The Clinical Neuropsychologist
2003, Vol. 17, No. 3, pp. 289–302

1385-4046/03/1703-289$16.00
© Swets & Zeitlinger

Neuropsychological Impairment Following Traumatic Brain Injury: A Dose-Response Analysis

Martin L. Rohling[1], John E. Meyers[2], and Scott R. Millis[3,4]

[1]Department of Psychology, University of South Alabama, Mobile, AL, USA, [2]Mercy Rehabilitation Center, Sioux City, IA, USA, [3]Kessler Medical Rehabilitation Research and Education Corporation, West Orange, NJ, USA, and [4]Department of Physical Medicine and Rehabilitation, University of Medicine and Dentistry of New Jersey, Newark, NJ, USA

ABSTRACT

Dikmen, Machamer, Winn, and Temkin (1995) administered the Halstead–Reitan Battery (HRB) to a sample of TBI patients. Similar patients were obtained from the second author (JEM) for two main purposes. First, we wished to determine if there is a dose-response relationship between TBI severity and residual cognitive deficit. Second, are Dikmen et al. results generalizable to other TBI samples? Analyses of the Meyers sample replicated the analyses of Dikmen sample. A significant dose-response relationship between loss of consciousness (LOC) and cognitive impairment was found using effect sizes for the Dikmen sample, as well as using regression-based normative T scores for the Meyers sample. The two methods were highly correlated with one another. Using mean scores for the six LOC-severity groups and the two samples resulted in a correlation coefficient $r = .97$, $p < .0001$. Results are presented for clinicians to use when assessing individual patients.

In a recent study, Dikmen, Machamer, Winn, and Temkin (1995) administered the Halstead–Reitan Battery (HRB; Reitan & Wolfson, 1993), as well as supplemental tests of memory, to traumatic brain injured (TBI) patients to assess their residual cognitive deficits. They examined 436 adult TBI patients, which they separated into six groups based on their severity of injury defined by patients' time to follow commands after the injury (TFC; e.g., raise your hand, stick out your tongue). Dikmen et al. (1995) found several associations between TFC and test scores, such that patients who had longer TFCs performed more poorly on their test battery than did patients with shorter TFCs. Dikmen et al. presented "expected" degrees of impairment on the each of 20 dependent variables, as well as on the Halstead Impairment Index (HII; Reitan & Wolfson, 1993). As a result, a patient with a specific TFC could be compared to a group of similar patients to determine whether his or her test scores were within the range of expectation. Dikmen et al. used the HII summary score to explore the influence of TFC, Glasgow Coma Scale (GCS), neurosurgical interventions, non-reactive pupils, and complications during hospitalization.

As Larrabee (1990) suggested, clinical history and background information from the medical record on initial injury severity can be used to estimate the expected range of scores for a particular level of severity. The particular patient's level of performance and pattern of performance must make "neuropsychological sense." This is one of the primary benefits of Dikmen et al.'s (1995) findings for clinicians involved in forensic cases. Once a clinician has ruled out complications

Address correspondence to: Martin L. Rohling, Department of Psychology, University of South Alabama, 381 LSCB, Mobile, AL 36688, USA. Tel.: +1-251-460-7675. E-mail: mrohling@usouthal.edu
Accepted for publication: July 7, 2003.

(i.e., cortical laceration by in-driven bone fragments, posttraumatic seizure disorder, and/or late onset hydrocephalus), a particular patient's set of scores can be compared to a normative group of similarly injured patients. If the scores obtained are far below expectation, there is reason to be suspicious of symptom exaggeration. For example, a patient who obtained a HII of 0.80 and whose TFC was less than 1 hr falls below the 10th percentile for patients with similar TFCs. Furthermore, this poor of a performance is near the mean for patients who have TFCs of 14–28 days (median = 0.72). A score this low, after a milder injury, is thus unlikely. This may indicate that the patient did not perform to the best of his or her ability. If financial incentives or other types of secondary gain could be identified in such a case, malingering should be considered as an alternative or additional explanation for this patient's poor performance.

To our knowledge, only one independent group of researchers has replicated the findings of Dikmen et al. (1995). Volbrecht, Meyers, and Kaster-Bundgaard (2000) used much of the same methodology as Dikmen et al. However, their primary purpose was to establish the validity of the Meyers Short Battery (MSB), not to replicate Dikmen et al.'s results. Furthermore, neither research group's findings are as frequently used by clinicians as they might be. For one, Dikmen et al.'s outcomes were generated using a battery of tests that are less commonly used in clinical practice as they once were (e.g., HRB, WAIS, & Buschke Selective Reminding Test; Sweet, Moberg, & Suchy, 2000). Second, because only 10% of the dependent variables for the two studies overlapped, Volbrecht et al. (2000) did not try to relate their results quantitatively to those of Dikmen et al. (1995). Such an analysis may have substantiated the associations between the severity of injury and degree of residual cognitive impairment between these two samples and test batteries.

Generating expected impairments on the HRB is most useful for the clinician who routinely administers the HRB. However, several recent practice surveys have found that the majority of neuropsychologists no longer use the HRB (Milberg, Hebben, & Kaplan, 1986; Sweet et al., 2000). Alternative batteries to the HRB have been praised for their individualized focus, psychometric ease, time efficiency, and utilization of advances in cognitive neuroscience (e.g., Lundin & DeFilippis, 1998). However, they have also been criticized both within the profession and in legal contexts (Hartlage, 2001; Reitan & Wolfson, 2001) for several reasons. First, these batteries do not yield a standard measure of impairment similar to what exists for the HRB. Second, results obtained with different combinations of tests are not easy to compare to empirically derived normative data, such as that which exists for the HRB (Heaton, Grant, & Matthews, 1991; Reitan & Wolfson, 1999; Russell, Neuringer, Goldstein, 1970). Finally, without global indices of performance and norm-referenced comparisons, test scores from alternative batteries for TBI patients are difficult to compare to the data in Dikmen et al.'s (1995) study. Therefore, the majority of clinicians are unable to determine whether the impairments obtained from a particular patient are consistent with the patient's medical history, as found by Dikmen et al. It is precisely this information that is essential for clinical diagnosis.

