ELSEVIER

Archives of Clinical Neuropsychology 21 (2006) 809–817

Archives of CLINICAL NEUROPSYCHOLOGY

Potential for interpretation disparities of Halstead–Reitan neuropsychological battery performances in a litigating sample☆,☆☆

Christine L. Yantz[a], Brandon E. Gavett[a],
Julie K. Lynch[b], Robert J. McCaffrey[a,b,*]

[a] University at Albany, State University of New York, NY, United States
[b] Albany Neuropsychological Associates, NY, United States

Accepted 6 September 2006

Abstract

The performances of 110 litigants on seven variables from the Halstead–Reitan neuropsychological battery (HRNB) were used to compare Heaton, Miller, Taylor, and Grant's (2004) Deficit Scale (DS) and Reitan and Wolfson's (1993) Neuropsychological Deficit Scale (NDS). Additional comparisons were made for people who passed or failed the Test of Memory Malingering (TOMM) to determine effects of effort on scores generated by either scoring system. Wilcoxon signed-rank tests revealed that all seven comparisons were significantly different for the full sample ($p \leq 0.001$). The NDS indicated greater levels of impairment compared to DS across all variables. These findings were also obtained when considering effort, though TOMM failure was related to non-significant differences for two variables. These findings suggest that the two scoring systems are not equivalent, with Heaton et al.'s DS resulting in consistently higher identification rates of normal brain functioning compared to those generated from Reitan and Wolfson's NDS system.

Keywords: Normative data; Neuropsychological assessment; Forensic neuropsychology; Demographic corrections; Age and education

The Halstead–Reitan neuropsychological battery for adults (HRNB; Reitan & Wolfson, 1993) is a widely used and researched test battery. In a recent survey, when 879 neuropsychology professionals responded about their typical neuropsychological practices, the HRNB was ranked sixth out of the top 40 most frequently used neuropsychological assessment instruments (Rabin, Barr, & Burton, 2005). The popularity of the HRNB stems from its ability to discriminate between brain-damaged and non-brain-damaged people and to identify the lateralization of damage (Reitan & Wolfson, 1993). One early study compared computerized tomography, electroencephalograms, and HRNB interpretation and found that HRNB interpretation had the highest accuracy rate at identifying brain damage of these three methods (Tsushima & Wedding, 1979). Russell (1995) reviewed 26 validation studies of the HRNB's ability to discriminate brain-damaged from normal subjects, using the Halstead Index ($n = 11$), the Average Impairment Rating (AIR; $n = 7$), the General Neuropsychological Deficit Scale (GNDS; $n = 4$), the Average Impairment Scale ($n = 1$), and brain 1 ($n = 1$);

☆ Portions of this paper were presented at the 33rd annual meeting of the International Neuropsychological Society at St. Louis, MO.
☆☆ Cecil R. Reynolds, Ph.D. served as the guest action editor for this manuscript.
* Corresponding author at: University at Albany, State University of New York, Department of Psychology, Social Sciences, 1400 Washington Avenue, Albany, NY 12222, United States. Tel.: +1 518 442 4841; fax: +1 518 442 5098.
E-mail address: rm188@albany.edu (R.J. McCaffrey).

0887-6177/$ – see front matter
doi:10.1016/j.acn.2006.09.001

in general, all of these HRNB scoring methods correctly classified about 80% of individuals. For the Halstead Index, the average sensitivity was 81.97 for brain-damaged individuals and 66.68 for non-brain injured individuals; its average specificity was 76.40 for brain-damaged individuals and 76.58 for non-brain injured individuals. For the AIR, the average sensitivity was 86.76 for brain-damaged individuals and 60.03 for non-brain injured individuals; its average specificity was 73.60 for brain-damaged individuals and 83.30 for non-brain injured individuals. Finally, for the GNDS, the average sensitivity was 79.03 for brain-damaged individuals and 80.65 for non-brain injured individuals; its average specificity was 89.56 for brain-damaged individuals and 67.56 for non-brain injured individuals. Recent research has demonstrated that dose-related declines in cognitive ability related to severity of traumatic brain injury can be measured by the HRNB (Reitan & Wolfson, 2000; Rohling, Meyers, & Millis, 2003).

