensitively measure. Dr. Bruno-Golden therefore determined that she needed further verbal testing to clearly understand how the plaintiff functions in her left hemisphere. Pursuant to the above language in the APA Standards, Dr. Bruno-Golden properly used literature to interpret the intertest scatter. In the context of a clinical evaluation, as opposed perhaps to a research study, Dr. Bruno-Golden did not need further literature to then validate her interpretation of the interaction between the WISC and the individual test examining the specific construct.

17

To conclude otherwise would require the field of neuropsychology to test, peer review, and calculate error rates for an infinite number of test combinations for the interpretations to be reliable. Each time a new validated and reliable test or battery of tests, such as the NEPSY, is developed or even updated, a clinical examiner could not use it as part of a comprehensive battery since it would be unknown how it interacted with the other tests within that battery. Since the flexible battery approach is the generally accepted approach to conducting neuropsychological assessments, the APA Standards could not logically mandate that a neuropsychologist always use a comprehensive test battery that is validated as a whole.

The amicus curiae supports this conclusion. Drawing a comparison to clinical medicine, the amicus curiae explains that "[t]here is no expectation that the specific battery of neurological exam procedures and diagnostic tests chosen by the neurologist be studied as a whole with regard to validity." Similarly, for neuropsychological testing, "test validity lies in individual tests, not in `test batteries.'" "Either the individual tests selected for inclusion in a flexible-battery are scientifically valid or they are not." Citing a "hallmark article" in an APA journal, the amicus curiae emphasizes that using large fixed batteries is a questionable practice, and that "a flexible, multimethod assessment battery using tests typically employed in practice and selected on the basis of idiographic referral questions" is recommended. See Meyer et al., Psychological Testing and Psychological Assessment: A Review of Evidence and Issues, 56 American Psychologist 128, 154 (2001). Accordingly, the defendants' position that Dr. Bruno-Golden's battery as a whole was required to have been tested, have been subject to peer review or publication, and have a known or potential error rate contravenes the flexible battery approach and the concept of hypothesis testing.

Moreover, under the defendants' position, no psychologist who uses a flexible battery would qualify as an expert, even though the flexible battery approach is the prevalent and well-accepted methodology for neuropsychological assessment. Further, while the approach proffered by the defendants, the fixed battery approach, has been tested, has been peer reviewed, and has known or potential error rates outside the context of lead poisoning, it seems to no longer be the generally accepted methodology for conducting neuropsychological assessments. See, e.g., Sweet et al., supra at 333; Rabin et al., supra at 48; Faust et al., supra at 8. Therefore, the implication of the defendants' position is that no neuropsychologist, or even psychiatrist or psychologist since, in their view, all combinations of tests need to be validated and reliable, could ever assist a trier of fact in a legal case.

However, "[t]he role of the Court when ruling on a Daubert motion is not to resolve the scientific debate, but to determine whether [the] plaintiff[s]'s experts have a reliable basis for their testimony." Palmer v. Asarco

18

Incorporated, 510 F. Supp. 2d 519, 527 (N.D. Okla. 2007). Regardless of whether the fixed battery approach was a better approach to evaluating the plaintiff, the relevant inquiry is whether Dr. Bruno-Golden used a reliable methodology to conduct her neuropsychological assessments. See Minner, 791 A.2d at 869 (noting that, while the plaintiffs argued that the fixed battery approach was the better approach, the "'flexible' approach appears to be an acceptable method for the evaluation of patients").

We agree with the trial court that the BPA is generally accepted in the scientific literature as a reliable method for clinically assessing children for cognitive and behavioral deficits. Cf. id. As Dr. Shaheen testified, "the battery of tests that [Dr. Bruno-Golden] employed followed the general guidelines for neuropsychological assessment for children and [her] qualitative analysis lends information that is used clinically and . . . is a standard clinical approach." See also L. Russ et al., supra §§ 23:9, 23:26, at 23-13, 23-25 to -26; McConnel, The Sevin Made Me Do It: Mental Non-Responsibility and the Neurotoxic Damage Defense, 14 Va. Env'tl. L.J. 151, 156-57 (1994). The defendants' own expert testified that qualitative indicators like those used in the BPA are "potentially important," and never refuted the notion that the BPA is a generally accepted methodology for clinically evaluating children.

Further, although another neuropsychologist may reach different conclusions, the evidence in the record indicates that the BPA as a flexible battery approach can be tested. See RSA 516:29-a, II(a)(1). Another examiner could conceivably administer the same core tests in the morning session, noting both the quantitative scores and the qualitative indicators, evaluate the results and form a hypothesis, and test that hypothesis in the afternoon session by conducting standardized tests that would identify any problem areas. Additionally, as Dr. Shaheen testified, the BPA has been subject to peer review and publication. See RSA 516:29-a, II(a)(2).