Recently, Rohling and colleagues have generated a strategy for deriving a global index from a more flexible patient battery (see, e.g., Miller & Rohling, 1998, 2001; Rohling & Miller, 1998; Rohling, Langhinrichsen-Rohling, & Miller, 2003). Specifically Rohling's Interpretive Method (RIM) was generated to respond to the recommendations of Wedding and Faust (1989) to improve upon the diagnostic judgments of neuropsychologists, particularly those who use flexible batteries. Furthermore, the RIM allows for the summarization of test data such that comparison can be more easily made to published comparison groups. The RIM is based on the principles of meta-analysis and is in keeping with the procedures articulated by Faith, Allison, and Gorman (1996). Miller and Rohling (2001) suggested that the linear combination of scores they recommend is supported by the research of Dawes and Corrigan (1974). Furthermore, it follows a model presented by Kiernan and Matthews (1976) and is similar to the methods used by Heaton et al. (2001) in their examination of the

stability and reliability of neuropsychological test scores in schizophrenia. Rohling and colleagues have claimed that their summary indices perform as well as or better than the HII, AIR, and the GNDS (Miller & Rohling, 2001; Rohling, Langhinrichsen-Rohling, & Miller, 2003; Rohling, Williamson, Miller, & Adams, in press). In the current study, it is proposed that these summary indices could be used to replicate the Dikmen et al. (1995) results using data generated from an alternative battery of more commonly used clinical instruments.

Specifically, these questions were addressed. First, is there a relationship between the time a patient has remained unconsciousness (i.e., loss of consciousness or LOC) and his or her residual cognitive deficits? Second, if a dose-response relationship exists, can the group statistics be summarized such that clinicians can better use these research findings to make better assessment decisions; increasing the accuracy of diagnosis and advance treatment planning? Because we expected to find a TBI dose-response relationship that was evident in two samples (i.e., Dikmen's and Meyers' samples), we wanted to generate regression lines that would provide information to neuropsychologists as to the nature of TBI and the process of recovery. Furthermore, regardless of the test battery administered, clinicians may be able to use the associations generated, and accompanying percentiles, to determine the probability that a particular patient's performance is causally linked to a TBI. Such data would be of particular importance in forensic settings.

Before our hypotheses can be examined, the issue of which battery should be examined needed to be considered. In the current study, we initially analyzed the Meyers Short Battery (MSB), presented in Volbrecht et al. (2000), as a proxy for the "typical" neuropsychological battery often used in clinical practice. The MSB has been described as a "core" battery of newer and more commonly administered neuropsychological tests. In the current sample, each patient was essentially administered a similar battery of tests in a standardized format. This allowed us to make comparisons across patients and groups with greater confidence in the results.

STUDY 1

METHOD

Participants: Dikmen Sample

Participants are well-described in Dikmen et al. (1995). Briefly, the sample was generated from a Level I Trauma Center and included 436 consecutive admissions of adult TBI patients. Patients were recruited at the time of their injury into one of three on-going studies (i.e., Behavioral Outcome Study, Patient Characteristics Study, and the Dilantin Prophylaxis Study). A weighting procedure was used to adjust for differences in the proportions of participants who were included from each of these three studies. Patients' age mean age was 28.9 ($SD = 12.5$), with an average of 12.0 years of education ($SD = 2.3$). Seventy-two percent of the sample was male. A trauma control (TC) group was also included. This sample's mean age was 31.2 ($SD = 12.5$), with an average of 12.0 years of education ($SD = 2.6$). Seventy percent of the TC group was male. There were no significant differences between the TBI sample and the Trauma Control (TC) group on any of the demographic variables.

Acute Indices of Severity for Dikmen Sample (Independent Variables)

Patients were assigned to one of six severity groups based on their time to follow commands (TFC). This was defined as the highest score on the motor portion of the Glasgow Coma Scale (GCS). Furthermore, following patients' presentation to the emergency room, each had to be admitted to the hospital. The six severity groups were defined as follows. Group 1's TFC was less than 1 hr. Group 2's TFC was greater than or equal to 1 hr but less than 1 day. Group 3 consisted of patients with TFCs that were greater than or equal to 1 day but less than 7 days. Group 4 consisted of patients with TFCs that were greater than or equal to 7 days but less than 14 days. Group 5 included patients with TFCs that were greater than or equal to 14 days, but less than or equal to 28 days. Finally, Group 6 included patients with TFCs that were greater than 28 days. Dikmen et al. included a control group that consisted of patients who were also injured in an accident, but whose injuries did not involve the head, neck, or central nervous system. These "trauma controls" (TC) were used to calculate effect sizes that are described below.

Time from TBI Until Assessment: Dikmen Sample

As required by the research protocols in which patients were enrolled, each patient had been assessed at 1 and 12 months postinjury. Dikmen et al. (1995) noted that

92% of the patients were assessed within 1 month of their 1-year injury anniversary and 99% were assessed within 2 months of their 1-year date. Our analyses were focused on test scores generated from patients' 1-year assessment. The fact that each patient was assessed twice likely differentially confounds our effect size and *T* score calculations. This point was also asserted by Dikmen et al., who were concerned that practice effects were not likely to be evenly distributed across severity group, with greater improvement in performance more likely to occur in the least severely injured patients (p. 87). They also noted the increase in patients scores from the 1-month to the 1-year assessment. For example, TC patients increased their PIQ score by 5 points ($g = 0.37$), their Category Test scores by 10 points ($g = 0.35$), their VIQ by 1.2 points ($g = 0.09$), and their HII by 0.05 points ($g = 0.18$). Using these four dependent variables, patients' average practice effect was 0.25 standard deviation units, which is equivalent to 2.8 *T* score points.

Objective Neuropsychological Tests for Dikmen Sample (Dependent Variables)

The test battery centered on the Halstead–Reitan Battery (HRB), with additional measures of memory and attention supplementing the basic battery. Specifically, several instruments were administered to patients and used to generate 20 dependent variables for the analyses. Intellectual verbal and performance skills were assessed using the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). Additional performance skills were assessed using the Tactual Performance Test (TPT), with the dependent variable being the total time per block variable (Reitan & Wolfson, 1993). Reasoning ability was assessed with the Category Test (CT), with the dependent variable being total number of errors obtained (Reitan & Wolfson, 1993). Verbal memory measures included the Wechsler Memory Scale (WMS; Wechsler, 1945) Logical Memory Immediate and 30 min. Delayed, and the Buschke-Selective Reminding Test (SRT; Buschke, 1973), which included the total number of items recalled (RCL), as well as free recall at 30-min and 4-hr delays. Visual memory was assessed using the WMS Visual Reproduction immediate and 30-min delay conditions. Measures of attention, concentration, and cognitive flexibility included the Seashore Rhythm Test (number correct), Stroop Test Parts 1 and 2 (time to complete; Dyer, 1973), Trail Making Test – Parts A and B (time to complete), and the Finger Tapping Test (FTT; number of taps for dominant and nondominant hands).