Despite the utility of the HRNB in differentiating brain injured and non-brain injured individuals, HRNB measures have been found to be influenced by age, sex, and education, and, to a lesser extent, race or ethnicity (Bernard, 1989; Elias, Podraza, Pierce, & Robbins, 1990; Kupke & Lewis, 1989; Long & Klein, 1990; Reitan & Wolfson, 1995; Vanderploeg, Axelrod, Sherer, Scott, & Adams, 1997). Samples in these studies range from healthy adults (Bernard, 1989; Elias et al., 1990; Reitan & Wolfson, 1995) to epileptic (Kupke & Lewis, 1989) 'pseudoneurologic' (Long & Klein, 1990), and brain-damaged groups (Vanderploeg et al., 1997). It also appears that HRNB performance is related to intelligence; higher IQs (WAIS and WAIS-R) are associated with higher scores on most of the battery's tests (Horton, 1999; Warner, Ernst, Townes, Peel, & Preston, 1987). Several researchers, including Steinmeyer (1986) and Tombaugh (2004), have created demographic corrections for HRNB tests to reduce the influence of these demographic variables on test scores.

In 1991, Heaton, Grant, and Matthews published regression-based norms for the HRNB based on a large sample of normal adults ($n = 378$) in order to control for the effects of age, educational level, and gender. While there is debate about whether or not demographic corrections should be made (Reitan & Wolfson, 2004), a number of researchers have criticized the development and utility of the Heaton, Grant, & Matthews (1991) norms in particular. Researchers have found fault with the statistical construction of the norms, such as inadequate numbers of people per cell in the regression analyses (Fastenau, 1998; Fastenau & Adams, 1996). Another concern focuses on the sample used in developing these demographically corrected norms. Russell (2005) found that the normal volunteers used for the normative sample had a WAIS full scale IQ of about one standard deviation above average ($M = 114$); use of this super-normal sample may result in an artificial inflation of normative data, thereby setting inappropriate cut-off scores for 'impaired' performances. Rather than recruiting volunteers, Russell recommends the use of individuals evaluated and found to be neurologically normal in the generation of normative data, since volunteers appear to be unrepresentative of a typical person who would be administered the HRNB. An additional concern about the normative sample is the sole use of a non-neurologically impaired sample; some researchers have reported findings that demographic variables affect brain-damaged and non-brain-damaged groups equivalently (Reitan & Wolfson, 1995, 2004). Reynolds and James (1997) state that "selecting a representative, balanced, demographically sound, and sufficient group on which to standardize an instrument is of the utmost importance" (p. 354). It is possible that demographic corrections may actually correct for brain injury itself (Golden, Espe-Pfeifer, & Wachsler-Felder, 2000).

Several studies have been completed to determine the utility of demographic corrections on test scores. One study has shown that when the raw scores of a diverse sample of psychiatric patients (diagnoses ranging from head trauma to adjustment disorder) are corrected based on the Heaton et al. (1991) norms, the variance that can be predicted by age and education is substantially reduced (from 10% before correction to <1% with the Heaton et al. norms), suggesting that these norms may be useful for cognitively impaired individuals (Moses, Pritchard, & Adams, 1999). Other investigators have shown that the Heaton et al. (1991) norms misclassify older and/or less educated individuals as neuropsychologically normal more frequently than younger and/or more educated individuals (Fastenau, 1998; Morgan & Caccappolo-van Vliet, 2001). Morgan and Caccappolo-van Vliet (2001) presented two cases of older, less educated adults, with both performances considered normal using the Heaton demographically corrected scoring system but impaired by Reitan and Wolfson's. Morgan and Caccappolo-van Vliet found that supplemental tests, complaints, and imaging corroborated the diagnosis of mild to moderate impairment that was found using Reitan and Wolfson's norms, demonstrating the tendency of Heaton et al.'s (1991) norms to produce false negatives in these populations. This suggests that use of Heaton et al.'s (1991) norms in clinical practice may result in an increased rate of missed diagnoses of brain dysfunction, with the scoring system normalizing scores of older and/or less educated people with or suspected to have brain damage. Sweeney (1999) compared Heaton et al. (1991) and Reitan and Wolfson (1993) HRNB scores of 33 motor-vehicle-accident victims, finding that the Heaton et al. generated scores were in the normal range while the Reitan–Wolfson scores were within the mildly impaired range. This suggests that the normalizing problems with Heaton et al.'s (1991) norms may extend into average age and educational ranges, as Sweeney's sample had a mean age of 41.6 (S.D. = 12.0) and mean educational level of 13.0 (S.D. = 2.2).

