Behavioral Sciences and the Law
Behav. Sci. Law **17**: 543–551 (1999)

Challenging a Flexible Neuropsychological Battery under *Kelly/Frye*: A Case Study

R. K. McKinzey, Ph.D,* and
Thomas G. Ziegler, J.D.

The ability of the flexible neuropsychological battery approach to withstand a challenge under California's evidentiary standard, *Kelly/Frye*, was tested in an actual trial. Despite repeating many criticisms offered by the literature (e.g., no malingering measures, unknown accuracy rates, ignoring statistical limitations, not using age norms), the battery was allowed in "for the weight of the evidence," rather than being excluded as unreliable. Copyright © 1999 John Wiley & Sons, Ltd.

In *Behavioral Sciences and the Law*, Joe Reed (1996) reviewed the effect the *Daubert* standard (*Daubert v. Merrell Dow Pharmaceuticals Inc.*, 1993) might have on the admissibility of neuropsychological evidence. He presented Ralph Reitan's testimony in *Chapple v. Ganger* (1994), noting that the case seemed to be the first time the issue of fixed versus flexible battery had been raised under *Daubert*. The court accepted the fixed battery and only portions of the flexible battery. Reed argued:

> Professionals who use the validated or fixed neuropsychological test batteries to obtain reliable and valid objective test results will generally find the *Daubert* standard an easy threshold to pass; however, professionals who use only flexible neuropsychological batteries to obtain valid and reliable objective test results will find the *Daubert* standard more imposing, if not impossible to pass. (p. 321)

In a coda, Reed then observed that California has kept to the *Frye* standard (*People v. Leahy*, 1994). "Consequently, in California general acceptance from the relevant scientific community is still a major criterion in determining the admissibility of scientific evidence, and thus in any California state court results obtained from flexible or standardized neuropsychological test batteries can be considered scientific." (p. 321)

In 1998, a flexible battery was offered as part of the defense evidence in a California criminal trial. We decided to test the assumption that the flexible battery approach would find easy acceptance when challenged.

* Correspondence to: R. K. McKinzey, Ph.D., 3045 Telegraph, Oakland, CA, USA; e-mail address: rkm@slip.net

CCC 0735–3936/99/040543–09$17.50

FACTS OF THE CASE

On 7 March 1997, Charles Sebastianelli (date of birth 17 May 1930) was hit on the head with a milk carton crate falling from an overhead storage area. He refused to call an ambulance, and was taken by private car to an emergency room, where he denied loss of consciousness (LOC). His Glasgow Coma Scale was 15. His scalp cuts were cleaned and stitched, and he drove himself home.

On 27 March 1997, Mr. Sebastianelli solicited the murder of a business partner, as well as the partner's wife and son. In fact, he had been asking the would-be hit man (DB) to take his assignment for six months. When he became more insistent, DB went to the police, who wired DB during the two planning conversations. In the recording, Mr. Sebastianelli can be heard calmly discussing the timing and place of the murders, the price of the hit, and which of several guns was to be used.

Mr. Sebastianelli was arrested and charged with a violation of California Penal Code Section 653f, solicitation of murder, which is a specific intent crime. There were allegations of a special circumstance, the possession of two assault weapons found during a search. During his interrogation, Mr. Sebastianelli denied all, and then invoked his Miranda rights.

Mr. Sebastianelli was referred in August to a neuropsychologist, Barbara Norton, Ph.D. He complained of LOC on 7 March 1997, and of a headache, dizziness, blurred vision, and poor memory on the next day. Medical history included past back surgery, diabetes, coronary artery disease, and obesity. He complained of continued poor memory, dizzy spells, tinnitus, and a sense of not being fully awake. His sister was doing the driving, as Mr. Sebastianelli was getting lost and confused. He had difficulty concentrating on reading and was less fluent. On mental status, he described reduced confidence, poor appetite, weight loss, insomnia, early morning wakening, depression, anxiety, and reduced motivation. Although he denied psychotic symptoms, he noted hallucinatory experiences and irritability when incarcerated. Social history included a year of college, a brief marriage, and years of real estate development.