While the BPA itself does not have a known or potential error rate, the Daubert factors "do not constitute `a definitive checklist or test'" that must be applied in all circumstances. Kumho Tire, 526 U.S. at 150 (quoting Daubert, 509 U.S. at 594); see also RSA 516:29-a, II. Rather, the factors must be applied with flexibility and in light of the proffered testimony. Baker Valley, 148 N.H. at 616. Given the nature of the BPA, and particularly that it inherently requires some level of flexibility, we find that the known or potential error rate factor is not an "appropriate" consideration in examining its reliability. See RSA 516:29-a, II(a).

However, we note that a critical component of our finding that the BPA meets three of the four Daubert factors is the use of standardized tests. To meet the threshold for reliability, a neuropsychologist applying the BPA must demonstrate that the individual tests he or she administered as part of the

19

battery, not the battery as a whole, have been tested, have been subject to peer review and publication, and have known or potential error rates. Cf. United States v. Efr, 461 F. Supp. 2d 529, 531, 533 (E.D. Tex. 2006) (although neuropsychologist's opinion of insanity was not a reliable conclusion, the battery of tests administered by the neuropsychologist to measure defendant's cognitive abilities, including the Wechsler Adult Intelligence Scale-III, ROCF, Boston Naming Test, Wisconsin Card Sorting Test, Wide Range Achievement Test-4, and Finger Tapping, was reliable because the individual tests could be repeated, had reasonable confidence levels, and had been widely administered).

The defendants do not dispute that each of the individual tests Dr. Bruno-Golden used to evaluate the plaintiff met these requirements. Notably, it might be a different case if Dr. Bruno-Golden had used the BPA in a manner inconsistent with the flexible battery approach, such as by specifically designing a set of tasks for the plaintiff that were not standardized, or by creating a new test to measure a particular deficit area. See W.P. Milberg et al., supra at 67; cf. Downs v. Perstorp Components, Inc., 126 F. Supp. 2d 1090, 1109-10, 1128 (E.D. Tenn. 1999) (excluding expert's testimony as unreliable partly because he developed and administered a completely self-selected battery that had no basis or consistency, and compared the results to norms that were based upon his own collective data, as opposed to standardized test data). In those circumstances, the lack of validity evidence for the individual tests administered by the neuropsychologist might affect the reliability of the methodology itself. See Langill, ___ N.H. at ___.

Accordingly, we find that, when the BPA is administered in a manner consistent with the flexible battery approach, as described above, it is generally a reliable approach to neuropsychological assessment, and is thus a reliable methodology for determining a person's cognitive status. Cf. Palmer, 510 F. Supp. 2d at 522, 524-25 (finding that, despite "the somewhat subjective nature of plaintiffs' neurocognitive injuries," neuropsychologist had a reliable basis to testify to test results of plaintiffs with lead exposure, and that plaintiffs suffered from certain neurocognitive deficits, where neuropsychologist administered a battery of tests to plaintiffs, including the WISC III "and selected subsets of the . . . Children's Memory Scale").

The defendants argue, and the trial court found, that, even if the BPA is reliable for clinical assessment, it is unreliable in the forensic context. The defendants assert that "Dr. Bruno-Golden[']s admi[ssion] that she combined quantitative data with her subjective, qualitative observations when reaching her conclusions," and her "admi[ssion] that her methodology . . . would not be appropriate or acceptable for a published research paper, . . . confirms the trial court's conclusion that the methodology was not sufficiently reliable for jury consideration."

20

The plaintiff, amicus curiae, Dr. Bruno-Golden, and Dr. Shaheen dispute that any different standard exists for the forensic setting as opposed to the clinical setting. According to the amicus curiae, "the clinical and `forensic' neuropsychologist are not distinguishable by their testing approach, by the scientific merits of their instruments, by balance of objectivity versus subjectivity, by standards of logical proof, or by training." Both "rely on situation-specific test batteries, and both strive to be objective and accurate in their characterization of an examinee's cognitive status. . . . Both neuropsychology roles . . . require scientific validation, but not fundamentally different kinds of scientific validation." We agree.

Besides Dr. Faust's unsupported opinion that clinical neuropsychologists are not objective examiners because their

apparent goal is to advance the patient's interest, nothing in the record indicates that the field of neuropsychology recognizes a relevant distinction between methodologies based upon their use in the clinical setting versus the forensic setting. Contrary to the defendants' assertion, Dr. Bruno-Golden's purpose was not to publish a research paper, which arguably would have required her to conduct a study similar to the one Dr. Faust has published, in which she standardized results against a control group. Instead, her purpose in evaluating the plaintiff was to determine the plaintiff's cognitive status; specifically, to "assess her current level of cognitive and behavioral functioning and serve as a basis from which to make recommendations with respect to her overall care, clinical management, and educational program."

To this end, Dr. Bruno-Golden administered a flexible battery based upon the plaintiff's presentation and history that consisted of standardized tests, scored those tests against published norms, and interpreted the results to determine the plaintiff's cognitive status. Although the BPA required Dr. Bruno-Golden to use and interpret both qualitative and quantitative measures to render her assessment, contrary to Dr. Faust's assertion, nothing in the record suggests that her goal as both a clinical and forensic neuropsychologist was not to render an objective determination of the plaintiff's cognitive status based upon those measures and her extensive clinical experience as a neuropsychologist. The subjectivity inherent in using the BPA to clinically assess a patient does not in and of itself render that methodology unreliable for determining whether a particular plaintiff in a legal proceeding suffers from or is at risk for cognitive or behavioral deficiencies. Cf. S.V. v. R.V., 933 S.W.2d 1, 42 (Tex. 1996) (Cornyn, J. concurring).

