Sensitivity to Brain Dysfunction of the Halstead-Reitan vs an Ability-Focused Neuropsychological Battery

Glenn J. Larrabee a , Scott R. Millis b & John E. Meyers c

a Independent Practice, Sarasota, FL
b Wayne State University School of Medicine, Detroit, MI
C Center for Neurosciences, Orthopedics, and Spine, Dakota Dunes, SD, USA

Published online: 11 Sep 2008.

To cite this article: Glenn J. Larrabee , Scott R. Millis & John E. Meyers (2008) Sensitivity to Brain Dysfunction of the Halstead-Reitan vs an Ability-Focused Neuropsychological Battery, The Clinical Neuropsychologist, 22:5, 813-825, DOI: 10.1080/13854040701625846

To link to this article: http://dx.doi.org/10.1080/13854040701625846

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
SENSITIVITY TO BRAIN DYSFUNCTION OF THE HALSTEAD-REITAN VS AN ABILITY-FOCUSED NEUROPSYCHOLOGICAL BATTERY

Glenn J. Larrabee1, Scott R. Millis2, and John E. Meyers3

1Independent Practice, Sarasota, FL, 2Wayne State University School of Medicine, Detroit, MI, and 3Center for Neurosciences, Orthopedics, and Spine, Dakota Dunes, SD, USA

We compared the sensitivity to brain dysfunction of an ability focused neuropsychological battery (AFB), as a proxy for the core of a flexible battery, to the Halstead-Reitan Battery (HRB). The AFB was designed to represent constructs of language function, fine motor skill, working memory, processing speed, verbal and visual memory, and verbal and visual abstraction and problem solving. Receiver operating characteristic analysis (ROC) yielded an area under curve (AUC) of .86 for the AFB, versus .83 for the HRB (p = .50), for discriminating 54 patients with brain dysfunction due to various etiologies, from 69 non-neurologic medical controls. Additionally, Bayesian Model Averaging selected four tests from the combined set of AFB and HRB subtests, plus Trail Making B, which optimally discriminated the brain dysfunction from medical control patients: H-Words, Grooved Pegboard, Finger Tapping, and Trail Making B. These data support the current mainstream practice in neuropsychology of using an AFB (flexible battery) to assess brain dysfunction. In particular, tests involving processing speed appear to be among the most sensitive measures of brain dysfunction. The data do not support the superiority of the HRB to AFB approaches.

Keywords: Ability-Focused; Brain Dysfunction; Flexible Battery; Halstead-Reitan.

INTRODUCTION

Sweet, Nelson, and Moberg (2006), in a survey of Division 40 members and other neuropsychologists, found that 76% of neuropsychologists used a flexible core battery, 18% used a totally flexible assessment, and only 7% used a standardized battery such as the Halstead-Reitan (HRB). Typically, persons employing a flexible core battery use individually normed measures of key neuropsychological constructs such as language, sensorimotor function, attention, memory, and problem-solving skills. This core battery is administered to all subjects, with additional flexible exploration of deficits that are identified on key neuropsychological constructs (Bauer, 2000). By contrast, proponents of standardized batteries such as the HRB, which contains a fixed series of procedures administered to all patients, argue that standardized batteries are the only valid procedures for use in forensic situations, because these procedures have been validated as a group, and have been co-normed...
Related to the issue of test validity is the claim by Halstead (1947) and later Reitan and Wolfson (1993), that the HRB measures “biologic” as opposed to “psychometric” intelligence and consequently is more sensitive to the presence of brain dysfunction than the Wechsler Adult Intelligence scales (WIS; [Wechsler-Bellevue], Wechsler, 1939; WAIS, Wechsler, 1955; WAIS-R, Wechsler, 1981), which are a common component of both AFB and HRB assessment.

