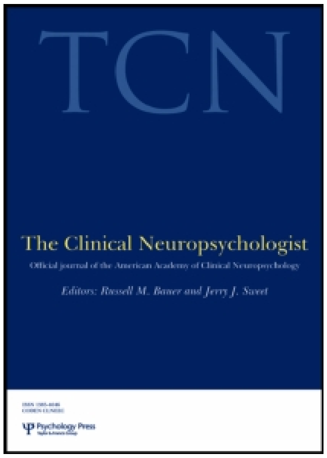

The Clinical Neuropsychologist
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713721659

Detecting Insufficient Effort Using the Seashore Rhythm and Speech-Sounds Perception Tests in Head Injury

Scott R. Ross [a]; Steven H. Putnam [b]; Scott R. Millis [c]; Kenneth M. Adams [d]; Rebecca A. Krukowski [e]
[a] DePauw University, Greencastle, IN, USA
[b] Wayne State University School of Medicine, Detroit, MI, USA
[c] Kessler Medical Rehabilitation Research and Education Corporation, University of Medicine and Dentistry of New Jersey - New Jersey Medical School, Newark, NJ, USA
[d] Ann Arbor VAMC and University of Michigan Medical Center, Ann Arbor, MI, USA
[e] University of Vermont, Burlington, VT, USA

Online Publication Date: 01 December 2006
To cite this Article: Ross, Scott R., Putnam, Steven H., Millis, Scott R., Adams, Kenneth M. and Krukowski, Rebecca A. (2006) 'Detecting Insufficient Effort Using the Seashore Rhythm and Speech-Sounds Perception Tests in Head Injury', The Clinical Neuropsychologist, 20:4, 798 - 815
To link to this article: DOI: 10.1080/13854040500328477
URL: http://dx.doi.org/10.1080/13854040500328477

The Clinical Neuropsychologist, 20: 798–815, 2006

ISSN: 1385-4046 print/1744-4144 online
DOI: 10.1080/13854040500328477

Psychology Press
Taylor & Francis Group

DETECTING INSUFFICIENT EFFORT USING THE SEASHORE RHYTHM AND SPEECH-SOUNDS PERCEPTION TESTS IN HEAD INJURY

Scott R. Ross[1], Steven H. Putnam[2], Scott R. Millis[3], Kenneth M. Adams[4], and Rebecca A. Krukowski[5]
[1]*DePauw University, Greencastle, IN, USA,* [2]*Wayne State University School of Medicine, Detroit, MI, USA,* [3]*Kessler Medical Rehabilitation Research and Education Corporation, University of Medicine and Dentistry of New Jersey—New Jersey Medical School, Newark, NJ, USA,* [4]*Ann Arbor VAMC and University of Michigan Medical Center, Ann Arbor, MI, USA, and* [5]*University of Vermont, Burlington, VT, USA*

This study examined the capacity of the Seashore Rhythm Test (SRT) and the Speech-Sounds Perception Test (SSPT) to detect insufficient effort in a clinical sample. Forty-six participants with financially compensable mild head injury who obtained scores indicative of insufficient effort on multiple measures were compared to 49 participants with brain injury who were not involved in litigation. Receiver operating characteristic (ROC) curve analysis indicated that both the SRT ($AUC = .84$) and SSPT ($AUC = .80$) were significant ($p < .001$) predictors of insufficient effort. Maximizing sensitivity and specificity, the optimal cutoff scores were 8 errors on the SRT and 10 errors on the SSPT. Combining both variables into a logistic regression function increased the diagnostic efficiency.

INTRODUCTION

The assessment of malingering, insufficient effort, or response bias in forensic settings has become an important part of the neuropsychological evaluation of patients with suspected head injury. A plethora of investigations examining the detection of persons attempting to feign cognitive impairment have been published over the last decade. Indeed, the assessment of malingering seems almost a preoccupation among neuropsychologists (Ross & Adams, 1999). Although malingering has become a rather high-profile diagnosis, this focus seems warranted. Despite considerable variability in estimates of malingering, the base rate appears high. In a recent examination of over 33,000 annual cases from numerous referral sources, Mittenberg, Patton, Canyock, and Condit (2002) estimate that as many as 39% of mild head injury patients may be malingering. These findings, though troublesome, are not inconsistent with previous studies. For instance, Schmand et al. (1998) found that 61% of litigating participants suffering whiplash injury appeared to be engaging

Address correspondence to: Scott R. Ross, DePauw University, Department of Psychology, 7 Larabee St., Greencastle, IN 46135. Tel.: 765-658-4576. Fax: 765-658-4572. E-mail: srross@depauw.edu
Accepted for publication: August 29, 2005.

in symptom overreporting. Additionally, Rogers, Salekin, Sewell, Goldstein, and Leonard (1998) estimated the incidence of malingering in forensic settings at 17%, based on a survey of forensic experts.