Mr. Sebastianelli argued that DB had manipulated him into the solicitation, as "he was not thinking right at the time." "He reiterates that [DB] set him up."

Dr. Norton administered a flexible neuropsychological battery: the tests (including acronyms) and results are presented in Table 1. She concluded that there were "some significant abnormalities in both cognitive and emotional functioning" of a mild nature (footnotes added):

> Performance on tests of attention and concentration was abnormally variable.[1] Language functioning tests suggested mild impairment in language comprehension and more serious impairment in expressive language.[2] Tests of new learning of both verbal and nonverbal material showed significant abnormalities with particularly impaired performance in memory of paragraph length stories[3] and abnormal loss of visual material after a 30 minute delay.[4] Mr. Sebastianelli also performed abnormally on a test of abstraction and concept formation, and this

[1] That is, an item of Mental Control (counting backwards from 20) was slow.
[2] That is, the BNT (naming fluency) score was low.
[3] WMS-R Logical Memory immediate and delayed.
[4] WMS-R Visual Reproduction delayed.

Table 1. Dr. Norton's evaluation of Charles Sebastianelli

Function	IQ equivalent						
	60	70	80	90	100	110	120
Overall intellectual functioning							
Verbal IQ[a] = 88				X			
Performance IQ[a] = 93					X		
Full scale IQ[a] = 89				X			
ATTENTION & CONCENTRATION							
Fixed/simple							
Digits forward[a] = 81			X				
Visual Memory forward[b] = 92				X			
Trails A = 98					X		
Mental tracking							
Digits backward[a] = 84			X				
Visual Memory backward[b] = 114							X
Mental Control[b] = 75		X					
Trails B = 87				X			
Complex							
Digit Symbol[a] = 95					X		
SDMT = 88				X			
LANGUAGE FUNCTIONS							
Receptive							
Tokens[d] = 93				X			
Expressive							
BNT = 67		X					
COWAT = 81			X				
Verbal abstraction							
Similarities[a] = 95					X		
Repetition							
Sentence repetition[d] = 90				X			
VISUAL/PERCEPTUAL/CONSTRUCTIONAL ABILITIES							
Visual/perceptual							
Picture completion[a] = 95					X		
HVOT = 104						X	
Visual constructional							
Block Design[a] = 100					X		
Rey–Osterrieth figure = 90				X			
MEMORY & NEW LEARNING							
Remote memory							
Information[a] = 100					X		
New learning—verbal							
Verbal Paired Associates immediate[b] = 84			X				
Verbal Paired Associates delayed[b] = 90				X			
Trial 5 (no interference)[c] = 106						X	
Trial 7–interference[c] = 104						X	
Logical Memory immediate[b] = 78			X				
Logical Memory delayed[b] = 78			X				
New learning—Nonverbal							
Visual Reproduction immediate[b] = 114							X
Visual Reproduction delayed[b] = 83			X				
SOCIAL REASONING & JUDGMENT							
Comprehension[a] = 85				X			
Picture Arrangement[a] = 105						X	
COMPLEX PROBLEM SOLVING							
WCST perservative R = 76		X					
ACADEMIC/ACHIEVEMENT							
WRAT–R Arithmetic = 95					X		

Note: This table is taken from Dr. Norton's records, with minimal modifications. Xs are not meant to be exact, and are only provided as a "visual summary of the pattern of performance." Tests given are partial Wechsler Adult Intelligence Scale—Revised (WAIS–R), partial Wechsler Memory Scale—Revised (WMS–R; indices were not calculated), Rey Auditory Verbal Learning (RAVLT), Multilingual Aphasia Examination (MAE), partial Wide Range Achievement Test—Revised (WRAT–R), Trail Making Test (A&B), Symbol Digit Modalities Test (SDMT), Boston Naming Test (BNT), Hooper Visual Organization Test (HVOT), Wisconsin Card Sorting Test (WCST), Controlled Oral Word Association Test (COWAT), Rey–Osterrieth Figure Drawing, and Coolidge Axis II Inventory. Barbara Norton is a pseudonym.
[a]WAIS–R.
[b]WMS–R.
[c]RAVLT.
[d]MAE.