Recently, Loring and Larabee (2006) reanalyzed Reitan’s original validation data for the HRB and Wechsler-Bellevue, and found that the Wechsler-Bellevue was nearly as sensitive to brain dysfunction as the HRB. Moreover, Loring and Larabee reviewed subsequent investigations employing revised WIS scales (WAIS, Wechsler, 1955; WAIS-R, Wechsler, 1981), and found these scales to be as sensitive, if not more so, than the HRB in discriminating patients with brain disorders from control subjects. Factor analysis of the HRB, WAIS-R, and Wechsler Memory Scale-Revised (Wechsler, 1987; see Larabee, 2000; Leonberger, Nicks, Larabee, & Goldfader, 1992) showed complete overlap of the HRB subtests loading on the same factors as the WAIS-R subtests—visual problem solving, attention, and psychomotor speed—leading Loring and Larabee (2006) to conclude that the WIS/HRB comparisons and factor analyses did not support the presence of separate “biologic” and “psychometric” intelligences. Last, both the HRB and WIS are less sensitive to brain dysfunction than measures of verbal supraspan learning and memory such as the Verbal Selective Reminding Test (VSRT; Buschke, & Fuld, 1974; Dikmen, Machamer, Winn, & Temkin, 1995) or the Rey Auditory Verbal Learning Test (AVLT; Lezak, Howieson, & Loring, 2004; Powell, Cripe, & Dodrill, 1991; Rey, 1964).

Rohling, Meyers, and Millis (2003) presented data demonstrating equivalent sensitivity to head trauma severity of the Meyers Neuropsychological Battery (MNB; an ability-focused core for a flexible battery) to an HRB augmented by additional measures of learning and memory (Dikmen et al., 1995). Moreover, mean T scores (collapsed over all functional domains) were essentially equivalent for the MNB and the augmented HRB for five levels of head trauma severity, ranging from less than 1 hour to follow commands to 14 to 28 days to follow commands. The correlation between the T scores for the five levels of impairment for the MNB and HRB samples was .97. Last, the mean overall T score, collapsed across all five severity groups, was 39.2 (SD 4.1) for the MNB and 38.9 (SD 5.1) for the HRB. This study demonstrated not only equivalent sensitivity of non-HRB tests to the HRB, but also the equivalence of independently normed tests, aggregated across the measures comprising the MNB, to co-normed tests comprising the expanded HRB employed by Dikmen et al. (1995).

In the present investigation we planned to directly compare the sensitivity to brain dysfunction of the original core HRB subtests: Category Test (CAT), Tactual Performance Test (TPT), Finger Tapping (FT), Seashore Rhythm Test (SRT), and Speech Sounds Perception Test (SSPT), to the sensitivity to brain dysfunction of an ability-focused neuropsychological battery designed to represent a proxy for currently employed core flexible batteries. Variable selection for the ability-focused battery (AFB) was constrained by the tests available from the Halstead-Russell Neuropsychological Evaluation System (HRNES; Russell & Starkey, 1993), with raw test data provided by Dr. Elbert Russell. The following measures comprised the
AFB: (1) language: H-Words Subtest (H-Words; 2 minutes to write words beginning with the letter H); (2) sensorimotor: Grooved Pegboard Test (GPB); (3) working memory: WAIS-R Arithmetic subtest (Arithmetic); (4) processing speed: WAIS-R Digit Symbol (Digit Symbol); (5) verbal memory: delayed recall of Wechsler Memory Scale (WMS, Wechsler, 1945) Logical Memory (LM); (6) visual memory: delayed recall of WMS Visual Reproduction (VR); (7) verbal intelligence and problem solving: WAIS-R Similarities subtest (Similarities); and (8) visual intelligence and problem solving: WAIS-R Block Design (Block Design).

For additional analyses beyond the HRB and AFB comparisons, we included Trail Making Test B (Trails B). Trails B, not a part of the original HRB, is frequently employed by clinicians using either the HRB or AFB approach to assessment. Indeed, Trails B was the third most widely used test, following the WAIS-R/III and WMS-R/III in a comprehensive survey of neuropsychological assessment practices (Rabin, Barr, & Burton, 2005).