Dependent Variables for Dikmen Sample: Effect Sizes

Similar to the methods used by Binder, Rohling, and Larrabee (1997), an effect size was calculated using Hedges' *g*. The *g* statistic is calculated by dividing the difference between the experimental and control groups' means by their pooled standard deviation. When standard deviations are pooled, it generally increases the denominator in this equation compared to what would result if only the control group's standard deviation were used. In effect, this causes the effect size estimator to be smaller than if only the control group's standard deviation was used. Effect sizes were used so that each raw score mean was on a common metric, facilitating the combining and contrasting of dependent variables. Effect sizes were then linearly combined, as has been suggested by Dawes (1979), Dawes and Corrigan (1974), and Heaton et al. (2001). The overall test battery mean (OTBM-*g*) was calculated by averaging all dependent variable effect sizes within a sample.

Data Analysis

Parametric statistics were chosen for the analyses. In particular, parametric analysis of variance and linear regression were used to compare severity groups and the dependent variables. When it was considered informative, nonparametric analyses were also conducted to ensure that the most appropriate methods were used (e.g., Pearson's *r* and Spearman's *Rho*). However, when results of these two procedures were similar in magnitude, we chose to report the results that were more user-friendly for clinicians.

RESULTS

Comparisons of TBI Severity Groups: Effect Sizes

Results for TBI severity from the Dikmen sample are shown in Table 1 and are graphically depicted in Figure 1. There was a significant main effect for severity, $F(5, 114) = 206.1$, $p < .0001$. The variances for effect sizes for each severity group were also compared statistically. There were significant differences revealed. The two mildest severity groups' variances were equivalent to one another. Both of these groups' variances were significantly smaller than the remaining four severity groups. Finally, the four more severely injured groups' variances were not significantly different from one another. Due to the heterogeneity of variance across the severity groups, the analysis was repeated using the nonparametric Kruskal–Wallis procedure. Results were similar, $H(5, 114) = 99.3$, $p < .0001$, with a significant

Table 1. Injury Severity Using Effect Sizes (Hedges' g) for Dikmen's Sample.

Severity group	n	Mean	SD	Minimum	Median	Maximum
Group 1: TFC < 1 hr	161	−0.02	(0.13)	−0.26	−0.02	−0.20
Group 2: TFC = 1–23 hr	100	−0.22	(0.14)	−0.47	−0.23	−0.03
Group 3: TFC = 1–6 days	52	−0.45	(0.22)	−0.94	−0.42	−0.09
Group 4: TFC = 7–13 days	37	−0.68	(0.27)	−1.21	−0.66	−0.15
Group 5: TFC = 14–28 days	32	−1.33	(0.26)	−1.67	−1.26	−0.89
Group 6: TFC > 28 days	53	−2.31	(0.41)	−2.88	−2.24	−0.85
Entire sample	435	−0.82	(0.80)	−2.88	−0.56	0.20

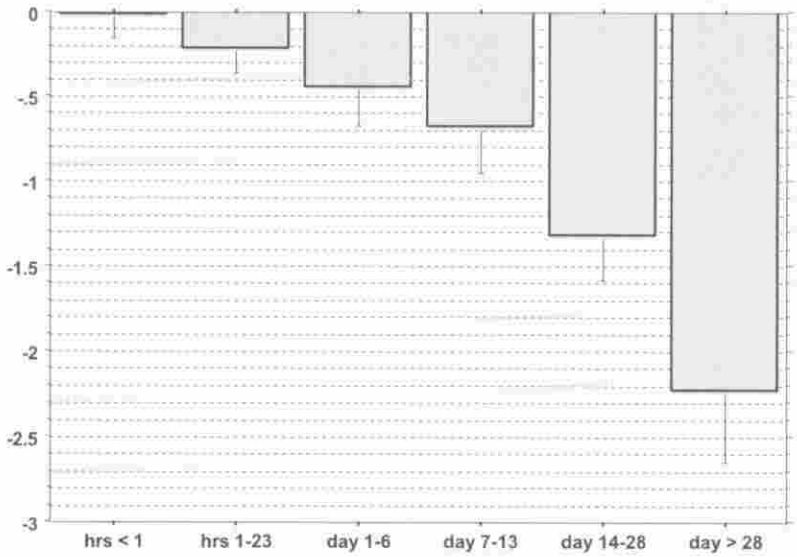

Fig. 1. Effect size mean from Dikmen sample based on severity defined by TFC.

omnibus test. Furthermore, comparisons amongst the six severity groups, using both parametric and nonparametric procedures, revealed significant differences amongst each group.

Simple Linear Regression of Severity Group on Effect Sizes

The effect size data were examined using simple linear regression, with the independent variable of injury severity, defined by group assignment (i.e., 1–6), regressed onto the effect sizes using the 20 dependent variables. The association was highly significant, $F(1, 119) = 665.1$, $p < .0001$, with Pearson r of −.93, which accounted for 86% of the variance. The slope of the data was -0.276 ± 0.021 (95% CI). Again, due to the heterogeneity of variance in effect sizes across severity groups and the categorical nature of the independent variable, the effect size data was also analyzed using the nonparametric Spearman procedure. Results were similar as with the parametric findings, with a *Rho* of −0.94, $Z = 9.65$, $p < .0001$.