To be sure,

[t]he ability of the clinician to supplement the actual test scores with observation of the patient during the test, as well as background qualitative information

21

obtained from the patient, and the patient's records and family, are generally believed to allow clinicians to make more informed diagnoses than would be possible by a simple mathematical calculation based on test scores alone.

L. Russ et al., supra § 23:26, at 23-26. Thus, "[t]he clinical judgment of the neuropsychologist is a critical part of the forensic evaluation and should be used in conjunction with" quantitative test scores. McConnel, supra at 156-57; see also L. Russ et al., supra § 23:9, at 23-13 ("when the purpose of the examination goes beyond merely identifying and quantifying impairments," such as "[w]hen there is any question as to the cause of the identified impairments, whether for purposes of treatment or litigation, or as to the 'stability' of the impairments (i.e., is the patient likely to get worse, and if so, over what time frame?), the bare scores on the tests will not be sufficient to answer the question"). If the BPA as used by Dr. Bruno-Golden, which employs both quantitative and qualitative measures, is reliable to diagnose the plaintiff with particular injuries, and to prescribe her future clinical and educational care as a result of that diagnosis, we do not see how that same methodology is unreliable for assisting a fact finder in understanding the plaintiff's cognitive and behavioral status. See, e.g., Benedi v. McNeil-P.P.C., Inc., 66 F.3d 1378, 1384 (4th Cir. 1995) ("We will not declare . . . methodologies invalid and unreliable in light of the medical community's daily use of the same methodologies in diagnosing patients."); State v. McMullen, 900 A.2d 103, 118-19 (Del. Super. Ct. 2006) (finding that if a particular methodology "provides a sufficient basis on which to prescribe medical treatment with potential life-or-death consequences, it should be considered reliable enough to assist a fact finder in understanding certain evidence or determining certain fact issues"); cf. Heller v. Shaw Industries, Inc., 167 F.3d 146, 155 (3d Cir. 1999).

Indeed, Daubert simply requires that an expert "employ[] in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." Kumho Tire, 526 U.S. at 152. The record reveals no evidence suggesting that Dr. Bruno-Golden used a different methodology to evaluate the plaintiff from that which she normally uses in clinically assessing patients with lead paint exposure. Moreover, because the BPA as a flexible battery approach has a reliable basis in the knowledge and experience of neuropsychology, see id. at 149, any alleged shortcomings in Dr. Bruno-Golden's interpretation of the quantitative and qualitative results affect the weight to be given her conclusions, not the reliability of the methodology itself. Langill, ___ N.H. at ___; Dahood, 148 N.H. at 723. Accordingly, we reject the defendants' assertion that Dr. Bruno-Golden's methodology, the BPA as a flexible battery approach, is not a sufficiently reliable methodology to assist a fact finder in understanding the plaintiff's neuropsychological status.

22

C. Reliability of Dr. Bruno-Golden's Application of the BPA to This Case

We now review the trial court's finding that Dr. Bruno-Golden did not reliably apply the BPA to this case because she failed to follow the rules of administration for certain tests she used during her assessments. See RSA 516:29-a,

I(c). We review this finding for an unsustainable exercise of discretion. Langill, ___ N.H. at ___.

RSA 516:29-a, I(c) precludes expert testimony by a witness unless "[t]he witness has applied the principles and methods reliably to the facts of the case." In Langill, we adopted the following standard for determining when a witness has met this requirement:

when the application of a scientific methodology is challenged as unreliable under Daubert and the methodology itself is otherwise sufficiently reliable, outright exclusion of the evidence in question is warranted only if the methodology was so altered by a deficient application as to skew the methodology itself. Where errors do not rise to the level of negating the basis for the reliability of the principle itself, the adversary process is available to highlight the errors and permit the fact-finder to assess the weight and credibility of the expert's conclusions. . . . As long as an expert's scientific testimony rests upon good grounds, . . . it should be tested by the adversary process -- competing expert testimony and active cross examination -- rather than excluded from jurors' scrutiny for fear that they will not grasp its complexities or satisfactorily weigh its inadequacies.

Langill, ___ N.H. at ___ (quotations and citations omitted). Thus, for the testimony to be inadmissible, the flaws in application must "so infect the procedure as to make the results unreliable." Id. (quotation and brackets omitted).

In its order, the trial court specifically found that, in 2004, Dr. Bruno-Golden erred in administering the Block Design subtest of the WISC. The court also noted that the defendants had pointed to five other tests or subtests "in which Dr. Bruno-Golden created her own rules of administration or procedure," and that "Dr. Bruno-Golden could only identify two tests in 2002 and two tests in 2004 in which she followed the rules of administration." Thus, because following test administration rules maintains the validity and reliability of testing and Dr. Bruno-Golden did not abide by the rules for the

23

cited tests, the trial court found that "as administered in this case, Dr. Bruno-Golden's methodology was not reliable."