One study found that age and education do influence General Neuropsychological Deficit Scale (GNDS) scores regardless of brain-damage status, when comparing demographically corrected and uncorrected scores of 73 brain injured people and those of 41 "pseudoneurologic" (psychologically distressed) individuals (Vanderploeg et al., 1997). Vanderploeg et al. found that use of the Heaton et al. (1991) corrected GNDS scores led to higher classification rates in both groups compared to the Reitan–Wolfson uncorrected GNDS scores (61% specificity and 86% sensitivity versus 54% specificity and 82% sensitivity, respectively). Vanderploeg et al. found that the demographically corrected GNDS had both higher positive predictive power (PPP) and negative predictive power (NPP) than the uncorrected GNDS (PPP = 80% versus 76% and NPP = 71% versus 63%). Sherrill-Pattison, Donders, and Thompson (2000) compared the accuracy of uncorrected or demographically corrected scores for the Trail Making Test A and B and Category Test in correctly classifying degree of brain impairment (severe or mild to moderate) in a sample of 136 patients with a traumatic brain injury (TBI). The authors compared both systems' classification rates of unimpaired and impaired functioning, though they used a cut-off of T-score <38, instead of the Heaton et al. suggested T-score of 40. With these changes, Sherrill-Pattison et al. found that Heaton et al.'s scoring system had a better rate at correctly classifying brain injured people (correctly classifying 79% of the mild–moderate and 73% of the severe brain damage) than Reitan and Wolfson's demographically uncorrected system (67% of the mild–moderate and 70% of the severe).

Reitan and Wolfson (1995), however, found that GNDS scores calculated with the Reitan–Wolfson system, while significantly related to age and education in the non-brain-damaged control group, differentiated between 50 brain injured subjects and 50 age- and education-matched control subjects. Further, these authors found that the effects of age and education on test performance were minimal among the brain-damaged group (age: NS, $r = -0.18$ and education: $p < 0.05$, $r = -0.27$), while the effects of these variables on performances of controls was much greater (age: $p < 0.001$, $r = 0.69$ and education: $p < 0.001$, $r = -0.66$). Recently, Reitan and Wolfson (2004) extended these findings, showing that the Heaton et al. (1991) scoring system correctly classified 87% of 26 normal controls and 63% of the 26 people with diffuse cerebral vascular disease based on their HRNB test scores. They found that the Reitan–Wolfson scoring method correctly classified 76% of the non-brain-damaged and 82% of the brain-damaged individuals based on their HRNB test scores. These findings suggest that the Heaton et al. (1991) norms better classify non-brain-damaged people, but the Reitan and Wolfson scoring system better classifies brain-damaged people, upholding suggestions that the age- and education-corrected norms may be inappropriate for brain-damaged samples.

Even more recently, Reitan and Wolfson (2005) compared the Heaton et al. (1991) scores with Reitan and Wolfson NDS scores for 52 people who had definitive diagnoses of brain disease or damage. For this clinical sample, Reitan and Wolfson found that the Heaton et al. (1991) norms were 1.74 times more likely to result in normal range scores than the NDS scores. The authors also compared the effects of age and education on the two scoring systems' classification of the sample. They found that the Heaton et al. (1991) norms, indeed, correct for age and education effects, classifying 66% of both the younger and older halves of the sample as impaired, and 64% of both the less educated and more educated halves of the sample as impaired. In contrast, the Reitan and Wolfson NDS resulted in 72% of the younger, 87% of the older, 85% of the less educated, 75% of the more educated sample as impaired. Therefore, in each category of the sample the Reitan and Wolfson NDS score was more likely to result in a score indicative of brain damage than the Heaton et al. (1991) (though the difference between percentage impaired in the younger category of the sample was non-significant with $p < 0.10$).

Finally, Golden and van den Broek (1998) compared scores generated by the Heaton et al. (1991) and Reitan–Wolfson scoring systems in their ability to distinguish brain injured patients from non-brain injured people, as well as their ability to correctly classify type of localized injury by brain quadrant. Golden and van den Broek found that each scoring system had strengths and weaknesses, leading to their suggestion that both should be used so that maximal information can be obtained. Therefore, it appears that research to date on the Heaton et al. (1991) and Reitan–Wolfson scoring systems is inconclusive, with several researchers favoring one over the other, and one suggesting the use of both. These studies have compared results from scoring systems across a variety of brain injured, psychiatric, and non-damaged samples.