STUDY 2

METHOD

Participants: Meyers Sample

Originally, the sample consisted of the 150 patients whose summary statistics were published in Volbrecht et al. (2000; see Rohling, Millis, & Meyers, 2000). However, to better characterize the sample, the first author (MLR) contacted the second author (JEM) to access to the original data. Raw test scores were

supplied for 323 consecutive patients, each of whom had been referred for assessment, diagnosis, and treatment planning. Access to the raw test scores provided us with an opportunity to statistically correct for symptom exaggeration in this sample. Following the recommendations of Iverson and Binder (2000; see also Iverson & Franzen, 1996), participants with invalid data were removed from the sample. Participants were excluded using procedures recommended by Meyers and Volbrecht (2003). The remaining sample of valid cases included 291 patients. All patients were treated at a Level I Trauma Center in the rural Midwest (Mercy Medical Center in Sioux City, IA). However, unlike the Dikmen sample, not all of these patients had been previously admitted to the hospital. Many were seen weeks or months following discharge from an inpatient unit, which may have been months or years following their TBI. Some patients had never been inpatients, as they had been discharged home from the emergency room. The mean age for the sample was 32.8 ($SD = 14.7$), with the average length of time from TBI until referral of 2.7 years. The typical patient was injured at the age of 30.1 years, which is similar to the mean age of 28.9 years for the Dikmen sample. Patients in the sample had an average of 12.0 years ($SD = 2.4$) of education, which is identical to that of the Dikmen sample. Overall, the sample was 69% male, which is nearly identical to the 70% observed in the Dikmen sample.

The Meyers sample did not include a Trauma Control (TC) group. However, each patient had two estimates of premorbid intelligence. The first estimate was based on the Barona, Reynolds, and Chastain (1985) demographic regression equation. The second estimate was based on patients' reading recognition scores on the National Adult Reading Test (NART). The mean estimated WAIS–R FSIQ using the Barona equation was 102.7 ($SD = 6.6$) and using the NART was 97.2 ($SD = 8.7$). The combined average of these two estimates of premorbid general ability was 100.7 ($SD = 6.4$).

Acute Indices of Severity for Meyers Sample (Independent Variables)

Patients were assigned to one of six severity groups, which were the same as those defined for the Dikmen sample. However, because the Meyers sample is one of "convenience," the data coded were not "Time to Follow Commands" but loss of consciousness (LOC). This is the more commonly recorded piece of information in patients' medical records. Each patient's LOC was retrieved from his or her hospital medical record. Table 2 presents descriptive statistics for the LOC, as well as for posttraumatic amnesia (PTA). PTA was defined as the time from initial injury until the patient could consistently recall episodic events following the accident. As often as possible, patients' self-report on this variable was verified by checking the medical record. However, for the majority of patients, PTA could not be validated, as it was not obvious from their medical record. As a result, there is considerable unreliability or error in this variable.

Time from TBI Until Assessment: Meyers Sample

Because patients in this sample were referred for clinical evaluation and management, rather than being a participant in a research protocol, there was greater variability in the amount of time that had passed from injury until assessment than found for the Dikmen sample. Refer to Table 2 for descriptive statistics

Table 2. Severity of Injury Information for the Meyers' Sample.

	Group 1 ($n = 138$)	Group 2 ($n = 52$)	Group 3 ($n = 47$)	Group 4 ($n = 19$)	Group 5 ($n = 14$)	Group 6 ($n = 21$)	All patients ($N = 291$)	H	p
LOC (days)								274.6	<.0001
Mean	0.01	0.17	2.5	9.0	17.3	44.6	5.4		
Median	0.01	0.09	2.0	8.0	16.0	30.0	0.02		
(*SD*)	(0.004)	(0.18)	(1.5)	(1.9)	(3.7)	(21.8)	(14.4)		
PTA (days)								143.8	<.0001
Mean	3.9	9.1	17.3	36.7	44.0	97.3	18.2		
Median	1.0	5.0	7.0	28.0	30.0	90.0	4.0		
(*SD*)	(8.6)	(10.9)	(26.3)	(55.1)	(34.5)	(74.0)	(40.7)		
Mo. since TBI								61.2	<.0001
Mean	13.1	10.7	26.4	37.2	81.3	126.5	32.6		
Median	4.0	3.0	6.0	10.0	21.0	65.5	3.0		
(*SD*)	(28.5)	(18.8)	(66.5)	(71.1)	(121.4)	(136.3)	(79.4)		

regarding this variable. The influence of this time on neuropsychological test scores is difficult to estimate. Some patients may still have been experiencing neurological recovery at the time of the assessment. The data for the "time since injury until assessment" were positively skewed (skewness = 3.81), with a mean of 32 months, which was far greater than the median of 3 months and the mode of 1 month. The standard deviations for each group on this variable were large. However, 54% of the patients were tested within the first 3 months of their TBI, and 68% of patients were assessed within the first 12 months of their TBI. Both parametric (Pearson's *r*) and nonparametric (Spearman's *Rho*) analyses failed to find any association between the time from TBI until assessment and test scores, so no patients were excluded from the analysis based on this measure.

Worth noting is that the majority of patients who were tested within three months of injury had been assigned to the two mildest injury-severity groups. Specifically, 47% of the patients tested within three months postinjury had been assigned to Group 1 (TFC < 1 hr) and 27% had been assigned to Group 2 (TFC = 1–23 hr). This totaled 74% of the entire sample. An additional 19% of patients assessed within 3 months postinjury had been assigned to Group 3 (TFC = 1–6 days). Only 7% of the sample assessed within 3 months postinjury had been assigned to the remaining three severity groups, with none of these patients being assigned to the most severely injured group (TFC > 28 days). As a result, the Meyers sample was likely a predominantly "neurologically-recovered" TBI sample.

Objective Neuropsychological Tests for Meyers Sample (Dependent Variables)

The sample was assessed using the MSB. The MSB includes many of the most commonly administered assessments instruments. There were 26 dependent variables analyzed from the MSB. For example, intellectual assessment included the Wechsler Adult Intelligence Scale – Revised or Third Edition (Pilgrim, Meyers, Bayless, & Whetstone, 1999; Wechsler, 1981, 1997). Verbal skills were typically assessed using the Information and Similarities subtests. Performance skills were assessed using the Picture Completion and Block Design subtest. Reasoning ability was assessed using the Booklet Category Test (CT; DeFilippis & McCambell, 1991; Kozel & Meyers, 1998). Auditory memory was assessed using the Auditory Verbal Learning Test (AVLT; Spreen & Strauss, 1998). Visual memory was assessed using the Rey Complex Figure Test (RCFT; Meyers & Meyers, 1995), which included the immediate and delayed recall scores, as well as the recognition score. Attention or working memory was assessed using Digit Span and Arithmetic subtests, as well as the Trail Making Test – Parts A and B. Processing and psychomotor speed were assessed using the Digit Symbol subtest and Finger Tapping.