THE SEASHORE RHYTHM AND SPEECH SOUNDS PERCEPTION TESTS

Attempts to identify patients who malinger have resulted in various measures specifically designed to assess insufficient effort (Validity Indicator Profile; VIP; Frederick, 1997; Ross & Adams, 1999; Portland Digit Recognition Test; PDRT; Binder & Willis, 1991; Test of Memory Malingering; TOMM; Rees, Tombaugh, Gansler, & Moczynski, 1998; Amsterdam Short-Term Memory Test; ASTM; Schagen, Schmand, de Sterke, & Lindeboom, 1997; and the Dot Counting Test; DCT; Hiscock, Branham, & Hiscock, 1994) and symptom validity (Structured Interview of Reported Symptoms; SIRS; Rogers, Gillis, & Bagby, 1990). Although such instruments may be highly useful, the need for economy in test selection and administration has given way to the development of indices for identifying malingering on tests typically used to assess neuropsychological status, rather than effort. These include tests of general intellectual ability (Wechsler Adult Intelligence Scale—Revised; WAIS-R; Millis, Ross, & Ricker, 1998; Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner, 1995), memory (California Verbal Learning Test; CVLT; Millis, Putnam, Adams, & Ricker, 1995; Baker, Donders, & Thompson, 2000; Sweet, Wolfe, et al., 2000; Memory Assessment Scales; MAS; Beetar & Williams, 1995; Ross, Krukowski, Putnam, & Adams, 2003; Wechsler Memory Scales—Revised; WMR-R; Bernard, McGrath, & Houston, 1993; Iverson, Slick, & Franzen, 2000; Mittenberg, Azrin, Millsaps, & Heilbronner, 1993) motor functioning (Greiffenstein, Baker, & Gola, 1996; Rapport, Farchione, Coleman, & Axelrod, 1998), executive functioning (Forrest, Allen, & Goldstein, 2004; Tenhula & Sweet, 1996), and psychological disturbance, such as the MMPI-2 (Larrabee, 1998, 2003; Miller & Donders, 2001; Ross, Millis, Krukowski, Putnam, & Adams, 2004; Tsushima & Tsushima, 2001). Two measures commonly administered as part of the Halstead-Reitan Neuropsychological Test Battery are the Speech-Sounds Perception Test (SSPT; Halstead, 1947) and Seashore Rhythm Test (SRT; Reitan & Wolfson, 1993). Both measures have a forced-choice response format that lends itself to the identification of insufficient effort. These tests appear more difficult than they are (Trueblood & Schmidt, 1993) and provide a statistical basis for determining chance responding and operationally defining insufficient effort in the neuropsychological exam (Charter, 1994; Millis, 1992).

Studies have generally supported the use of the SSPT and SRT in the detection of probable malingering and insufficient effort. In an early study, Heaton, Smith, Lehman, and Vogt (1978) found that clinically identified malingerers performed significantly worse compared to nonlitigating head trauma patients on the Speech-Sounds Perception Test (SSPT). In an examination of the Halstead Impairment Index, Mutchnick, Ross, and Long (1991) similarly reported positive findings for the SRT and SSPT; of the six tests included in this index, they were the two that had the best predictive validity for detecting malingering. Further, Trueblood and Schmidt (1993) examined the utility of 63 neuropsychological variables in the detection of malingering. They found that errors on the SRT were most useful in discriminating between malingerers and those judged to have test scores of questionable validity, from control participants. Trueblood and Schmidt identified an optimal cutoff score of 9 errors, which accurately classified 56% of the malingerers and 50% of those in the questionable validity group. In addition, malingering participants were found to differ significantly from controls on the SSPT, as expected. Similar classification rates to the SRT were found for a cutoff score of 18 errors on the SSPT and resulted in optimal diagnostic accuracy.