> test[5] is particularly sensitive to brain dysfunction ... [His] pattern of performance on tests of cognitive functioning is suggestive of mild brain dysfunction.

Dr. Norton opined that the dysfunctions were due to the 7 March concussion. She noted the presence of the brain deficits was more serious because of Mr. Sebastianelli's age, health condition, and previous injury (a head blow at age 3). It would have been worse immediately afterwards with slow improvement. ''Since he is continuing to show deficits at present, he would have had considerably more significant upset in his thinking ability in the first weeks following his injury.''

Dr. Norton noted depression, anxiety, and a mild thought disorder. She noted headaches, neck, and shoulder pain are common consequences of head trauma, which can also lead to significant depression, and recommended supportive psychotherapy and medication, providing the former.[6]

Defense counsel offered Dr. Norton's report and proposed testimony as part of the defense theory that Mr. Sebastianelli's concussion prevented him from attaining the required level of intent. The doctor's records and conclusion were therefore closely reviewed for rebuttal testimony potential.

There were many problems with Dr. Norton's testing and conclusions. Given Reed's arguments, we decided to challenge the admissibility of Dr. Norton's testimony.[7]

RELEVANT CASE LAW

The California Supreme Court adopted *Frye* in *Kelly* (*Frye v. U.S.*, 1923; *People v. Kelly*, 1976), and explicitly adhered to *Frye* rather than *Daubert* in *People v. Leahy*, 1994. In *Leahy*, the court returned to *Kelly* in rejecting the approach which left:

> questions of admissibility to the discretion of the trial court in the first instance ''in which event objections, if any, to the reliability of the evidence (or of the underlying scientific techniques on which it is based) might lessen the weight of the evidence but would not necessarily prevent its admissibility. (p. 31)'' (p. 599).

In *Adams*, ''the proponent of the evidence must demonstrate that current scientific procedures were used in the *particular* case'' (*People v. Adams*, 1975, as cited in *Kelly*, p. 31, emphasis added). On the other hand, in both *People. v. McDonald* (1984) and *People v. Stoll (1989)*, the court refused to consider psychiatric/psychological evidence to need review: ''We have never applied the *Kelly–Frye* rule to expert medical testimony.'' (*People v. McDonald*, 1984, p. 374) Nevertheless, we argued, under *Adams* the judge had to rule on whether *this* doctor had done an admissibly proper evaluation, even if the general field of neuropsychology did not need to be reviewed.

[5] WCST.

[6] The appropriateness of this dual role is beyond the scope of this article.

[7] We had tactical considerations as well. If the judge ruled in favor of the prosecutor, an appellate issue would be raised, but not conversely. The local judges were known to rule in ways to avoid providing such a chance of being overturned. However, we had our own risk/benefit consideration: the evidentiary hearing would provide us an otherwise unavailable chance for discovery, and might eviscerate the defense.

THE HEARING

In a pretrial motion, the admissibility of Dr. Norton's testimony was challenged under California's statute for evidence relevancy, EC Section 402. The defense had the burden of proof, preponderance of the evidence. In the 402 hearing, we planned to make five points:

(1) The doctor's testing had no control for faking.
(2) It was done without known accuracy rates.
(3) The doctor ignored the proper statistical checks.
(4) The doctor did not use the proper norms.
(5) The doctor ignored the literature in interpretation.

Therefore, her testimony should be excluded as not "reliable" or scientifically done, and was likely to unduly sway the lay jury with the mantle of "expert".