We address each cited error in turn to determine whether the errors rise to the level of negating the basis for the reliability of the methodology itself. Id.

1. WISC III and WISC IV

The WISC provides a global picture of the brain's processing. It assesses both the left and right hemispheres of the brain by incorporating subtests that analyze various verbal and non-verbal constructs. The subtests are individually scored in accordance with age norms contained in a manual. The manual also contains standards for administration of the test.

In 2002, when administering the WISC III, Dr. Bruno-Golden deviated from the manual's standards of administration for two subtests: Block Design and Object Assembly. In the Block Design subtest, a child is presented with a picture with a design on it and cubes that have the same design on them. The child is instructed to arrange the cubes so that they match the picture. Similarly, the Object Assembly subtest requires the child to complete a puzzle. Both subtests are timed tests in which the child is allotted more points for completing the tasks early, and receives a zero if she does not finish within the time limits.

On those occasions when the plaintiff did not complete the Block Design or Object Assembly tasks within the allotted time, Dr. Bruno-Golden gave the plaintiff a score of zero, but allowed her to continue trying to complete the tasks past the time limits. Dr. Bruno-Golden gave the plaintiff more time primarily because she did not want to disrupt her rapport with the plaintiff since the plaintiff had asked to continue the tasks, and secondarily because she wanted to observe whether, if given more time, the plaintiff would be able to finish the tasks. On one occasion, Dr. Bruno-Golden gave the plaintiff more time because the plaintiff was nearing completion of the task; on others, unless the plaintiff wanted to stop, she allowed her to continue.

Similarly, in 2004, when administering the WISC IV, Dr. Bruno-Golden deviated from the manual's administration standards for the subtests Block Design and Coding. The Coding subtest is administered after the Block Design subtest. In the Coding subtest, the child is presented with a code where the numbers one through nine are matched with specific geometric shapes. The numbers are then listed in a random order, and the child must fill in the corresponding geometric shapes. The child is first given a sample so that the child understands the task, and then is provided two minutes in which to complete the subtest. The Coding subtest is scored based upon what the child completes in the two-minute time limit.

24

For the Block Design subtest of the WISC IV, Dr. Bruno-Golden gave the plaintiff a score of zero, but again

allowed the plaintiff to continue after the time limits for the tasks had expired because the plaintiff was highly motivated to complete the tasks. Further, in the WISC IV version of the Block Design subtest, when a child obtains three consecutive scores of zero, the manual instructs the examiner to discontinue the subtest. Dr. Bruno-Golden allowed the plaintiff to continue with the fourth task, even though the plaintiff had received three consecutive scores of zero, because the plaintiff impulsively flipped the subtest over and insisted upon attempting the next problem. Dr. Bruno-Golden, however, did not score the additional time that the plaintiff spent on the subtest.

Similarly, for the Coding subtest of the WISC IV, Dr. Bruno-Golden allowed the plaintiff to continue for three minutes after the two-minute time limit. She scored, however, only the portion completed within the time limit. The Coding subtest on the WISC IV was the same as the Coding subtest for the WISC III.

The defendants argue that Dr. Bruno-Golden's failure to follow the time limits for the foregoing WISC subtests rendered her entire testimony unreliable. Relying upon the APA Standards, the trial court found that "the plaintiff ha[d] not shown that Dr. Bruno-Golden's decision to provide her with additional time to complete the [Block-Design] test was an approved change in test format, or more importantly, what effect the changes that Dr. Bruno-Golden made in test administration had upon the validity, reliability and appropriateness of norms."

The APA Standards allow an examiner to make an approved change in test format or mode of administration, but note that in these instances the examiner "should have a sound rationale for concluding that validity, reliability, and appropriateness of norms will not be compromised." APA Standards, supra at 117. The APA Standards comment:

In some instances, minor changes in format or mode of administration may be reasonably expected, without evidence, to have little or no effect on validity, reliability, and appropriateness of norms. In other instances, however, changes in format or administrative procedures can be assumed a priori to have significant effects. When a given modification becomes widespread, consideration should be given to validation and norming under the modified conditions.

Id. However, the APA Standards also explain:

25

When conducting psychological testing, standardized test administration procedures should be followed. When nonstandard administration procedures are needed, they are to be described and justified. . . .

One advantage of individually administered measures is the opportunity to observe and adjust testing conditions as needed. In some circumstances, test administration may provide the opportunity for skilled examiners to carefully observe the performance of persons under standardized conditions. For example, their observations may allow them to more accurately record behaviors being assessed, to understand better the manner in which persons arrive at their answers, to identify personal strengths and weaknesses, and to make modifications in the testing process. Thus, the observations of trained professionals can be important to all aspects of test use.

Id. at 120-21 (emphasis added).