Dependent Variable for Meyers Sample: *T* Scores

Regression-based *T* scores were calculated, resulting in each dependent variable raw score being converted to a *T* score ($M = 50$ and $SD = 10$). Specifically, each raw score was adjusted for the patient's age, education, sex, and handedness. The regression procedures used are the same as those used by Heaton et al. (1991) in their HRB scoring and interpretation system. Specifically, regression-based norms for each dependent variable were generated from an independent sample of patients and control subjects. The resulting equations were used to generate *T* scores for the TBI participants to correct for the influence of age, education, gender, and handedness. The results were compared to norms generated by using the Heaton et al. (1991) when possible. Correlation coefficients between the two different methods, examining six dependent variables (i.e., FSIQ, Trails, FTT, and CT), ranged between .82 and .95, with an average of .92. None of the intercepts significantly differed from 0 and none of the regression line slopes significantly differed from 1.0. Furthermore, the mean and standard deviation for all TBI patients in the Meyers sample ($M = 41.1$, $SD = 7.2$) was quite similar to that obtained by Heaton et al. (2001) for a sample of patients who suffered from chronic schizophrenia ($M = 41.9$, $SD = 6.5$). Therefore, we proceeded to demographically correct the remainder of the dependent variables using the same regression-based procedure. Specifically, we transformed each patient's raw scores into *T* scores for all 26 dependent variables. These 26 *T* scores were linearly combined to generate an overall test battery mean (OTBM-T) for each patient that was demographically corrected.

RESULTS

Comparisons of TBI Severity Groups: *T* scores

Means and standard deviations for each of the six severity groups are shown in Table 3 and are graphically illustrated in Figure 2. An ANOVA of these data revealed a significant main effect for severity, $F(5, 285) = 29.5$, $p < .0001$. Due to the heterogeneity of variance across groups and the categorical nature of the independent variable, the analysis was repeated using a nonparametric procedure (Kruskal–Wallis). As with the Dikmen

Table 3. Injury Severity Using *T* Scores for Meyers' Sample.

Severity group	*n*	*M*	*SD*	Minimum	Median	Maximum
Group 1: LOC < 1 hr	138	44.3	(5.0)	31.8	44.9	54.6
Group 2: LOC = 1–23 hr	52	41.3	(5.7)	30.7	42.0	52.4
Group 3: LOC = 1–6 days	47	40.0	(6.6)	28.3	39.4	49.7
Group 4: LOC = 7–13 days	19	36.5	(7.1)	24.2	38.4	48.5
Group 5: LOC = 14–28 days	14	33.7	(5.3)	22.6	34.5	42.3
Group 6: LOC > 28 days	21	30.3	(8.6)	10.6	31.2	40.7
Entire sample	291	41.1	(7.2)	10.6	42.3	54.6

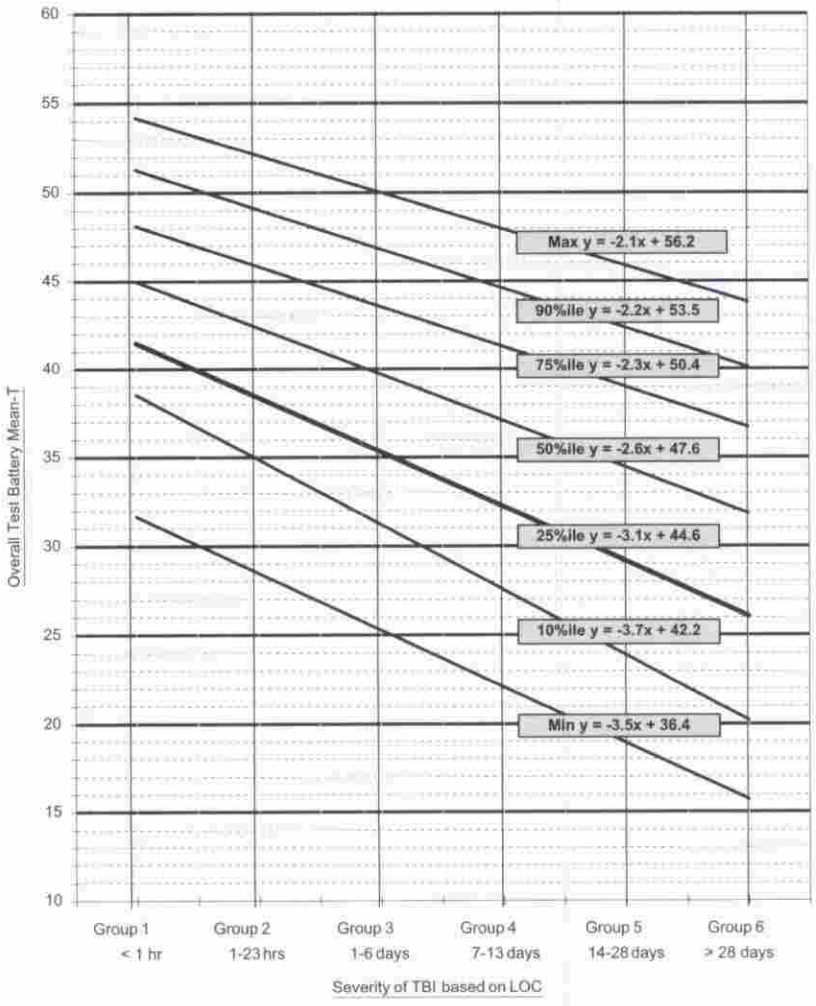

Fig. 2. Overall Test Battery Mean (OTBM) using *T* scores from the Meyers sample.

sample, results were quite similar between the parametric and nonparametric procedures, $H(5, 285) = 84.1$, $p < .0001$, with a significant difference in both omnibus tests. Planned pairwise comparisons amongst the severity groups using both procedures were also consistent, and

with four exceptions, revealed 26 significant differences out of 30 pairwise group comparisons. The four exceptions were that for both procedures, Groups 2 and 3 were not significantly different from one another, nor were Groups 4 and 5 significantly different from one another. The differences between these two sets of scores were all in the expected direction, but these last four were not of sufficient magnitude to reach statistical significance.