During questioning, Dr. Norton testified that she administered widely accepted tests, using the flexible battery approach taught her at the University of California, and as propounded by Muriel Lezak (Lezak, 1995), which requires the presence of a coherent pattern of scores, rather than merely one abnormal score:

> Marked quantitative discrepancies in a person's performance suggest that some abnormal condition is interfering with that person's overall ability to perform at the characteristic level of cognitive functioning ... Any single discrepant score or response error can usually be disregarded as a chance deviation. A number of errors or test score deviations, however, may form a pattern that can then be analyzed in terms of whether it makes neurological sense. (Lezak, 1995, pp. 165–166)

The doctor explained that she had found that pattern in the deficits of concentration, recent memory, verbal fluency, and complex problem solving. Dr. Norton gave 13 tests (three of them partial), used means and standard deviations to produce equivalent standard scores, and compared them all to each other. This procedure yielded 88 comparisons.

The doctor agreed that there was some literature on the need to control for neuropsychological malingering (cf., Faust, Ziskin, & Hiers, 1991), but did not use any test designed for that purpose since she found the literature unsettled. She did note that:

(1) She gave a personality test, the Coolidge Axis II Inventory, which did not have an elevated validity scale. On rebuttal, we pointed out that such personality measures cannot be used to rule out neuropsychological faking. We did not add that the literature on the point has been settled for the last two decades (Faust *et al.*, 1991; Heaton, Smith, Lehman, & Vogt, 1978; Rogers, Harrell, & Liff, 1993).
(2) The patient was unlikely to be faking since he had produced a welter of normal scores. We replied that this method too was well known to be invalid, as many studies of malingerers had substantial numbers of subjects who produced normal test scores even when trying to fake (see, e.g., Heaton *et al.*, 1978; McKinzey, Podd, Krehbiel, Mensch, & Trombka, 1997).
(3) The patient could not possibly be faking since the doctor had found a pattern that fit the injury. We replied that this method of ruling out faking has never been validated. In discussing the problem, Lezak recommends a specific procedure (Rey's Fifteen Item Test) be used instead.

The doctor had no answer to the problem of unknown accuracy rates. As Dr. Reitan argued (*Chapple v. Ganger*, 1994; Reed, 1996), the biggest problem the flexible battery approach has is its complete lack of known accuracy rates. The name derives from Lezak's recommendation that each patient requires a different set of tests to explore hypotheses about the patient's deficit. Thus, the set of tests given (the battery) is flexible, rather than a fixed set. However, as Halstead found five decades ago, the more tests given, the higher the probability that some of the tests would yield abnormal results, as can be seen in the manual to one set of Halstead–Reitan Battery (HRB) norms (Heaton, Grant, & Matthews, 1991, p. 37). This tendency can be controlled for when a fixed battery is given to large numbers of normal people and to those with neuropsychological deficits, but the flexible battery, by definition, cannot do so (Faust *et al.*, 1991; Russell, 1998). Since this particular battery has never been given to a large group of normals, no one knows how many of them might produce a "pattern of scores" that the doctor would find significant. The doctor was thus testifying to a scientific technique that is without a known accuracy rate, surely not a generally accepted procedure.

In fact, some of the comparisons the doctor made are known to be unusable. For example, she compared the WAIS–R subtests to each other (as well as to the other test scores), a commonly used procedure known as "scatter analysis". However, large differences amongst these subtests are quite common in normal people (i.e., the false positive rate is high) (Matarazzo, Daniel, Prifitera, & Herman, 1988). She also compared the subtests of the WMS–R to each other, a procedure the manual cautions against, since the subtests have uneven test–retest reliabilities (Wechsler, 1987). In technical terms, she made comparisons without allowing for different standard errors of measurement (SEMs). There are also no studies validating such subtest comparisons.

Mr. Sebastianelli was 67 at time of testing. All of the tests used have norms corrected for age, but the doctor did not always use them! The doctor used the WCST (Heaton, 1981; Heaton, Chelune, Talley, Kay, & Curtiss, 1993) as a measure of complex problem solving. This test has two manuals: the doctor used the older of them. This older manual has two sets of norms: the doctor used the set for younger patients, *despite explicit manual instructions to the contrary*. She also did not use the more recent age-corrected norms. When either of the age norms were used, Mr. Sebastianelli's WCST score became normal.