Moreover, the WISC III manual states that if the child is nearing the completion of an item when the time limit expires, the child should be allowed to finish in the interest of maintaining rapport with the child. The administrator scores, however, only that work that is completed within the allotted time.

Finally, literature on the BPA indicates that in most cases, a patient is allowed additional time to complete the problem at hand when she is near a solution. W.P. Milberg et al., supra at 69. According to this literature, "[r]esponse slowing often accompanies brain damage, and its effects on test performance need to be examined separately from the actual loss of information-processing ability." Id. Thus, it is important to distinguish between patients who work too slowly from those who cannot complete problems no matter how much time is given. Id.

At the Daubert hearing, Dr. Shaheen, a qualified pediatric neuropsychologist, testified that a neuropsychologist may allow a child to continue past the time limits to observe the process the child uses to solve the problem and to see whether the child is persistent. These observations allow the clinician to determine whether the child is capable of succeeding on that task. Dr. Shaheen also testified that it is unknown what effect disrupting the rapport Dr. Bruno-Golden had with the plaintiff by not allowing her to complete the tasks would have had on subsequent tests. She agreed, however, that she

26

also did not know what effect extending the time limit on these subtests had upon the plaintiff's performance on later subtests, but opined that the additional time is unlikely to have affected the validity of the subsequent tests.

The foregoing evidence demonstrates that Dr. Bruno-Golden was permitted to exceed the time limits of subtests

when the plaintiff was nearing completion of a test, so long as she scored the plaintiff within the time limits. Dr. Bruno-Golden appears to have given the plaintiff more time to complete some of the WISC subtests even in those instances where the plaintiff was not nearing a solution. However, given that Dr. Bruno-Golden always scored the plaintiff in accordance with the time limits, and that she gave the plaintiff more time for the permissible purposes of maintaining her rapport with the plaintiff, sustaining the plaintiff's motivation, and observing the plaintiff's behavior, such a slight deviation from the rules of administration does not rise to the level of negating the basis of reliability for the BPA itself. Langill, ___ N.H. at ___.

Indeed, both the WISC manual and the BPA literature emphasize the importance of an examiner's rapport with the child. In 2002 and 2004, Dr. Bruno-Golden administered the WISC tests in the morning, and second only to the human figure drawing. Thus, since Dr. Bruno-Golden had planned for a full day of tests, it would have been particularly important for Dr. Bruno-Golden to maintain her rapport with the plaintiff at this early stage.

Further, the BPA itself includes a qualitative component that requires the examiner to observe a child's behavior and draw conclusions from that behavior based upon the examiner's expertise and published literature. See, e.g., W.P. Milberg et al., supra at 67-69. The APA Standards also note the importance of an examiner's observations for the accurate recording of the patient's behavior. APA Standards, supra at 121. Thus, in the context of the BPA, Dr. Bruno-Golden's failure to stop the plaintiff at the time limits for two of the WISC III and two of the WISC IV subtests could not have so infected the procedure as to skew the reliability of the BPA itself. Langill, ___ N.H. at ___. Rather, it was for the jury to resolve the dispute in the testimony and determine what effect, if any, the increased time on the subtests had on the results of the subsequently administered subtests. Id. Accordingly, the trial court erred in excluding Dr. Bruno-Golden's testimony based upon her deviation from the WISC time limits.

2. NEPSY

The NEPSY test is a battery of neuropsychological tests specifically designed for children. Korkman et al., NEPSY, A Developmental Neuropsychological Assessment: Manual 45 (The Psychological Corp. 1998). It consists of numerous subtests that examine various cognitive domains,

27

including memory. Id. at 45-48. Each subtest has been individually standardized and is independently scored.

In the 2002 afternoon session, Dr. Bruno-Golden selected and administered certain NEPSY subtests designed to measure verbal working memory. She did not administer all the core and expanded subtests for memory as "recommended" by the manual. Id. at 46. Otherwise, Dr. Bruno-Golden followed the manual's instructions for administration. In the 2004 afternoon session, Dr. Bruno-Golden repeated the NEPSY subtests she had administered in 2002, except for those subtests that were no longer age-appropriate. In those instances, she selected other tests that examined the same constructs.

The trial court based its decision in part upon Dr. Bruno-Golden's failure to "administer[] the entire NEPSY battery as recommended in the manual." The NEPSY manual states, in pertinent part:

The NEPSY can be used as an assessment tool at a variety of levels. Subtests are selected on the basis of age, the referral question, the needs of the child, time constraints, and the setting in which the assessment takes place. A Core Assessment, which is composed of selected subtests from each domain, provides an overview of a child's neuropsychological status. An Expanded or Selective Assessment allows a more thorough analysis of specific cognitive disorders, and consists of selected subtests--generally beyond those in the Core. The following sections present recommendations for subtest selection and levels of assessment. A comprehensive neuropsychological evaluation can be completed using the Full NEPSY, which includes all subtests. There is no prescribed set of subtests that must be administered to every child.

. . . .