METHOD

Participants

Participant data were provided courtesy of Elbert W. Russell, Ph.D., tested at the Miami VA Medical Center from 1971 to 1989. Since the subjects were examined in a setting wherein external incentive (Veterans Disability) was a possible factor, and testing was conducted prior to the widespread availability of measures of response bias (symptom validity testing or SVT), we screened subjects, excluding those who either (a) obtained an age-scaled score of 5 or less on Digit Span (Babikian, Boone, Lu, & Arnold, 2006) or (b) obtained a difference of 5 or greater between the age-scaled scores of Vocabulary and Digit Span (Mittenberg et al., 2001). The initial sample of 204 brain dysfunction patients was reduced to 101 by the SVT screening, with the initial sample of 113 non-neurologic medical controls reduced to 95 following SVT screening. We further eliminated individuals with missing data to yield 69 non-neurologic medical controls and 54 brain-impaired participants. In the brain-impaired group (assumed to have brain dysfunction on the basis of the following diagnoses and neurodiagnostic procedures available at the time these data were collected), 15 had traumatic brain injury, 20 had cerebrovascular disease (including CVA and TIA), 5 had multiple sclerosis, 4 had seizure disorder, 2 had tumor, 2 had AVM or aneurysm, 2 had infectious disease, 2 had congenital brain damage, 1 had degenerative CNS disease, and 1 had damage due to toxic exposure. There were 64 males and 5 females in the control group, and 49 males and 5 females in the brain-impaired group ($\chi^2 = 1.64, p < .747$). Mean age was 48.03 ($SD = 14.36$) for controls, and 47.33 (15.90) for brain impaired ($F = 1.02, p < .32$). Mean education was 12.13 ($SD = 2.71$) for controls, and 12.74 ($SD = 2.49$) for the brain impaired ($F = 0.049, p < .824$).

Procedures

All tests were administered following standard clinical practice. Test scores available for analysis were constrained by how these archival data were
originally recorded. Raw scores were available for the HRB subtests, Trails B, H-Words, LM, and VR. For the WAIS-R variables, only the scaled scores determined from the reference group sample (age 20–34; Wechsler, 1981), and the age-corrected scaled scores were available. The only common normative base for all of the data was that available through the HRNES (Russell & Starkey, 1993). Use of these normative data would have been problematic, however, since the norms for all of the HRNES tests are derived from multiple regression capturing the variance each test shares with the core HRNES measures: CAT, Trails B, FT, Digit Symbol, TPT (Time and Memory), SSPT, Block Design, Aphasia Screening Test, and Perceptual Disorders Examination. As can be seen, these reference tests are heavily weighted with measures of visuospatial processing and psychomotor speed. Consequently, rescaling the AFB tests using the HRNES normative data would have only weighted variance shared with visuospatial processing and psychomotor speed, and would not have captured AFB variance shared with memory or verbal skills. As a result, we chose to analyze raw scores for all non-WAIS-R HRB and AFB tests, and non-age-corrected WAIS-R scaled scores, since the brain dysfunction and control participants were naturally matched on age, education, and gender. Age-corrected WAIS-R scaled scores were utilized for the symptom validity screening, since the original cut-off scores were derived on age-corrected scaled scores.

HRB subtests were scored to yield the traditional measures of total errors for the CAT, a weighted Total Time variable for the TPT (see Russell & Starkey, 1993), Memory and Location for the TPT, dominant FT score, and total errors for the SRT and SSPT. For the AFB, raw scores were analyzed for H-Words, total time for the GPB, non-age scaled scores for the WAIS-R variables (Arithmetic, Digit Symbol, Similarities, Block Design), and raw scores for LM and VR. We also included data from Trail Making B (Trails B), for additional analyses beyond those planned for the AFB and HRB comparison. Trails B was scored for time to complete this task.