A study on test use in forensic neuropsychological practice reported that "neuropsychological evaluations were pervasively influenced by the HRNB" (p. 48) though truncated (e.g., using five to eight subtests) or modified forms (e.g., computerized or booklet category tests) of the test battery were frequent (Lees-Haley, Smith, Williams, & Dunn, 1996). Thus, it appears that litigants are frequently administered some components of the HRNB during forensic neuropsychological evaluations. It is likely that many of the professionals that administer these HRNB components use either Heaton et al. or Reitan and Wolfson's normative data to interpret the results. If the choice between these scoring systems is made without consideration of the potential different interpretations that may result, the neuropsychologist could wrongly conclude that the client is either more or less impaired than he/she actually is.

Recently Heaton, Miller, Taylor, and Grant (2004a) have published a revision of their 1991 norms. One addition to the revised norms is the inclusion of separate data for African–Americans and Caucasians. Another improvement is the inclusion of deficit scores for tests and a Global Deficit score, similar to the neuropsychological deficit scores and GNDS of the Reitan–Wolfson scoring system. The revised norms also are based on a larger sample ($n = 1.212$ for the HRNB) and have more sophisticated analyses (see Heaton et al., 2004a). To date, no researchers have examined these revised demographically adjusted norms in comparison to the Reitan and Wolfson's scoring system. This study compared the scores of litigating adults obtained with each scoring system to determine the degree to which the two scoring systems agree or disagree. To test the potential effect of inadequate effort on the relationship between scores generated by either scoring system, separate comparisons were also made for people who passed the Test of Memory Malingering (TOMM; Tombaugh, 1996) and people who failed the TOMM.

1. Method

1.1. Participants

Approval for analysis of archival data was obtained from the human subjects institutional review board. The 110 participants who were seen in the context of litigation involving allegations of traumatic brain injury were all Caucasian and ranged in age from 20 to 78 ($M = 43.13$, S.D. = 12.47), in years of education from 7 to 22 ($M = 13.05$, S.D. = 2.68), and WAIS-R FSIQ from 61 to 129 ($M = 93.31$, S.D. = 12.25). All subjects were administered the HRNB in its entirety and the TOMM as well as other tests as part of each evaluation. Fifty-three percent of the sample was male, and 88% were right-handed. Sixty-six percent of the sample passed the TOMM. The performance of these litigating adults on seven HRNB variables was examined: Category Test, Trails A and B, Seashore Rhythm Test, Speech-sounds Perception Test, and Memory and Location scores for the Tactual Performance Test (TPT). The TPT time to completion score was not included in these analyses due to different data being used by Reitan and Wolfson (total time) and Heaton et al. (time for each block).

1.2. Procedure

Raw scores from these seven variables were converted into the Deficit Scale (DS) of Heaton et al. (2004a) (scores range from 0 = normal to 5 = severe impairment) and the Neuropsychological Deficit Scale (NDS) of Reitan and Wolfson (1993) (scores range from 0 = normal to 3 = severe impairment). The NDS conversion was done by hand, using Reitan and Wolfson's values; the DS values were obtained by using the scoring program companion to Heaton et al.'s revised norms (Heaton, Miller, Taylor, & Grant, 2004b). In order to directly compare DS and NDS scores, three composite severity scores were created: none to mild (0 and 1 for Reitan and Heaton), mild to moderate (2 for Reitan and 2 and 3 for Heaton), and moderate to severe (3 for Reitan and 4 and 5 for Heaton). Group means for the transformed DS and NDS scores can be found in Table 1. Scoring procedures were the same for both sets of norms. The Reitan and Wolfson GNDS and Heaton et al. GDS were not compared since the GDS includes tests that were not administered to this sample.

1.3. Analyses

The non-parametric Wilcoxon signed-ranks test was used in this study because of its utility in comparing paired, rather than independent, scores (Gravetter & Wallnau, 2000). The difference between the paired scores for each individual is computed, and then positive and negative difference scores and their rankings are compared to determine if there is a significant difference between the paired scores. Transformed DS and NDS scores for the seven variables were compared using Wilcoxon signed-ranks test for the entire sample. Additionally, score pairs were compared separately for people passing and failing the TOMM.