The doctor used the BNT as a measure of verbal fluency. The doctor used the non-age-corrected norms in the manual, despite three age-corrected sets being available (Gorp, Satz, Kiersch, & Henry, 1986; Neils *et al.*, 1995; Welch, Doineau, Johnson, & King, 1996). Again, when age-corrected norms were used, Mr. Sebastianelli's BNT score became normal.

The doctor used the WMS–R Mental Control as a measure of concentration. The manual provides some age norms done on small samples. However, the MOANS project (Ivnik *et al.*, 1992) provides much larger, better structured norms. Mr. Sebastianelli's Mental Control score became normal when these age norms were used.[8]

[8] Multiple norms for neuropsychological tests is another problem for forensics. The problem has been solved for one fixed battery, the Luria–Nebraska (McKinzey, Roecker, Puente, & Rogers, 1998), but not for the Halstead–Reitan (Golden & Broek, 1998) or the flexible battery (McKinzey, 1997).

Out of the 88 comparisons made, only two remained abnormal, the WMS–R Logical Memory immediate and delayed, which the doctor used as measures of recent memory. However, these have been identified as the most likely to be elevated in litigating patients (Fox, 1994). When these two tests are therefore ignored, the entire test battery becomes quite normal.

Given these glaring deficiencies, we argued, the doctor's testing was not done properly, according to the profession's generally accepted scientific principles. Her opinions, which relied on her testing, were therefore equally unreliable, and should not be allowed.

The judge refused to agree. He pointed out that both neuropsychologists (Drs. Norton and McKinzey) had impressive vitae, so how was he to judge between them? But, we countered, if he could not distinguish between good and bad science after a hearing, how was a lay jury to do so? He remained firm, and allowed the doctor's testimony, noting that the jury would consider "the weight of the evidence."

THE CASE AT TRIAL

In trial, Dr. Norton testified as to the findings of her report, and repeated her assertion that Mr. Sebastianelli had a concussion. She did not correct her misscoring.

In rebuttal, Dr. McKinzey opined that Mr. Sebastianelli did not have a concussion, and was fully capable of forming the necessary intent to solicit the deaths of all three people. He offered three independent reasons the defense theory should be rejected:

(a) Dr. Norton had not used any accurate method of ruling out faking, despite several methods available.
(b) He also noted that, when properly scored, the test battery became completely normal.
(c) More importantly, the transcript of the taped solicitation did not show any signs of the deficits Dr. Norton had suggested. There were no signs of poor concentration, word finding, fluency, or grammar problems. Mr. Sebastianelli was able to keep track of various conversational threads, and was able to discuss his personal and national history, including recent events. He could give and take instructions. He was driving without getting lost, contrary to his self-report. There were no signs of manipulation by the would-be hit man. There were no signs of impulsivity.

The jury returned a verdict of Guilty of Solicitation on only one count, that relating to the business partner, and refused to convict Mr. Sebastianelli of soliciting the murders of the wife and son. In post-trial interviews, they said they wanted to let him go altogether (he was, after all, just an old man, and hadn't actually hurt anyone[9]), but there was simply too much evidence against him.

Mr. Sebastianelli was sentenced to six years in state prison.

[9] There were, of course, jury instructions to the contrary.

DISCUSSION

It would appear that, in this California judge's mind, psychologists get to present their evidence without regard to either *Frye* or *Daubert*. Far from being worried about an expert unduly swaying a jury with the mantle of expert, the judge (backed by *McDonald* and *Stoll*) preferred to let the adversarial process take its course in the form of rebuttal. Given this mind set, further speculation on the effect of *Daubert*/*Kelly*/*Frye* would seem to be purely academic.

There is good news and bad news in this case example. The judge's decision to allow the flexible battery in, despite the criticisms, means psychologists will rarely have any conventional testimony excluded. It looks as if psychologists can say anything: judges will continue to let it in, allowing psychological information to be considered according to the weight of the evidence, if only to avoid chancing being overturned on appeal. If this case is any guide to the future, judges will not use *Frye or Daubert* to keep juries from hearing testimony, even when the psychologist's procedures are clearly flawed. In fact, a rereading of Reed's description of Reitan's testimony in *Chapple* shows that very thing seems to have happened: the judge did not exclude evidence, she merely considered which of two experts she believed more. The more cynical of us would wonder how an expert with Dr. Reitan's credentials could *not* be the favored one.