Simple Linear Regression of Severity Group on *T* Scores

Refer to Figure 2 for a graphic illustration of the regression analyses, which includes the typical percentile-divisions used in box plots. As with the Dikmen sample, the graph was generated using simple linear regression, with the independent variable of injury severity, as defined by Group assignment (i.e., 1–6), was regressed onto the OTBM of the 291 patients from the Meyers sample. The results were highly significant, $F(1, 290) = 146.6$, $p < .0001$, with a Pearson r of $-.58$, which accounted for 34% of the variance. Using the 95% confidence interval, the intercept was $47.1 + 1.2$, and the slope was -2.7 ± 0.4. The corresponding Spearman's *Rho* for this relationship was $-.72$, which was also highly significant, $Z = -8.81$, $p < .0001$. Worth noting is that these results are quite similar to that reported by Dikmen et al. (1995) with the HII ($r = -.59$).

Comparison of Results from Study 1 and Study 2

Two problems exist that makes it difficult to compare the results of these two samples. First, there is only summary data from the Dikmen sample, whereas the Meyers sample includes patient's raw scores. We attempted to obtain raw data from Study 1's authors, but were unsuccessful. Furthermore, we do not have a trauma control group to generate effect sizes for the Meyers sample, as was available for the Dikmen sample. Second, the operational definition of the most severely impaired group was not the same between the two studies. In the Dikmen sample, patients' who were sufficiently impaired that they could not complete the assessment battery had their scores estimated. However, these types of patients were excluded from the Meyers sample. As a result, Group 6 is more severely impaired in the Dikmen sample than in the Meyers Sample. To make the two samples comparable, we excluded the results of Group 6 and assumed that the effect size for the least injured group was essentially zero (Binder et al., 1997). Then, we set the OTBM-T for the trauma control group of Dikmen et al. (1995) to be equivalent to that of the least injured group of the Meyers sample. Using this estimated OTBM-T for the trauma control group, we then generated a series of estimated OTBM-T's for the first five groups of the Dikmen sample. These results are presented in Table 4. The mean level of performance was not

Table 4. Comparison Between Mean OTBM-T for Meyers' Sample and Estimated OTBM-T for Dikmen Sample, Excluding the Most Severely Impaired Group.

Severity group	Meyers sample	Dikmen sample
Group 1: LOC < 1 hr	44.3	44.1
Group 2: LOC = 1–23 hr	41.3	42.1
Group 3: LOC = 1–6 days	40.0	39.8
Group 4: LOC = 7–13 days	36.5	37.5
Group 5: LOC = 14–28 days	33.7	31.0
Sample mean of six groups (*SD*)	39.2 (4.1)	38.9 (5.1)
Correlation coefficient[a]	.99	.96
Slope of regression line (95% CI)	−2.6 (±.6)	−3.1 (±1.6)
Intercept of regression line	47.0 (±2.0)	48.1 (±5.3)

Note. [a]Simple linear regression, with the independent variable being severity group assignment and the dependent variable being level of performance, as operationally defined by the OTBM-T.

significantly different between the Meyers and Dikmen samples (39.1 vs. 38.9, respectively). The correlation coefficient between the two samples was .97 ($p = .0067$), with a slope that was not significantly different from 1.0 and an intercept that is not significantly different from 0. This suggests that the two samples are in a high degree of agreement and are likely assessing the same dose-response relationship.

Percentile Classification Accuracy Using T Scores

T scores from the Meyers sample were used to calculate percentiles for each severity group. The percentiles plotted were chosen because they are the same as those used in the box plots of the HII presented by Dikmen et al. (1995). Specifically, there are seven cut points provided on the graph (i.e., minimum score, 10th, 25th, 50th, 75th, and 90th percentiles, and the maximum score). The actual patient data were smoothed using regression procedures to remove random variability across severity groups. The regression lines generated by these analyses created seven regions on the graph: (1) less than the minimum score, (2) the minimum score to 10th percentile, (3) 10th to 25th percentile, (4) 25th to 75th percentile, (5) 75th to 90th percentile, (6) 90th to maximum score, and (7) greater than the maximum score.

Actual percentiles for each patient were then checked against their assignment into the graphic regions. Results revealed that 81% of the patients had been assigned to the correct region of the graph based on their true percentile within the sample. Of the patients who were incorrectly assigned, 11% were assigned to a graphic region that was one lower than the one in which they belonged. Of these lower-placed patients, half scored below the 25th percentile and four (1.4% of the total sample) were assigned to the region that was below the minimum score of their severity group. Of these 4 patients, 2 were in Group 1, 1 was in Group 2, and 1 was in Group 6. Furthermore, the 4 patients were only slightly below each minimum score, with each being less than 1.5 *T* score points below the minimum line.

Regarding the remaining incorrectly assigned patients, 8% were assigned to a region that was one higher than they belonged. However, of these higher-placed patients, only 3 (1% of the total sample) scored below the 25th percentile. Of these 3, all were from Group 3. Finally, 9% of patients were incorrectly assigned to a percentile region that was one farther from the sample median; whereas 10% of the patients were incorrectly assigned to a percentile region that was one closer to the sample median. These "farther" incorrectly assigned patients were distributed evenly across the six severity groups.

DISCUSSION

The results of the current study are clear. When TBI researchers or clinicians use similar operational definitions for the independent variable (e.g., time to follow commands) and the dependent variable (e.g., memory loss as measured by a published, standardized, and well-normed psychometric instrument), the association between TBI severity and residual cognitive deficits will be revealed.

One reason why our results are of particular importance is that most neuropsychologists do not use the HRB (see, e.g., Sweet et al., 2000). This limits what many of them can do with the findings of Dikmen et al. (1995) regarding the dose-response relationship of TBI. The majority of the clinicians use a "flexible battery" approach. However, the major limitation of the flexible battery approach is that it is difficult to compare test scores across individuals and groups. The current study found that generating an OTBM provides a way to compare TBI patients' global residual cognitive ability with that which would be expected based on the severity of their injury.

In addition, the current study serves as a validation of the OTBM as a global index of neuropsychological functioning generated by the RIM (Heaton et al., 2001; Miller & Rohling, 2001; Rohling, Langhinrichsen-Rohling, & Miller, 2003). The correlation between the effect sizes generated from the HII and the OTBM-T from the Meyers sample confirms that the OTBM serves the same purpose for the MSB as does the HII does for the HRB. Heaton et al. (2001) used a similar method to substantiate the stability of neuropsychological

deficits in schizophrenia. In their study, the OTBM (i.e., Global Neuropsychological *T* score) was found to have greater reliability than both the AIR and the FSIQ. It is possible that the OTBM will provide the most useful estimate of current cognitive functioning for the clinician working with TBI patients. Further research will be needed to substantiate this.