Table 1
Descriptive statistics of the transformed DS and NDS scores for HRNB variables by TOMM status

Variable	Scale	Total sample		TOMM passed		TOMM failed	
		n	*M* (S.D.)	*n*	*M* (S.D.)	*n*	*M* (S.D.)
Category	DS	106	1.30 (0.57)	72	1.26 (0.56)	34	1.38 (0.60)
	NDS	106	2.02 (0.86)	72	2.03 (0.86)	34	2.00 (0.89)
Trails A	DS	108	1.65 (0.81)	73	1.48 (0.71)	35	2.00 (0.91)
	NDS	108	1.80 (0.89)	73	1.64 (0.84)	35	2.11 (0.93)
Trails B	DS	106	1.47 (0.73)	72	1.39 (0.66)	34	1.65 (0.85)
	NDS	107	1.79 (0.88)	73	1.66 (0.82)	34	2.06 (0.95)
Seashore Rhythm	DS	105	1.48 (0.65)	71	1.38 (0.62)	34	1.68 (0.68)
	NDS	105	1.70 (0.80)	71	1.55 (0.77)	34	2.03 (0.76)
Speech-sounds	DS	106	1.56 (0.77)	73	1.41 (0.68)	33	1.88 (0.86)
	NDS	106	1.72 (0.86)	73	1.56 (0.80)	33	2.06 (0.90)
TPT Memory	DS	100	1.28 (0.49)	70	1.26 (0.47)	30	1.33 (0.55)
	NDS	100	1.45 (0.59)	70	1.37 (0.54)	30	1.63 (0.67)
TPT Location	DS	100	1.13 (0.37)	70	1.16 (0.40)	30	1.07 (0.25)
	NDS	100	2.39 (0.68)	70	2.33 (0.7)	30	2.53 (0.57)

Note: HRNB, Halstead–Reitan neuropsychological battery; DS, Transformed Heaton et al. Deficit Scale; NDS, Transformed Reitan–Wolfson Neuropsychological Deficit Scale; TPT, Tactual Performance Test. Transformed DS and NDS scores range in value from 1 (none to mild) to 3 (moderate to severe).

2. Results

Correlations between the untransformed DS and NDS for the seven variables ranged from 0.46 (TPT Location) to 0.87 (Speech-sounds), while the correlations for the transformed DS and NDS variables ranged from 0.32 (TPT Location) to 0.86 (Speech-sounds). Wilcoxon signed-ranks tests revealed that all seven comparisons of transformed NDS and DS scores were significantly different for the full sample (Table 2). NDS scores indicated greater levels of impaired functioning compared to DS scores across all of the variables.

When considering effortful performance, the pattern of results was similar to that of the full sample (Table 2). Transformed NDS and DS scores for people who passed the TOMM were all significantly different. Like the full sample, NDS scores were greater than DS scores for all variables. Participants who failed the TOMM had significantly different NDS and DS scores for five of the seven variables; NDS and DS values for Trails A and Seashore Rhythm Test were not significantly different. The pattern of the significant differences was similar to the rest of the sample, with all tests showing the NDS higher than the DS. Additionally, independent samples *t*-tests showed that education level is significantly higher for those who passed the TOMM compared to those who failed (13.6 and 12.0 years, respectively, $t(84) = 3.31$, $p \leq 0.001$). There were no age or gender differences between the groups passing and failing the TOMM.

3. Discussion

The current analyses revealed that, for all seven HRNB subtests examined, Heaton et al.'s (2004a) DS scores revealed a significantly lesser degree of neuropsychological impairment relative to Reitan and Wolfson's (1993) NDS scores. This suggests that professionals utilizing Heaton et al.'s (2004a) norms may have a strong tendency to interpret test results of individual clients as less impaired compared to those using Reitan and Wolfson's norms. The lowered significance of results when considering people with questionable effort (i.e., TOMM failures) may be impacted by decreased statistical power, due to the small sample size of this group. Alternatively, the non-significant differences found for the TOMM-failing group in contrast to the TOMM-passing group may be due to the tendency for the TOMM-failing group to exaggerate brain dysfunction (Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005); if so, a ceiling effect for cognitive impairment may be seen in this sub-sample. The general results remain important, however, with one scoring system consistently resulting in lower impairment ratings than another. Illustration of these results

Table 2
Wilcoxon signed-ranks test statistics for Transformed NDS and DS comparisons, arranged by TOMM status