When a Core Domain Score, a subtest or Supplemental Score, the referral question, or a previous diagnosis indicates the presence of a problem in a certain domain, it is recommended that an *Expanded Assessment*--the administration of all Core and Expanded subtests in a domain--be administered in order to investigate the problem in greater depth. .

. .

28

A *Selective Assessment* involves choosing additional subtests across domains as part of an evaluation. When the Core Assessment suggests the presence of a disorder of a complex function that may involve or affect components from several domains, continuation of testing using the pertinent Expanded subtests as well as additional subtests from other domains is recommended. . . . In these cases, the assessment involves the administration of subtests that

assess subcomponents of the capacity in question.

Id. at 45-47 (emphases added). The NEPSY manual also delineates a recommended order in which subtests should be administered if they are individually selected for the particular child being tested. Id. at 50-51.

Dr. Bruno-Golden administered several NEPSY subtests, but did not administer the Full NEPSY or the entire battery for each domain she tested. However, Dr. Bruno-Golden explained that she selected the particular subtests based upon the results of the morning tests. The NEPSY manual specifically allows a neuropsychologist to select pertinent subtests examining the domain in question when the core assessment suggests the presence of a particular issue.

Dr. Bruno-Golden used the WISC III to obtain the plaintiff's "core assessment"; that is, an overview of the plaintiff's neuropsychological status. Based upon this core assessment, Dr. Bruno-Golden selected pertinent NEPSY subtests that examined specific domains that required further analysis. For example, Dr. Bruno-Golden selected the Sentence Repetition subtest to measure verbal working memory for structured linguistic material because, during the morning session, the plaintiff had frequently requested that Dr. Bruno-Golden repeat herself and the directions. While Dr. Bruno-Golden did not completely follow all the recommendations of the NEPSY manual, contrary to the defendants' assertion, she did not contravene the "manual requirements." Accordingly, her failure to administer the Full NEPSY was not an administration error that could have altered the reliability of the methodology. Instead, it was for the jury to assess the weight of Dr. Bruno-Golden's testimony in light of her failure to follow the NEPSY manual's recommendations. Langill, ___ N.H. at ___; Dahood, 148 N.H. at 723. Thus, the trial court's reliance upon this failure was misplaced.

3. Wide Range Assessment of Memory and Learning (WRAML)

The WRAML is a battery designed to examine memory and learning functions. D. Sheslow & W. Adams, Wide Range Assessment of Memory and

29

Learning: Administration Manual 13 (1990). In addition to the complete battery, a shorter form of the WRAML consisting of four subtests may also be administered for screening purposes. Id. The WRAML manual instructs the examiner to administer the standard battery in the order presented. Id. at 13-14. It further states:

Preliminary investigations suggest that for many of the subtests, the level of performance is not significantly affected if taken out of sequence or if a specific subtest is administered alone. However, until such investigations are complete, the examiner is encouraged to administer the test in the order presented in this manual. If the examiner is using the Screening Form of the battery, administer only the first 4 WRAML subtests.

Id. at 14 (emphasis added).

In 2002, Dr. Bruno-Golden administered only the Story Memory subtest of the WRAML. The Story Memory subtest consists of two stories and examines the recall of narrative information. The trial court cited as an administration error Dr. Bruno-Golden's administration of only portions of the WRAML. It also found that Dr. Bruno-Golden "did not record the time limits she used, with the result that she could not say whether or not she had adhered to the time limits in the WRAML manual."

While the WRAML manual encourages an examiner to administer the entire WRAML in the order presented, it does not require the examiner to administer the entire WRAML or the screening version of the WRAML. Further, the WRAML was not a critical test upon which Dr. Bruno-Golden based her conclusions. Thus, even if Dr. Bruno-Golden erroneously administered only one of the subtests instead of the entire battery or the screening portion, this error could not have risen to the level of negating the basis for the reliability of the BPA itself. Langill, ___ N.H. at ___. Again, as with the NEPSY, it was for the jury to determine the weight of Dr. Bruno-Golden's testimony given her failure to follow the WRAML manual's recommendations. Id.

Moreover, the record does not support the trial court's finding that Dr. Bruno-Golden may not have adhered to certain time limits in the WRAML manual. The relevant portion of the WRAML manual provides:

Do observe these additional guidelines when using the WRAML:

30

1. Use the directions exactly as written. Do not paraphrase, adapt or add.
2. Adhere to the time limits associated with the 2 timed subtests (Picture Memory and Design Memory). . . . On no subtest is the child's performance timed, but the Picture Memory and Design Memory subtests require the examiner to expose materials for a specified amount of time. These time intervals should be exact. . . .

Sheslow & Adams, *supra* at 14.

Dr. Bruno-Golden only administered the Story Memory subtest of the WRAML, not the Picture Memory or Design Memory subtests. Thus, it is not clear that the WRAML manual required her to adhere to certain time limits, particularly since the manual specifically states that the child's performance is not timed on any subtest except the Picture Memory and Design Memory subtests. The record reveals no evidence indicating that the Story Memory subtest had precise time limits. Notably, Dr. Faust did not cite this alleged failure as an administration error. Accordingly, since no evidence indicated that the Story Memory subtest was subject to time limits, the trial court unsustainably exercised its discretion in finding that Dr. Bruno-Golden failed to adhere to "the time limits in the WRAML manual."