Application of Findings to Clinical Assessment

Use of the OTBM-T is generally simple in its calculation and application. First, each dependent measure is translated to a similar metric (e.g., *T* scores). These are then averaged across the dependent measures that were administered in the test battery.[1] Then, the results of the present study can be used to determine the relative congruity between acute injury and residual cognitive deficits, which is independent of the assessment battery used by a clinician. However, this last statement is true only under certain conditions. The reader might wonder, can a clinician calculate "any old OTBM" and get the same results? No! We believe that the similarity between the two samples used in the present study occurred only because both batteries comprehensively assessed the primary domains of cognition with normed instruments that are well standardized. We have conducted research to substantiate this fact, with two manuscripts in press (Rohling, Langhinrichsen-Rohling, & Miller, 2003; Rohling et al., in press) to substantiate our claim.

[1]A simpler method was also advocated by Rohling and colleagues (Miller & Rohling, 2001; Rohling et al., in press). The Instrument Test Battery Mean (ITBM) is the average of a single measure from each instrument administered. The most appropriate measure selected for inclusion into the ITBM is either a summary score (e.g., FSIQ for the WAIS-III) or the most sensitive subtest score of the instrument (e.g., total score across five trials for the AVLT). Research conducted by Rohling et al. (in submission) has found correlations between the OTBM and the ITBM of .95. One can adequately substitute the ITBM for the OTBM with little concern about its effect on the subsequent analyses.

Regarding our results, one must keep in mind that before a patients' set of scores can be compared to our results, other complications must be ruled out. For example, such complicating factors such as cortical laceration by in-driven bone fragments, posttraumatic seizure disorder, and/or late onset hydrocephalus can all cause greater impairment in cognition than one might predict based on a simple uncomplicated TBI. If these factors can be ruled out, then comparison to the present results is feasible.

The following example will illustrate how our results might be used in an individual case to determine the probability of the level of performance occurring with the level of acute injury. Using a patient's LOC, which can usually be obtained from the medical record, a clinician can find the severity group in which the patient best belongs. Using Figure 2, the patient's group can be found on the horizontal axis (*X*-axis). Then, the clinician should move upward on the graph until the group's median OTBM *T* score is found. This will be the most likely degree of impairment for a patient falling in the patient's severity group. Next, the 10th and 90th percentiles should be located for the patient's particular severity group. The clinician can then check to see if their patient's OTBM falls within this range of scores. If the patient's OTBM falls below the 10th percentile, and particularly if it falls below the minimum score for the entire severity group, the clinician should investigate other potential variables that may have impacted the patient's test scores (e.g., suboptimal performance, psychiatric disorder, or premorbid dysfunction).

For example, if a patient were unconscious for 18 hr following a TBI, this would place him or her in the second severity group. Using the equations supplied on Figure 2, the median OTBM *T* score for this group is 42.4, and the 10th to 90th percentile range is 34.8 and 49.3, which is a range of ± 7.3 *T* score points. If the patient's OTBM equaled 27.0, this is less than the minimum score for all patients in this severity group. The clinician may then suggest that this degree of impairment is statistically unlikely, based on the patient's severity of injury.

A similar procedure can be followed if the neuropsychologist knows the patient's OTBM-T,

based on his or her performance on a comprehensive test battery, and wants to calculate the severity of TBI that would most likely result in such a score. Using Figure 2, this can be accomplished by finding the patient's OTBM-T on the vertical axis (*Y*-axis). This *T* score can then be followed to the left until it intersects the 50th percentile regression line. At this point, go down the graph vertically to the horizontal axis (*X*-axis) to determine the severity group to which the patient is most likely to belong. This procedure can be made more precise by algebraically rearranging the regression equations in Figure 2, so that the severity group number is isolated on the left side of the equation and the patient's OTBM is entered into the right side of the equation. For example, using the 50th percentile line, the equation is as follows: severity group number = [(47.6 − OTBM)/2.6]. Therefore, if a patient obtained an OTBM-T of 27.0, as given in the above example, the calculation for a group assignment would be greater than six. This severity group had an LOC of greater than 28 days. Furthermore, between 10% and 25% of the patients in the next less severely injured group (i.e., LOC = 14–28 days) scored this low and less than 10% of the patients in the fourth severity group (i.e., LOC = 7–13 days) scored this low. Finally, no one in the first or second severity group should be expected to score this low. Thus, the probability of belonging to a particular severity group can be compared to the patient's medical records to see how much of a discrepancy there is between expectation and reality. In the example above, the patient's medical record documented an LOC of 18 hr. According to the graph, the expected LOC for this amount of deficit should be greater than 28 days. Given that the patient was unconscious for a much shorter amount of time than was anticipated, the clinician should be encouraged to consider other factors that might explain the patient's remarkably poor performance.

Summary

Overall, the current study also provides some validity data for the use of the MSB. The comparison between the HRB and the MSB indicates that both batteries measure similar neuropsychological functions and have similar sensitivities to neurological lesions. We believe this to have been the result of the comprehensive nature of both assessment batteries, which includes measures of the primary domains of cognition (Larrabee, 1990). There were no substantial differences between the MSB and HRB, as utilized by Dikmen et al. (1995), across the TBI severity groups. Specifically, the slope of the regression line, when effect sizes were transformed into an OTBM-T estimate, indicated that the two samples were highly congruent in their results. In addition, the two test batteries were similar in terms of their variance in effect sizes and *T* scores across the severity groups. The correlation between the samples (.97) suggests that the batteries are equivalent in their ability to assess different levels of cognitive impairment is similar to the findings of Volbrecht et al. (2000). It is anticipated that any comparable battery with same level of normative data as the MSB and assessing the same cognitive skills would produce similar results. Of course, further research is required to substantiate this claim, as would be true with any flexible or semi-flexible batteries. What minor differences exist in the slope and intercept of the regression equations generated with the Meyers and Dikmen sample are likely attributable to practice effects in the Dikmen sample, which likely differentially influenced the more mildly injured patients, which was not evident in the Meyers sample.