Variable	TOMM status	*n*	Negative ranks (*n*)	Positive ranks (*n*)	Ties (*n*)	*Z*-score
Category	Total	106	0	60	46	−7.10[a]
	Pass	72	0	41	31	−5.81[a]
	Fail	34	0	19	15	−4.19[a]
Trails A	Total	108	5	19	84	−2.92[a]
	Pass	73	3	14	56	−2.68[b]
	Fail	35	2	5	28	−1.27
Trails B	Total	106	1	30	75	−4.96[a]
	Pass	72	1	18	53	−3.75[a]
	Fail	34	0	12	22	−3.28[a]
Seashore Rhythm	Total	105	0	24	81	−4.90[a]
	Pass	71	0	12	59	−3.46[a]
	Fail	34	0	12	22	−3.46[a]
Speech-sounds	Total	106	2	18	86	−3.55[a]
	Pass	73	0	10	63	−3.05[b]
	Fail	33	2	8	23	−1.90
TPT Memory	Total	100	1	17	82	−3.71[a]
	Pass	70	1	8	61	−2.31[b]
	Fail	30	0	9	21	−3.00[b]
TPT Location	Total	100	0	88	12	−8.43[a]
	Pass	70	0	59	11	−6.93[a]
	Fail	30	0	29	1	−4.85[a]

Note: DS, Transformed Heaton et al. Deficit Scale; NDS, Transformed Reitan–Wolfson Neuropsychological Deficit Scale; TOMM, Test of Memory Malingering; TPT, Tactual Performance Test.
[a] NDS > DS, $p \leq 0.001$ (2-tailed).
[b] NDS > DS, $p \leq 0.01$ (2-tailed).

can be found when comparing the DS and NDS scores for an individual. Table 3 shows the raw score, *T* score, and transformed DS and NDS values for the seven variables in two cases.

3.1. Case 1

Case 1 is a 56-year-old man with 8 years of education. This patient scored in the mild to severe range of impairment on six of seven transformed NDS scores. Based on these results, one would expect Case 1 to have a diagnosis of brain damage. However, if transformed DS scores were used, Case 1 would fall within the normal to mild range on five

Table 3
Case comparisons of transformed DS and NDS ratings

	Case 1: Male, aged 56 with 8 years of education				Case 2: Female, aged 65 with 9 years of education			
	Raw score[a]	*T*-score[a]	DS[a]	NDS[a]	Raw score[a]	*T*-score[a]	DS[a]	NDS[a]
Category Test	60	48	1	2	108	40	1	3
Trails A	82	25	2	3	73	30	2	3
Trails B	194	39	1	3	100	52	1	2
Seashore Rhythm Test	25	48	1	1	26	52	1	1
Speech-sounds test	21	30	2	3	6	52	1	1
TPT Memory	5	39	1	2	8	62	1	1
TPT Localization	1	44	1	3	4	58	1	2

Note: HRNB, Halstead–Reitan neuropsychological battery; DS, Transformed Heaton et al. Deficit Scale; NDS, Transformed Reitan–Wolfson Neuropsychological Deficit Scale; TPT, Tactual Performance Test. DS and NDS values: 1 = none to mild impairment; 2 = mild to moderate impairment; 3 = moderate to severe impairment.
[a] HRNB variable.

measures and in the mild to moderate range on two measures. These transformed scores lead to a much more favorable view of the man's neuropsychological functioning. A clinician using one or the other of these methods to interpret the test results of Case 1 may consider the client to have fairly severe damage or only mild impairments, depending on the scoring system used.

3.2. Case 2

Case 2 is a 65-year-old woman with 9 years of education. She would likely be thought to have brain damage based on her four mild to severe NDS scores. Conversely, little impairment would be expected if the practitioner used her DS scores (with only one score in the mild to moderate impairment range). This large discrepancy between the scoring systems for this woman's test results corresponds with the findings of previous researchers that the Heaton et al. method tends to underestimate impairments in people of advanced age and lower educational attainment (Fastenau, 1998; Morgan & Caccappolo-van Vliet, 2001) relative to the Reitan–Wolfson norms. These cases exemplify the problematic pattern of significant differences between NDS and DS scores across HRNB tests; depending on the scoring system used by a neuropsychologist, vastly different conclusions can be drawn based on one set of test results.