RESULTS

Two separate logistic regression models were fitted: one model containing the HRB measures and the other containing the AFB measures. Logistic regression is used when the dependent variable (response variable) is binary. In this case, the models were used to discriminate brain-dysfunction from control participants. Although sensitivity and specificity are commonly used to assess the diagnostic efficiency of test or group of tests in a logistic regression model, both sensitivity and specificity rely on a single cutoff score. A more complete description of classification accuracy is given by the area under the receiver operating characteristic (ROC) curve (AUC) (Zhou, Obuchowski, & McClish, 2002). The curve plots the probability of detecting a disorder (sensitivity) and false signal (1 – specificity or false positive) for an entire range of possible cutoff scores (Hsiao, Bartko, & Potter, 1989). The AUC provides a measure of the model to discriminate between persons with brain dysfunction versus the control participants. Perfect discrimination is achieved at an AUC of 1.00, with chance falling at an AUC of .50, represented as the diagonal line traversing from zero false positive rate, and zero sensitivity, to perfect sensitivity and 100% false positives. AUC of 0.7 to < 0.8 have
been characterized as acceptable, 0.8 to < 0.9 as excellent, and = 0.9 as outstanding (Hosmer & Lemeshow, 2000).

Because the logistic regression model contains a set of tests and not a single test score, the logistic linear predictors from each model are used to construct and calculate the AUC. There are different methods to calculate and compare AUCs. Parametric methods, such as the Metz and Kronman (1980) method, are based on the bivariate normal distribution, which assumes a normal distribution for cases with the disease and a normal distribution for cases without, or that the data have been monotonically transformed to normal. Parametric methods also assume homoscedasticity. The assumptions can be restrictive and, thus, we elected to use a nonparametric approach (DeLong, DeLong, & Clarke-Pearson, 1988).

Figure 1 shows the ROC for the AFB, which had an AUC of .86 (95% CI .79–0.93), and the ROC for the HRB, which had an AUC of .83 (95% CI .76–0.91). The difference in AUCs for AFB and HRB was not statistically significant, \( \chi^2 = 0.45, p = .50 \). Using this criterion, neither battery of tests is superior from a diagnostic discrimination perspective, and both batteries showed excellent discrimination.

To supplement the AUC analysis, we also examined the Bayesian information criterion (BIC) (Raftery, 1996) associated with the HRB and AFB models. The BIC is another measure that assesses the overall fit of a model and allows the comparison of both nested and non-nested models. The BIC is intended to provide a measure of the weight of evidence favoring one model over another, or
Bayes factor. Suppose that one has two models. Under the assumption that one has no prior preference for one model over the other, BIC identifies the model that is more likely to have generated the observed data. BIC is defined as (Hardin & Hilbe, 2007):

$$\text{BIC} = D(M_K) - (df) \ln(n)$$

When comparing non-nested models, as we have in this study, one may assess the degree of model preference on the basis of the absolute values of the difference between the BIC values of the two models. The BIC for the HRB was 163.44 compared to 159.87 for AFB. The scale for determining the relative preference for one model over another (Raftery, 1996) with regard to the magnitude of the absolute difference in BIC is as follows: 0–2 = weak; 2–8 = positive; 6–10 = strong; >10 = very strong. Hence, the discrepancy in BICs of 3.57 falls in the “positive” range of support for the AFB over the HRB.

A third method was used to examine the relative contributions of the HRB and AFB in differentiating the brain dysfunction group from the control group. A full model was fitted with all tests from the HRB, AFB, and Trails B entered into the model, yielding an AUC of .92 (95% CI .87–.97). Next, the HRB tests were removed from the model and a reduced model was fitted. A likelihood ratio test was conducted to compare the full model with the reduced model and found to be significant, $$\chi^2 = 20.61, p = .004$$. This indicated that the HRB made a significant contribution to the model. Similarly, a reduced model was fitted with the AFB tests removed. This likelihood ratio test was also statistically significant, $$\chi^2 = 34.47, p = .0001$$, which indicated that the AFB tests also made significant contributions to the model.