Mittenberg, Rotholc, Russell, and Heilbronner (1996) also reported that simulators obtained significantly more errors on the SSPT and SRT, but did not provide specific guidelines for the use of the SSPT and SRT for identifying probable malingering. In an effort to more closely examine the diagnostic utility of the SRT, Gfeller and Cradock (1998) employed an analogue dissimulation paradigm. They found that both sophisticated simulators (given information regarding typical symptoms of traumatic brain injury) and naïve simulators (receiving no information) had significantly more errors on the SRT as compared to control and bona fide patient groups. However, sophisticated simulators and naïve simulators did not differ significantly. A cut score of 7 was recommended by these researchers; this cut score was found to have an overall classification rate of 80.7%, correctly classifying 72.5% of the simulating and 85% of control and head injury participants.

CURRENT STUDY

Although previous investigations of the SRT and SSPT support the use of these tests in the detection of malingering, these studies are not without their limitations. For example, Trueblood and Schmidt's study was based on a sample size of only 8 patients suspected of malingering. Although Gfeller and Cradock (1998) and Mittenberg et al. (1996) employed larger samples in their investigations, it is unclear to what extent their results generalize to clinical patients. Both studies employed "dissimulators," or persons instructed by the experimenter to act like a malingerer. The problem with analogue studies, however, is that it is unclear to what extent persons *instructed* to malinger are effective in simulating the performance of real-life patients who are likely to feign symptoms. To address previous limitations regarding sample size and composition, we examined the SRT and SSPT in a sample of 95 participants who had been referred for clinical evaluation. This sample size allowed for the determination of classificatory rates for the SRT and SSPT using receiver operating characteristic (ROC) curve analysis. ROC curves allow one to examine the performance of a test over a wide range of scores and select a cutoff score with optimal diagnostic accuracy. Consequently, we report the classificatory rates for a number of previously suggested cutoff scores in identifying insufficient effort in MHI. Additionally, as advocated by Millis and Volinsky (2001) and Rohling, Langhinrichsen-Rohling, and Miller (2003), we used *multiple* measures to identify participants who are exhibiting insufficient effort in the neuropsychological exam. Although a number of indices have been developed for identifying insufficient effort or otherwise invalid performance in the neuropsychological exam, we used indices from instruments commonly employed by clinicians (e.g., the MMPI-2 and WAIS-R). In addition to performance on the Recognition Memory Test (RMT; Warrington, 1984), participants were also selected on the basis of their performance on the Wechsler Adult Intelligence Scale—Revised (WAIS-R; Wechsler, 1981) and pattern of symptom reporting on the Minnesota Multiphasic Personality Inventory, 2nd ed. (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Although "insufficient effort" is certainly not synonymous with "malingering," we believed that including multiple indices (e.g., multiple test scores, compensation-seeking status, injury severity) in participant selection would increase confidence in our selection procedure for identifying probable malingerers.

METHODS

Participants

The probable malingering group (PM) was composed of 46 participants (age: $M = 40.2$ years; $SD = 12.0$; education: $M = 12.0$; $SD = 2.4$; race: 63.0% were White; gender: 50.0% were men) with alleged head injury who had external incentives to present as cognitively impaired and who were referred for neuropsychological examination to a private practice center in the Midwest. All participants had mild to very mild head injuries (PTA < 60 min; Bigler, 1990). Although no participants admitted to us that they were malingering, our characterization of this group as demonstrating "insufficient effort" appeared to be reasonable. All PM participants were actively pursuing personal injury or workers' compensation litigation due to alleged impairment following head injury at the time that they were examined: 76% were in a motor vehicle accident, 11% had fallen, and 13% claimed injury under other conditions. Further, all participants appeared to exhibit a pattern of performance consistent with insufficient effort on multiple measures. Only PM participants were included who obtained: (1) scores at or below 32 on the Words or Faces subtest, and at or below 40 (< 1st percentile of the normative sample) on the other subtest of the RMT; (2) a predicted score greater than .1011 using Mittenberg et al.'s (1995) discriminant function to identify malingering based on multiple WAIS-R subtests; and, (3) a raw score greater than or equal to 21 on the MMPI-2 Fake Bad Scale, consistent with recent findings by Ross et al. (2004). Descriptive statistics for selected MMPI-2, WAIS-R, and RMT indices are reported in Table 1. Overall, participants obtained a mean WAIS-R Full Scale IQ in the low average range ($M = 78.6$; $SD = 11.7$).