REFERENCES

Barona, A., Reynolds, C.R., & Chastain, R. (1984). A demographically based index of pre-morbid intelligence for the WAIS–R. *Journal of Consulting and Clinical Psychology, 52*, 885–887.

Binder, L.M., Rohling, M.L., & Larrabee, G.J. (1997). A review of mild head trauma: Part I. A meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology, 19*, 421–431.

Buschke, H. (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior, 12*, 543–550.

Dawes, R.M. (1979). The robust beauty of improper linear models in decision-making. *American Psychologist, 34*, 571–582.

Dawes, R.M., & Corrigan, B. (1974). Linear models in decision-making. *Psychological Bulletin, 81*, 95–106.

DeFilippis, N., & McCambell, E. (1991). *The Booklet Category Test manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Dikmen, S.S., Machamer, J.E., Winn, H.R., & Temkin, N.R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology, 9*, 80–90.

Dyer, F.N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory and Cognition, 1*, 106–120.

Faith, M.S., Allison, D.B., & Gorman, B.S. (1996). Meta-analysis of single case research. In R.D. Franklin, D.B. Allison, & B.S. Gorman (Eds.), *Design and Analysis of Single-Case Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hartlage, L.C. (2001). Neuropsychological testing of adults: Further considerations for neurologists. *Archives of Clinical Neuropsychology, 16*, 201–214.

Heaton, R.K., Gladsjo, J.A., Palmer, B.W., Kick, J., Marcotte, T.D., & Jeste, D.V. (2001). Stability and course of neuropsychological deficits in schizophrenia. *Archives of General Psychiatry, 58*, 24–32.

Heaton, R.K., Grant, I., & Matthews, C.G. (1991). *Comprehensive Norms for an Extended Halstead-Reitan battery*. Odessa, FL: Psychological Assessment Resources, Inc.

Iverson, G.L., & Binder, L.M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation, 15*, 829–858.

Iverson, G.L., & Franzen, M.D. (1996). Using multiple objective procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology, 18*, 38–51.

Kiernan, R.J., & Matthews, C.G. (1976). Impairment index versus *T*-score averaging in neuropsychological assessment. *Journal of Consulting and Clinical Psychology, 44*, 951–957.

Kozel, J.J., & Meyers, J.E. (1998). A Cross-Validation Study of the Victoria Revision of the Category Test. *Archives of Clinical Neuropsychology, 13*, 327–332.

Larrabee, G.J. (1990). Cautions in the use of neuropsychological evaluation in legal settings. *Neuropsychology, 4*, 239–247.

Lundin, K.A., & DeFilippis, N.A. (1998). Proposed schedule of usual and customary test administration times. *The Clinical Neuropsychologist, 13*, 433–436.

Meyers, J.E., & Meyers, K.R. (1995). *Rey Complex Figure Test and Recognition Trial: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Meyers, J.E., & Volbrecht, M.E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology, 18*, 261–276.

Milberg, W.P., Hebben, N., & Kaplan, E. (1986). The Boston process approach to neuropsychological assessment. In I. Grant & K.M. Adams (Eds.), *Assessment of Neuropsychiatric Disorders*. New York: Oxford University Press.

Miller, L.S., & Rohling, M.L. (1998). Statistical interpretative methods for difficult neuropsychological test data. *Archives of Clinical Neuropsychology Abstract, 12*, 86–87.

Miller, L.S., & Rohling, M.L. (2001). A statistical interpretive method for neuropsychological test data. *Neuropsychology Review, 11*, 143–169.

Pilgrim, B.M., Meyers, J.E., Bayless, J.J., & Whetstone, M.M. (1999). Validity of the ward seven-subtest WAIS-III short form in a neuropsychological population. *Applied Neuropsychology, 6*, 243–246.

Reitan, R.M., & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation* (2nd ed.). Tucson, AZ: Neuropsychological Press.

Reitan, R.M., & Wolfson, D. (1999). *Issues in forensic neuropsychology*. Workshop presented at the annual National Academy of Neuropsychology.

Reitan, R.M., & Wolfson, D. (2001). Clinical evaluation of "Assessment: neuropsychological testing of adults". *Archives of Clinical Neuropsychology, 16*, 215–226.

Rohling, M.L., Langhinrichsen-Rohling, J., & Miller, L.S. (2003). Statistical methods for determining malingering. In R. Franklin (Ed.), *Prediction in forensic and neuropsychology: Sound statistical practices*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rohling, M.L., & Miller, L.S., (1998). Rohling's interpretive method. *Journal of the International Neuropsychological Society Abstract, 4*, 26.

Rohling, M.L., Miller, L.S., & Langhinrichsen-Rohling (in press). Rohling's interpretive method for neuropsychological data analysis: A response to critics. *Neuropsychology Review*.

Rohling, M.L., Millis, S.R., & Meyers, J.E. (2000). Generating a linear function of residual impairment from TBI: A comparison between the HRB and a flexible battery. *Archives of Clinical Neuropsychology Abstract, 15*, 821–822.

Rohling, M.L., Williamson, D.J., Miller, L.S., & Adams, R. (in press). Using the Halstead Reitan battery to diagnose brain damage: A comparison of the predictive power of traditional techniques to Rohling's interpretive method. *The Clinical Neuropsychologist*.

Russell, E.W., Neuringer, C., & Goldstein, G. (1970). *Assessment of brain damage: A neuropsychological key approach*. New York: Wiley–Interscience.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms,*

and commentary (2nd ed.). New York: Oxford University Press.

Sweet, J., Moberg, P., & Suchy, Y. (2000). Ten-Year follow-up survey of clinical neuropsychologists: Part I. Practices and beliefs. *The Clinical Neuropsychologist, 14*, 38–55.

Volbrecht, M.E., Meyers, J.E., & Kaster-Bundgaard, J. (2000). Neuropsychological outcome of head injury using a short battery. *Archives of Clinical Neuropsychology, 15*, 251–265.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19*, 87–95.

Wechsler, D. (1955). *Manual for Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale – Revised*. New York: Psychological Corporation.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale-III*. New York: Psychological Corporation.

Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4*, 233–265.

Copyright of Clinical Neuropsychologist is the property of Swets & Zeitlinger, BV and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.