Hence, our next step was to determine the “best” subset of tests from both batteries. We employed Bayesian Model Averaging analysis to select the most diagnostically sensitive subtests from the AFB and HRB, plus Trails B. Bayesian model averaging (BMA) provides an alternative approach to variable and model selection (Wasserman, 2000). Unlike stepwise methods and conventional frequentist approaches, BMA is able to account for model uncertainty that overcomes the difficulties associated with standard model selection procedures. BMA approaches the problem of model selection by finding a collection of the best models, and averaging over them in accordance with their posterior model probabilities. The different models and variables are incorporated into the predictions with weights proportional to the evidence we have for their utility. BMA has been shown to provide superior out-of-sample predictive performance compared to stepwise methods (Viallefont, Raftery, & Richardson, 2001; Wang, Zhang, & Bakhai, 2004). In most cases BMA selects the correct model and out-performs stepwise approaches at predicting an event of interest. The Appendix presents a brief discussion of the mathematical foundations of BMA.

BMA for logistic regression was implemented with the statistical software, R, using the bic.glm function. This function accounts for uncertainty about the variables to be included in the model using the Bayesian information criterion (BIC) approximation to the posterior model probabilities. The function does an exhaustive search over the model space using the fast leaps and bounds algorithm.
Our goal was to select and evaluate models composed of variables from the AFB, HRB, and Trails B to differentiate the control group from the neurologic group. FT, GPB, and H-Words appeared in every model. Consistent statistical support (with posterior probabilities of model inclusion) was found for only FT (100%), GPB (100%), H Words (100%), and Trails B (57%). When considered in the context of this set of tests, there was minimal support for the HRB tests typically considered to be among the best sensitive to brain dysfunction, e.g., the Category Test and TPT.

Table 1 displays the means and standard deviations for the subtests and scores of the AFB, and Table 2 displays these data for the HRB, and Trails B. Both tables include Cohen’s $d$ (pooled) for the group differences. Cohen’s $d$ ranged from a low of .26 for Similarities to a high of 1.08 for GPB for the tests comprising the AFB. Cohen’s $d$ ranged from −0.18 for FT to .87 for TPT time for the tests comprising the HRB. Trails B yielded an effect size of .89.

### DISCUSSION

We have demonstrated that an AFB, comprised of subtests to represent eight domains of function commonly evaluated in a flexible approach to neuropsychological evaluation, is equivalent in sensitivity to brain dysfunction, compared to the more traditional HRB. This replicates prior research demonstrating equivalent, if not superior, sensitivity to brain dysfunction of non-HRB measures such as WIS subtests, and measures of memory, in comparison to HRB scores (Loring & Larrabee, 2006; Powell et al., 1991; Rohling et al., 2003). Moreover, our current results likely represent a lower-bound estimate of AFB sensitivity, given our use of older, less sensitive measures of working memory, and
verbal and visual learning and memory. Had we employed more sensitive measures of working memory such as Auditory Consonant Trigrams (Stuss, Stethem, & Poirier, 1987) or Letter–Number Sequencing (Wechsler, 1997a, 1997b), more sensitive measures of verbal supraspan learning such as the AVLT or California Verbal Learning Test-II (CVLT-II, Delis, Kramer, Kaplan, & Ober, 2000), and more sensitive measures of visual memory such as the Rey Osterreith Complex Figure (Meyers & Meyers, 1995; Osterreith, 1944) or CVMT (Trahan & Larrabee, 1988), we anticipate that an even greater advantage of the AFB over the HRB would have been clearly demonstrated.

Additional analysis directed at identifying the most sensitive subset of test procedures specified four measures: H-Words, GPB, FT, and Trails B. It is noteworthy that all four tests require speeded cognitive or psychomotor performance. Measures of reaction time have a long history of sensitivity to brain dysfunction in neuropsychology (cf. Benton & Joynt, 1959; Blackburn & Benton, 1955). Consequently, procedures requiring speed of mental processing are among the most sensitive neuropsychological tasks (Lezak et al., 2004). Of note, the largest effect sizes in Tables 1 and 2 are for GPB and Trails B.