The head injury group (HI) was composed of 49 persons with moderate to severe traumatic brain injuries (age: $M = 37.5$ years, $SD = 14.1$; education: $M = 13.4$ years, $SD = 2.7$; race: 77.6% were White; gender: 59.2% were men) who were evaluated at an outpatient private practice clinic. At the time of examination, none of the HI participants were involved in litigation or seeking compensation. Mean WAIS-R Full Scale IQ was in the average range ($M = 95.4$; $SD = 14.2$). Of the HI participants, 62.2% had mild to very mild head injuries (PTA < 60 min), 32.7% had moderate injuries [post-traumatic amnesia (PTA) 1 to 24 hours]; and 8.1% had severe injuries (PTA greater than 1 day) (Bigler, 1990). PTA was estimated from emergency room and other hospital records, and in a minority of cases by patient self-report. Seventy-nine percent were injured in a motor vehicle accident, 10% were injured in a fall, and the remaining 11% were injured in a motorcycle accident, as a pedestrian in an auto accident, or by an explosion. HI participants were examined a mean of 27.1 months post-injury ($SD = 43.5$) with a median of 14.0

Table 1 Means and standard deviations for selected validity indices from the MMPI-2, WAIS-R, and RMT by group

	M	*SD*	*t*	*p*
MMPI-2 Validity Index				
Lie scale			2.63	.010
PM	62.54	10.81		
HI	56.49	11.42		
F-Infrequency scale			4.58	.000
PM	83.07	27.22		
HI	60.00	21.01		
K-Correction scale			.74	.460
PM	47.13	12.34		
HI	48.85	10.10		
F–K			3.52	.001
PM	.82	12.15		
HI	−7.11	9.33		
Fake-Bad Scale			11.44	.000
PM	29.57	5.05		
HI	17.37	5.00		
WAIS-R Validity Index				
Mittenberg Discriminant Function			10.31	.000
PM	.94	.59		
HI	−.77	.84		
Vocabulary-Digit Span			8.26	.000
PM	3.26	2.14		
HI	−1.56	2.88		
RMT Validity Index				
Words Subtest			15.32	.000
PM	26.74	6.91		
HI	45.45	4.68		
Faces Subtest			11.43	.000
PM	27.61	6.86		
HI	42.33	4.02		

Note. PM = Probable malingering group. HI = Head injury comparison group. MMPI-2 = Minnesota Multiphasic Personality Inventory (2nd ed.). WAIS-R = Wechsler Adult Intelligence Scale—Revised. RMT = Recognition Memory Test.

months. None obtained a score that fell within the 95% confidence interval for random responding (i.e., less than 32 correct) on the RMT Words subtest (see Table 1). However, some HI patients not included in the final sample did obtain scores within chance levels on the RMT. Additionally, while the vast majority of HI participants in the final sample obtained scores on the *FBS* and WAIS-R discriminant function that were below cutoffs for the PM group, not all did. For instance, 7 of the 49 HI participants obtained scores at or above 21 on the *FBS*, whereas 5 of the 49 HI participants obtained WAIS-R discriminant function scores above .1011. Of these participants, only 2 obtained scores above cutoffs on both the *FBS* and WAIS-R. Although eliminating these participants would result in a cleaner sample, it could potentially limit the generalizability of decision rules derived in the current study. We wanted our HI sample to be reasonably representative of patients with bona fide head injury.

Measures

Seashore rhythm test (Reitan & Wolfson, 1993). The SRT consists of 30 pairs of rhythmic beats presented on a tape recording where the participant is to determine whether the pair of rhythms is similar or different (Gfeller & Cradock, 1998). Bornstein (1982) reported adequate internal consistency, with split-half reliability coefficients ranging from .74 to .87.

Speech-sounds perception test (Halstead, 1947). The SSPT consists of 60 nonsense words that have the "ee" sound in the middle of the word. The words are spoken on a tape recording, and the participant chooses one of four written responses in an attempt to match a word to the sound (Reitan & Wolfson, 1990). Bornstein (1983) reported adequate internal consistency (Cronbach's $\alpha = .74$) across all 60 trials.