4. ROCF

In 2002 and 2004, pursuant to her usual practice, Dr. Bruno-Golden administered the ROCF test to the plaintiff after the WISC. As stated earlier, the ROCF test examines visual construction skills and non-verbal memory. The ROCF manual instructs:

Administer verbal tasks to the respondent during the interval between completion of the Immediate Recall trial and the Delayed Recall trial. Tasks involving visuospatial stimuli should not be administered between the Copy trial and the Delayed Recall trial. It is important that the respondent is engaged and actively performing a verbal task during the delay interval.

Meyers & Meyers, *supra* at 8. As examples of verbal tasks, the manual cites "time estimation, controlled verbal fluency, temporal orientation, or . . . clinical interview[s]." *Id.* Examples of tasks involving visuospatial stimuli include "the

31

Benton Judgment of Line Orientation" and "the Visual Reproduction subtest of the WMS-R." *Id.*

During the delay interval, Dr. Bruno-Golden first obtained a written language sample from the plaintiff by handing her a complex standardized picture from an aphasia screening exam, called the Cookie Theft Test, and instructing her to write a story about the picture. Dr. Bruno-Golden testified that she used the Cookie Theft Test picture as an academic screening measure to observe whether the plaintiff's written language arts and written language expression were commensurate with her grade level and chronological age. She did not administer or score the Cookie Theft Test itself. Dr. Bruno-Golden also administered the CCPT during the interval. This test measures attention and required the plaintiff to click a button on a computer when a letter other than "X" flashed on the screen.

Dr. Bruno-Golden testified that neither test was "visuospatial." She explained that a visuospatial task would have involved giving the plaintiff "a task where she had to do visual construction skills, or a psycho or a visual spatial task." She testified that the task must not disrupt the child's visual construction of spatial relationships.

Dr. Faust testified that Dr. Bruno-Golden violated the ROCF manual when she used the Cookie Theft Test to obtain a written language sample and administered the CCPT during the interval between the recall trials. He disagreed with Dr. Bruno-Golden's interpretation of the term "visuospatial," and opined that visuospatial "means basically visual stimuli." Based upon Dr. Faust's testimony, the defendants assert that Dr. Bruno-Golden erred by administering a "visuospatial" test during the thirty-minute interval.

In its order, the trial court summarized the defendants' argument that "Dr. Bruno-Golden administered inappropriate tests following first half of the [ROCF] test," and summarized some of Dr. Bruno-Golden's testimony relating to her administration of the ROCF. The trial court, however, never adopted the defendants' contention and resolved the conflicting testimony to actually find that Dr. Bruno-Golden improperly administered the ROCF. Accordingly, because the trial court did not find that Dr. Bruno-Golden erred in administering the ROCF, we decline to consider this alleged error on appeal.

5. Finger Tapping and Grip Strength Tests

The Finger Tapping and Grip Strength tests are two subtests of the Halstead-Reitan battery that measure pure motor function. In 2002 and 2004, Dr. Bruno-Golden administered both the Finger Tapping and Grip Strength tests as part of her hypothesis testing to rule out potential issues with pure motor function.

32

The Finger Tapping test measures the number of times a child taps certain fingers on each hand in ten seconds. The results are compared with standardized rates of other individuals in the same age group. Although neither party has provided the Finger Tapping test manual, the testimony elicited at the Daubert hearing seems to indicate that the

manual requires the examiner to have the child do five consecutive "finger tap trials" that are within "five points" of each other. Each "finger tap trial" consists of the examiner counting the number of taps the child does in ten seconds. Consecutive "finger tap trials" are within "five points" of each other when the number of taps in each ten-second period differs by five or less. For example, if on the first finger tap trial, the child taps his or her finger twenty-two times, on the second trial, taps his or her finger twenty-eight times, and on the third trial, taps his or her finger thirty times, the first two consecutive finger tap trials are not within five points of each other, but the second two consecutive finger tap trials are within five points of each other.

Either in 2002 or 2004, or in both years, Dr. Bruno-Golden asked the plaintiff to perform only three finger tap trials on one of her hands because she consistently went slower each time. Dr. Bruno-Golden also may not have properly obtained five consecutive finger tap trials on the plaintiff's left hand that were within five points of each other.

The Grip Strength test measures a child's strength, which in turn provides information concerning the behavior of the left and right hemispheres of the brain. For the Grip Strength test, the manual instructs the examiner to perform the test two times unless the first is not properly performed. In either 2002 or 2004, or in both years, Dr. Bruno-Golden administered the test to the plaintiff three times and then averaged the results.

In its order, the trial court noted that the defendants "contend[ed] that Dr. Bruno-Golden failed to follow the instructions with regard" to these tests, but never explicitly found that the tests were administered improperly. As with the ROCF, because the trial court never found that Dr. Bruno-Golden improperly administered either the Finger Tapping or the Grip Strength test, we decline to rely upon these alleged errors on appeal.