REFERENCES

Chapple v. Ganger, 851 F. Supp 1481. (E. D. Wash. 1999).

Daubert v. Merrell Dow Pharmaceuticals Inc. (1993). 509 U.S. 579.

Faust, D., Ziskin, J., & Hiers, J. (1991). *Brain damage claims: Coping with neuropsychological evidence* (Vol. 1). Los Angeles: Law and Psychology.

Fox, D. D. (1994). Normative problems for the Wechsler Memory Scale—Revised Logical Memory test when used in litigation. *Archives of Clinical Neuropsychology*, *9*, 211–214.

Frye v. U.S., 54 Appl. D.C. 46, 293 Fed. 1013 (1923).

Golden, C. J., & Broek, A. (1998). Potential impact of age- and education-corrected scores on HRNB score patterns in participants with focal brain injury. *Archives of Clinical Neuropsychology*, *13*, 683–694.

Gorp, W. G. V., Satz, P., Kiersch, M., & Henry, R. (1986). Normative data on the Boston Naming Test for a group of normal older adults. *Journal of Clinical and Experimental Neuropsychology*, *8*, 702–705.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead–Reitan Battery*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*, 892–900.

Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G., Petersen, R. C., Kokmen, E., & Kurland, L. T. (1992). Mayo's Older Americans Normative Studies: WMS–R norms for ages 56 to 94. *The Clinical Neuropsychologist*, *6*(Supplement), 49–82.

Lezak, M. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Matarazzo, J. D., Daniel, M. H., Prifitera, A., & Herman, D. O. (1988). Intersubtest scatter in the WAIS–R standardization sample. *Journal of Clinical Psychology*, *44*, 940–950.

McKinzey, R. K. (1997). The cross-examination of neuropsychologists: Countering the claim of brain damage. *Prosecutor's Brief*, *19*, 13–20.

McKinzey, R. K., Podd, M. H., Krehbiel, M. A., Mensch, A. J., & Trombka, C. C. (1997). Detection of malingering on the Luria–Nebraska Neuropsychological Battery: An initial and cross-validation. *Archives of Clinical Neuropsychology*, *12*, 505–512.

McKinzey, R. K., Roecker, C. E., Puente, A. E., & Rogers, E. B. (1998). Performance of normal adults on the Luria–Nebraska Neuropsychological Battery, Form I. *Archives of Clinical Neuropsychology*, *13*, 397–413.

Neils, J., Baris, J. M., Carter, C., Dell'aira, A. L., Nordloh, S. J., Weiler, E., & Weisiger, B. (1995). Effects of age, education, and living environment on Boston Naming performance. *Journal of Speech and Hearing Research*, *38*, 1143–1149.

People v. Adams, 53 Cal. App. 3d 109 (1975).

People v. Kelly, 549 P.2d 1240 (1976).

People v. Leahy, (1976). 882 P.2d 321 (1976).

People v. McDonald, 690 P.2d (1984).

People v. Stoll, 783 P.2d 698 (1989).

Reed, J. E. (1996). Fixed vs. flexible neuropsychological test batteries under the *Daubert* standard for the admissibility of scientific evidence. *Behavioral Sciences and the Law*, *14*, 315–322.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, *13*, 255–274.

Russell, E. W. (1998). In defense of the Halstead Reitan Battery: A critique of Lezak's review. *Archives of Clinical Neuropsychology*, *13*, 365–381.

Wechsler, D. (1987). *Wechsler Memory Scale—Revised manual*. New York: Psychological Corporation.

Welch, L., Doineau, D., Johnson, S., & King, D. (1996). Educational and gender normative data for the Boston Naming Test in a group of older adults. *Brain and Language*, *53*, 260–266.