Recognition memory test (Warrington, 1984). The RMT consists of 100 dichotomous forced-choice items divided evenly over two subtests, Words and Faces. The RMT uses a recognition paradigm for assessing memory and has been found useful in identifying insufficient effort in patients presenting with symptoms of MHI (Millis, 1992, 1994; Millis & Putnam, 1994; Millis et al., 1995). In a recent review of the RMT in the detection of insufficient effort, Millis (2002) gives the following reasons for the utility of the RMT: (1) standard neuropsychological tests using a forced-choice response format have been used to detect response bias in a manner similar to symptom validity tests (e.g., Iverson & Franzen, 1994; Millis et al., 1995; Millis et al., 1998); (2) application of the binomial curve allows for the statistical determination of chance levels of performance; (3) the RMT demonstrates a robust ceiling effect in normative samples; and, (4) the RMT is generally insensitive to brain injury (Millis & Djikers, 1993; Bigler et al., 1996), including lateralized brain dysfunction (Kneebom, Chelune, & Luders, 1997; Sweet, Demakis, Ricker, & Millis, 2000). Using the binomial curve, the confidence interval for determining random responding yields a 95% confidence interval (two-tailed test) of 18 to 32 for a 50-item test (Charter, 1994). In the current study, an important selection criterion for the PM group was performance at chance (at or below 32) on one subtest and below the 1st percentile (below 40) on the other test of the RMT. These cutoff scores are consistent with original findings by Millis (1992, 1994) and Millis and Putnam (1994) and are somewhat more conservative than other studies using the RMT to identify insufficient effort (see Millis et al., 1995, 1998; Ross et al., 2004).

Minnesota multiphasic personality inventory-2 (MMPI-2; Butcher et al., 1989). A broad-band self-report measure of psychopathology and clinically relevant personality traits, the MMPI-2 is comprised of 13 Basic scales including three validity and ten clinical scales. Like the original MMPI, the MMPI-2 has been the subject of the development of a number of actuarial scales designed to discriminate between groups of interest. One validity scale that shows particular promise as an indicator of malingering in personal injury litigants is the Fake Bad Scale (*FBS*; Lees-Haley, English, & Glenn, 1991). Developed by Lees-Haley et al. to specifically detect somatic malingering in personal injury cases, the *FBS* consists of 43 items, selected on the basis of their content, using unpublished frequency counts of malingerers' MMPI test responses and observations of personal injury malingerers.

A number of studies support the validity of the *FBS* and find that the *FBS* is better than traditional validity scales such as *F* and *FBack* at identifying probable malingerers involved in personal injury litigation (Lees-Haley et al., 1991; Larrabee, 1998; Miller & Donders, 2001; Tsushima & Tsushima, 2001). Additionally, studies by Larrabee (2003) and Slick, Hopp, Strauss, and Spellacy (1996) have reported that elevations on *FBS* are related to poorer performance on neuropsychological tests. Most recently, in a large sample of participants with bona fide head injury and litigating MHI who obtained scores within chance levels on the RMT, Ross et al. (2004) found that a cutoff score of 21 provided high sensitivity (90%) and specificity (90%), using ROC curve analysis. The *FBS* and the cutoff score suggested by Ross et al. were used as a second test-based selection criterion in identifying PM participants.

Wechsler adult intelligence scale—revised (WAIS-R; Wechsler, 1981). The WAIS-R is composed of 11 subtests of various abilities and is used to determine an overall general intelligence quotient or Full Scale IQ (FSIQ). Five of the subtests (e.g., Vocabulary, Comprehension, Similarities) comprise a Verbal Intelligence Quotient (VIQ), whereas the remaining six comprise a Performance (or nonverbal) Intelligence Quotient (PIQ). The most widely used, individually administered test of intelligence, the WAIS and its revisions have been and continue to be a mainstay of cognitive assessment following head injury and other neurologic insult. Based on previous studies, Mittenberg et al. (1995) developed two indices from the WAIS-R that were effective in discriminating between patients with bona fide head injury and participants instructed to simulate impairment following head injury. The index that has shown most promise is a discriminant function based on scores for 8 subtests of the WAIS-R (Greve, Bianchini, Mathias, Houston, & Crouch, 2003; Axelrod & Rawlings, 1999). Millis, Ross, and Ricker (1998) found that this function generalized to clinical patients engaging in insufficient effort, with Axelrod and Rawlings (1999) reporting high levels of specificity in a longitudinal examination of traumatically brain-injured patients. We included this index as a final test-based selection criterion in determining PM group membership.