Moreover, neither the Finger Tapping nor the Grip Strength tests were critical to Dr. Bruno-Golden's conclusions. Indeed, the plaintiff performed within normal limits on both tests, and Dr. Bruno-Golden used the tests to rule out potential issues. Accordingly, even if Dr. Bruno-Golden administered these tests improperly, such errors did not skew the methodology itself. Langill, ___ N.H. at ___.

33

6. Collective Impact of Errors

The defendants assert in their brief, and the trial court ultimately found, that "Dr. Bruno-Golden could only identify two tests in 2002 and two tests in 2004 in which she followed the rules of administration." This finding is based upon Dr. Bruno-Golden's answer to a long, multi-part question posed by defense counsel. The question was:

Can you identify for me any test that you administered in 2002, 2004, that is current, that is up-to-date, that was administered entirely, that is, administered all the sub-tests as required by the test publisher, that were administered according to the manual, without deviation? That were recorded correctly and according to the manual, and that is, we don't have to rely upon your word to tell us that the child completed the test or didn't complete the test? That were scored correctly and that you applied the proper norms and appropriate description to the performance of the child?

Ultimately, after some clarification of the question and defense counsel discounting certain tests for not meeting the question's requirements, Dr. Bruno-Golden answered: "Well, [in 2002,] there's the Peabody picture vocabulary test, which we haven't discussed. . . . [and] the California verbal learning test," and, in 2004, the "Peabody picture book vocabulary test, . . . [and] the Hooper [Visual Organization Test]."

Although defense counsel's excellent cross-examination certainly highlighted in a summary fashion weaknesses in Dr. Bruno-Golden's testimony, Dr. Bruno-Golden's answer to this question, while perhaps technically correct, is misleading. Her answer considers issues relating not simply to the methodology, but also to her ultimate conclusions, such as errors in score calculation, which the trial court properly did not consider as administration errors. Such issues affecting the weight of the evidence are better left to the determination of the fact finder. Langill, ___ N.H. at ___. Dr. Bruno-Golden's answer also includes reference to alleged errors that may in reality not have been errors at all (e.g., the question presumes that Dr. Bruno-Golden was required to administer all the subtests of a particular test battery), and to tests that were not critical to her conclusions. Thus, the trial court erred in relying upon this misleading answer to support its finding that Dr. Bruno-Golden's administration errors rendered her testimony unreliable.

Furthermore, even considering Dr. Bruno-Golden's administration errors as a whole, the evidence does not support a finding that these errors so altered

the BPA as to skew the methodology itself. *Id.* Notably, the evidence reveals that Dr. Bruno-Golden's only arguably true error was exceeding the time limits for certain of the WISC subtests. As for the other errors cited by the defendants, either: (1) the trial court did not explicitly find that they were errors; (2) they occurred on tests that were not critical to Dr. Bruno-Golden's conclusions, *i.e.*, tests that she used to eliminate certain hypotheses; or (3) the evidence did not support a finding that they were in fact errors. Because we have already found that Dr. Bruno-Golden's errors on the WISC subtests did not render the entire methodology unreliable, we reverse the trial court's finding that Dr. Bruno-Golden unreliably applied the methodology to the facts of this case, see RSA 516:29-a, I(c), and its orders excluding her testimony and dismissing the plaintiff's writ.

IV. Conclusion

The trial court excluded Dr. Bithoney's testimony because his testimony was based almost entirely upon Dr. Bruno-Golden's proposed testimony, which it found to be unreliable. Given our conclusion that Dr. Bruno-Golden's testimony is reliable and admissible, we do not reach the plaintiff's remaining arguments concerning Dr. Bithoney's testimony. We vacate the court's ruling excluding Dr. Bithoney's testimony and remand for further proceedings consistent with this opinion.

Reversed in part; vacated in part; and remanded.

BRODERICK, C.J., and DALIANIS and GALWAY, JJ., concurred.

Footnotes:

FN1. We note that, although this particular source is not in the record, consistent with the practice of other courts, we may consider it to fully understand the science at issue in this case. See, e.g., Ballew v. Georgia, 435 U.S. 223, 231-32 n. 10, 232-39 (1978) (noting that although only "some" of the studies "ha[d] been pressed upon [the Court] by the parties," the Court considered all of the cited sources "carefully because they provide[d] the only basis, besides judicial hunch, for a decision"); State v. O'Key, 899 P.2d 663, 682, 686 (Or. 1995) (en banc) (evaluating the reliability of the Horizontal Gaze Nystagmus test based partly upon its "own research" and "numerous other sources" not in the record); see also Monahan & Walker, Judicial Use of Social Science Research, 15 Law & Hum. Behav. 571, 571 (1991) ("Increasingly in recent decades, courts have sought out research data on their own when the parties have failed to provide them."). But see Ballew, 435 U.S. at 246 (Powell, J. concurring) (criticizing majority's "heavy reliance" upon studies not "subjected to the traditional testing mechanisms of the adversary process.").