The results of this study show consistent differences between the scores generated by these different methods, which on its own, suggests that the common use of both scoring systems in the field could be an important problem. The trend that DS scores tend to show less impairment than the NDS scores means that preferential use of one or the other scoring systems can lead to differing rates of diagnoses between clinicians, causing difficulties communicating test data with professionals within and outside the field (Lezak, 2002). Additionally, the knowledge of the pattern of results that are obtained by either method could lead to these normative data being misused in forensic settings (van Gorp & McMullen, 1997). If a neuropsychologist has a predetermined bias to find brain damage or not, he/she may choose to use one of these scoring systems based on its tendency to classify people as normal or impaired more than the other method.

This study has limitations, including the use of a sample of litigating adults. Though the HRNB was not designed for litigating adults, neuropsychologists frequently include HRNB tests in forensic evaluations (Lees-Haley et al., 1996). This study provides information to those clinicians who may have not considered the potential gross differences between scoring systems. Another limitation in this study is that the data of diagnostic category (i.e., brain-damaged or non-brain-damaged) was not available, disallowing the results to be used to support either scoring system over the other. The utility of the current study is to raise awareness of the significant differences of results obtained by two major scoring systems for the HRNB, which should make clinicians take pause before arbitrarily choosing one set of norms over the other. Despite these limitations, the results of the current study are consistent with findings of other research that has been completed with clinical samples and has found that Heaton et al. (1991) norms underestimate brain damage as compared to Reitan and Wolfson norms (Fastenau, 1998; Morgan & Caccappolo-van Vliet, 2001; Reitan & Wolfson, 2005; Sweeney, 1999).

Since there has been evidence favoring either method at this time, it is unclear how these issues should be resolved. It is our recommendation that, until research demonstrates clear superiority of one scoring system over the other, clinicians should use the Reitan and Wolfson norms so that the results from separate professionals can be accurately compared. There are several reasons for this: first, there has been a strong history of research demonstrating the utility of the norms created by Reitan and Wolfson; second, most of the research on the Heaton et al. (1991) norms have not been compared across the same patient populations used by Reitan and Wolfson in their classification studies; third, investigators have not yet compared Heaton et al.'s (2004a) new norms to the Reitan and Wolfson norms with multiple diagnostic groups of patients across multiple settings. Additionally, it has been suggested that the Reitan and Wolfson norms may be more appropriate for brain-damaged samples (Reitan & Wolfson, 2004), which tends to be the population of concern for neuropsychologists (rather than neurologically normal people). The present study demonstrates consistent, statistically significant differences between scores as a function of the norms, with those generated by the Heaton et al. (2004a) norms indicating a lesser degree of brain impairment compared to those from the Reitan and Wolfson norms.

References

Bernard, L. C. (1989). Halstead–Reitan neuropsychological test performance of black, Hispanic, and white young adult males from poor academic backgrounds. *Archives of Clinical Neuropsychology*, *4*, 267–274.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M., & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology*, *20*, 191–198.

Elias, M. F., Podraza, A. M., Pierce, T. W., & Robbins, M. A. (1990). Determining neuropsychological cut scores for older, healthy adults. *Experimental Aging Research*, *16*, 209–220.

Fastenau, P. S. (1998). Validity of regression-based norms: An empirical test of the comprehensive norms with older adults. *Journal of Clinical and Experimental Neuropsychology*, *20*, 906–916.

Fastenau, P. S., & Adams, K. M. (1996). Heaton, Grant and Matthews' Comprehensive Norms: An overzealous attempt. *Journal of Clinical and Experimental Neuropsychology*, *18*, 444–448.

Golden, C. J., Espe-Pfeifer, P., & Wachsler-Felder, J. (2000). *Neuropsychological interpretations of objective psychological tests*. New York: Kluwer Academic/Plenum Publishers.

Golden, C. J., & van den 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.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead–Reitan battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004a). *Revised comprehensive norms for an expanded Halstead–Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources, Inc.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004b). *Revised comprehensive norms for an expanded Halstead–Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults scoring program*. Lutz, FL: Psychological Assessment Resources, Inc.

Horton, A. M., Jr. (1999). Above-average intelligence and neuropsychological test score performance. *International Journal of Neuroscience*, *99*, 221–231.

Kupke, T., & Lewis, R. (1989). Relative influence of subject variables and neurological parameters on neuropsychological performance of adult seizure patients. *Archives of Clinical Neuropsychology*, *4*, 351–363.

Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology*, *11*, 45–51.

Lezak, M. D. (2002). Responsive assessment and the freedom to think for ourselves. *Rehabilitation Psychology*, *47*, 339–353.

Long, C. J., & Klein, K. (1990). Decision strategies in neuropsychology II: Determination of age effects on neuropsychological performance. *Archives of Clinical Neuropsychology*, *5*, 335–345.

Morgan, J. E., & Caccappolo-van Vliet, E. (2001). Advanced years and low education: The case *against* the comprehensive norms. *Journal of Forensic Neuropsychology*, *2*, 53–69.

Moses, J. A., Pritchard, D. A., & Adams, R. L. (1999). Normative corrections for the Halstead Reitan neuropsychological battery. *Archives of Clinical Neuropsychology*, *14*, 445–454.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*, 33–65.

Reitan, R. M., & Wolfson, D. (1993). *The Halstead–Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). South Tucson, Arizona: Neuropsychology Press.

Reitan, R. M., & Wolfson, D. (1995). Influence of age and education on neuropsychological test results. *The Clinical Neuropsychologist*, *9*, 151–158.

Reitan, R. M., & Wolfson, D. (2000). The neuropsychological similarities of mild and more severe head injury. *Archives of Clinical Neuropsychology*, *15*, 433–442.

Reitan, R. M., & Wolfson, D. (2004). Clinical and forensic issues regarding age, education, and the validity of neuropsychological test results: A review and presentation of a new study. *Journal of Forensic Neuropsychology*, *4*, 1–32.

Reitan, R. M., & Wolfson, D. (2005). The effect of age and education transformations and neuropsychological test scores of persons with diffuse or bilateral brain damage. *Applied Neuropsychology*, *12*, 181–189.

Reynolds, C. R., & James, E. M. (1997). Development of neuropsychological measures. In M. E. Maruish, & J. A. Moses (Eds.), *Clinical neuropsychology: Theoretical foundations for practitioners* (pp. 347–370). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rohling, M. L., Meyers, J. E., & Millis, S. R. (2003). Neuropsychological impairment following traumatic brain injury: A dose–response analysis. *The Clinical Neuropsychologist*, *17*, 289–302.

Russell, E. W. (1995). The accuracy of automated and clinical detection of brain damage and lateralization in neuropsychology. *Neuropsychology Review*, *5*, 1–68.

Russell, E. W. (2005). Norming subjects for the Halstead Reitan battery. *Archives of Clinical Neuropsychology*, *20*, 479–484.

Sherrill-Pattison, S., Donders, J., & Thompson, E. (2000). Influence of demographic variables on neuropsychological test performance after traumatic brain injury. *The Clinical Neuropsychologist*, *14*, 496–503.

Steinmeyer, C. H. (1986). A meta-analysis of Halstead–Reitan test performances of non-brain damaged subjects. *Archives of Clinical Neuropsychology*, *1*, 301–307.

Sweeney, J. E. (1999). Raw, demographically altered, and composite Halstead–Reitan Battery data in the evaluation of adult victims of nonimpact acceleration forces in motor vehicle accidents. *Applied Neuropsychology*, *6*, 79–87.

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. New York: Multi Health Systems.

Tombaugh, T. N. (2004). Trail making test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*, 203–214.

Tsushima, W. T., & Wedding, D. (1979). A comparison of the Halstead–Reitan neuropsychological battery and computerized tomography in the identification of brain disorder. *Journal of Nervous and Mental Disease, 167*, 704–707.

van Gorp, W. G., & McMullen, W. J. (1997). Potential sources of bias in forensic neuropsychological evaluations. *The Clinical Neuropsychologist, 11*, 180–187.

Vanderploeg, R. D., Axelrod, B. N., Sherer, M., Scott, J. G., & Adams, R. L. (1997). The importance of demographic adjustments on neuropsychological test performance: A response to Reitan and Wolfson (1995). *The Clinical Neuropsychologist, 11*, 210–217.

Warner, M. H., Ernst, J., Townes, B. D., Peel, J., & Preston, M. (1987). Relationships between IQ and neuropsychological measures in neuropsychiatric populations: Within-laboratory and cross-cultural replications using WAIS and WAIS-R. *Journal of Clinical and Experimental Neuropsychology, 9*, 545–562.