Also of interest is the identification of FT in the four-variable model. This is somewhat puzzling, given the data in Table 2 showing that FT has the smallest effect size, \(d' = -0.18\), which is also in the opposite direction of all other effect sizes, with the brain-dysfunction participants performing slightly better than the controls. By contrast, Table 1 shows that the controls outperformed the brain-dysfunction participants on GPB. This atypical pattern (poorer simple compared to complex motor performance) for the control vs brain-impaired participants is reminiscent of the atypical motor performance pattern in persons suspected of malingering (Greiffenstein, Baker, & Gola, 1996). Thus, although we attempted to control for

<table>
<thead>
<tr>
<th>Variable</th>
<th>Controls</th>
<th>Brain dysfunction</th>
<th>Cohen’s (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category Test</td>
<td>(M = 56.12)</td>
<td>(M = 75.21)</td>
<td>0.83</td>
</tr>
<tr>
<td>((SD = 21.83))</td>
<td>((SD = 24.49))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPT(^b) Time</td>
<td>(M = 18.53)</td>
<td>(M = 27.36)</td>
<td>0.87</td>
</tr>
<tr>
<td>((SD = 9.65))</td>
<td>((SD = 10.70))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPT Memory</td>
<td>(M = 6.90)</td>
<td>(M = 5.66)</td>
<td>0.74</td>
</tr>
<tr>
<td>((SD = 1.71))</td>
<td>((SD = 1.59))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPT Location</td>
<td>(M = 3.04)</td>
<td>(M = 2.06)</td>
<td>0.44</td>
</tr>
<tr>
<td>((SD = 2.44))</td>
<td>((SD = 1.94))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finger Tapping</td>
<td>(M = 43.74)</td>
<td>(M = 45.57)</td>
<td>-0.18</td>
</tr>
<tr>
<td>((SD = 8.88))</td>
<td>((SD = 12.18))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech Sounds</td>
<td>(M = 6.46)</td>
<td>(M = 10.35)</td>
<td>0.76</td>
</tr>
<tr>
<td>((SD = 4.81))</td>
<td>((SD = 5.53))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seashore Rhythm</td>
<td>(M = 3.78)</td>
<td>(M = 5.63)</td>
<td>0.50</td>
</tr>
<tr>
<td>((SD = 3.09))</td>
<td>((SD = 4.39))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trails B</td>
<td>(M = 87.57)</td>
<td>(M = 127.78)</td>
<td>0.89</td>
</tr>
<tr>
<td>((SD = 34.77))</td>
<td>((SD = 55.63))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) HRB = Halstead Reitan Battery; \(^b\)TPT = Tactual Performance Test.
poor effort excluding individuals with abnormally low Digit Span, or abnormally large differences between Vocabulary and Digit Span, it appears that participants with reduced effort may remain in the present sample. This suggests that FT is functioning as a moderator variable (sensitive to poor effort) per its selection as one of the four best discriminators in the Bayesian Model Averaging analysis.

As noted in the introduction, Rohling et al. (2003) found that an AFB, the Meyers Neuropsychological Battery, yielded essentially identical T scores in TBI patients of varying severity, collapsed across all subtests, to the T scores found for an expanded HRB. This showed the equivalence, on a global basis, of an aggregation of independently normed tests, to a conormed HRB. This still does not address the issue of determining patterns of differential strengths and weaknesses between and among various dimensions of ability; in other words, the frequency at which a particular difference occurs in the population. Fortunately, a statistical formula exists for determination of the rarity of a difference between two test scores, A and B, which is dependent on the standard deviation of difference scores (Payne & Jones, 1957):

$$\sqrt{\sigma_A^2 + \sigma_B^2 - 2\rho_{AB}\sigma_A\sigma_B}$$

To employ this formula, one needs to know the intercorrelations of a set of scores, as well as the standard (z or T) scores for the various scores. With these data, the difference between scores A and B is compared to the standard deviation of difference scores, to yield a z score representative of the frequency at which that difference score occurs in the population. The T or z scores are available for measures comprising an AFB, based on the individual test norms. What is typically not available is the intercorrelation among the various measures of memory, attention, processing speed, etc. that comprise an AFB. There is, however, a data set for which information is available on the intercorrelation of a number of key neuropsychological constructs: the WAIS-III/WMS-III Technical Manual (Wechsler, 1997c). Various tables present data on the intercorrelation of the Verbal Comprehension Index, Perceptual Organization Index, Working Memory Index, and Processing Speed Index of the WAIS-III with one another, and with the various WMS-III indexes. Moreover, WAIS-III and WMS-III correlations with measures of executive function, fine motor speed and dexterity, language and spatial processing, are also reported in the WAIS-III/WMS-III technical manual. We suggest that these correlational data can be employed to provide estimates of the intercorrelation of the constructs measured by a particular collection of well-validated measures comprising an AFB. These estimates can then be employed in the formula for the rarity of a difference score between two domains of function, (e.g. verbal intelligence and problem solving vs verbal memory), for those clinicians interested in obtaining such information.

In summary, we provide additional evidence that an AFB characteristic of the core of a flexible battery is equal and slightly superior to the HRB in sensitivity to brain dysfunction. Moreover, selection of the best four subtests from the combined data set of AFB and HRB subtests, plus Trails B, only identified one of the HRB measures, which appeared to be functioning more as a measure of effort and response validity, than as a measure sensitive to brain dysfunction. Clearly, our data
are consistent with those of Rohling et al. (2003) in supporting the current mainstream practice of the AFB (flexible) approach to neuropsychological evaluation. By contrast, the present data, prior research by Powell et al. (1991), and research reviewed by Loring and Larrabee (2006) do not demonstrate that the HRB is superior to well-standardized measures of memory and intellectual function. We conclude that the clinician utilizing an AFB approach that captures key constructs of language, sensorimotor function, working memory, processing speed, verbal and visual memory, and verbal and visual intelligence and problem-solving skills is practising in an equally valid manner as someone who relies on the HRB that is augmented with measures of language, verbal abilities, learning, and memory function (e.g., Dikmen et al., 1995; Heaton, Miller, Taylor, & Grant, 2004; Russell & Starkey, 1993). An AFB approach following the above constraints is most likely superior to approaches that do not augment the HRB with measures of language and memory function (e.g., Reitan & Wolfson, 1993), particularly given the superior sensitivity to brain dysfunction of verbal supraspan learning tests such as the AVLT or Selective Reminding, compared to the HRB, as demonstrated by Powell et al. (1991), and Dikmen et al. (1995).

ACKNOWLEDGMENT

We are grateful to Elbert W. Russell, Ph.D., who provided us with the data to conduct the present investigation.

REFERENCES


**APPENDIX**

Following the notation of Hoeting, Madigan, Raftery, and Volinsky (1999), let $\Delta$ be the quantity of interest, such as the probability of cognitive impairment. We are interested in the posterior distribution of $\Delta$ given the data, $D$. We average over all possible models, $M_k; k = 1 \ldots K$, where $M_k$ indicates a unique subset of neuropsychological tests:

$$pr(\Delta|D) = \sum_{k=1}^{K} pr(\Delta|M_k, D)pr(M_k|D).$$ (1)
This is a weighted average of the posterior distributions under each of the models considered, where the weights are the posterior model probabilities, \( pr(M_k | D) \). These posterior probabilities are given by:

\[
pr(M_k | D) = \frac{pr(D | M_k) pr(M_k)}{\sum_{i=1}^{I} pr(D | M_i) pr(M_i)} ,
\]

where \( pr(M_i) \) is the prior for model \( M_i \), and

\[
pr(D | M_k) = \int pr(D | \theta_k, M_k) pr(\theta_k | M_k) d\theta_k
\]

is the marginal likelihood under model \( M_k \) (with parameters \( \theta_k \)). Although Equation (1) suggests that we will be averaging over all of the models in our model space, Madigan and Raftery (1994) showed that Equation (1) is well approximated by averaging over a small group of parsimonious data-supported models, substantially reducing computational complexity.