RESULTS

There were no significant differences between the PM and HI groups with regard to age, $t(93) = -1.00$, $p > .30$; sex, $\chi^2(1) = .80$, $p > .30$; race coded as "White" and "non-White" $\chi^2(1) = .27$, $p > .60$; months post-injury, $t(85) = .48$, $p > .60$. However, the PM group had fewer years of educational attainment than the HI group [$t(93 = 2.71, p < .01)$]. In addition, the PM group had obtained significantly lower Full Scale IQ scores than the HI group on the WAIS-R ($t(93) = 4.65$, $p < .05$). Consequently, we examined whether education or WAIS-R FSIQ was an important confound in our selection of participants. To this end, we conducted four ANCOVAs with group membership as our categorical variable, either the SRT or SSPT as the continuous variable, and either education or WAIS-R FSIQ as a covariate. Education did not achieve significance ($p > .05$) in relation to either the SRT or SSPT. However, WAIS-R FSIQ was significant in relation to the SRT ($p < .001$) as well as SSPT ($p < .005$). Despite differences between groups on FSIQ, premorbid estimates of IQ using the Barona Estimate (Barona, Reynolds,

& Chastain, 1984) and the North American Adult Reading Test (NAART; Blair & Spreen, 1989) did not significantly differ ($p > .05$) between PM and HI groups. Moreover, given that a validity index derived from the WAIS-R was used as a criterion for defining insufficient effort, it is understandable that the PM group's mean IQ would be lower than the mean for the HI group. Presumably, poor effort depressed scores on the WAIS-R. Consequently, we found no reason to "correct" for differences in IQ and believe that these findings speak favorably to our internal validity and study design.

We took two approaches to examining the use of the SRT and SSPT as measures of probable malingering. Inspired by previous research, we first sought to determine the ability of the SRT and SSPT to identify participants engaging in insufficient effort. In order to determine the diagnostic efficiency of these measures across a wide range of potential cutoff scores, Receiver operating characteristic (ROC) curve analysis was conducted for each test. When examining ROC curves, cutoff scores were selected such that both sensitivity and specificity were maximized. The selected cutoff scores will not correspond to the maximum value of sensitivity or specificity but will correspond to a point that minimizes the probability of both false positive and false negative classifications. In this study, sensitivity referred to the number of PM participants correctly classified; specificity was defined as the number of HI participants correctly classified.

It should be noted that there are several methods for determining the optimal operating point (OOP) or optimal decision rule for ROC curves (Gallop, Crits-Christoph, Muenz, & Tu, 2003). In addition to the simultaneous maximization of sensitivity and specificity, as in the present study, the OOP can be based on a preassigned value for sensivity or specificity. For example, the negative consequence of mistakingly diagnosing an individual as malingering may warrant setting specificity at a high level; e.g., 95%. Tables will be presented that will allow readers to use this second method for choosing preassigned values for sensitivity and specificity to determine the OOP for different diagnostic situations. A third method explicitly incorporates prevalence rates and the "cost" of making misclassfications. In this context, cost is expressed as a ratio of cost for false-positive errors divided by cost for false-negative errors (Gallop et al., 2003). Determing the OOP with this method requires collaboration with clinicians and other content experts to estimate the cost factor and prevalence reates. This method is known as the slope of isoutility.

A cutoff score greater than or equal to 8 errors on the SRT yielded a sensitivity of 76.1% and specificity of 73.5% (area under the curve or $AUC = .843$, $p < .001$; asymptotic 95% CI = .764 to .923; see Table 2). This cut score provided the highest overall classificatory accuracy of 74%. In contrast, a cutoff score greater than or equal to 6 errors as suggested by Gfeller and Cradock (1998) resulted in a reduced specificity of 65% though high sensitivity of 85%.

Consistent with the results of Trueblood and Schmidt (1993), the optimal cut score for the SSPT was greater than or equal to 10 errors ($AUC = .801$, $p < .001$; asymptotic 95% CI = .706 to .896; see Table 3). This resulted in 72% of the PM participants (sensitivity) and 76% of the HI participants (specificity) correctly classified, for an overall correct classification rate of 73%.

Our second approach used logistic regression analysis in order to evaluate the conjoint but unique contributions of the SRT and SSPT in the detection of

Table 2 Diagnostic classification rates from the receiver operating characteristic curve for seashore rhythm test (errors)

Cutoff score	Sensitivity	Specificity
0	1.000	0.000
1	1.000	0.020
2	1.000	0.163
3	0.935	0.265
4	0.913	0.429
5	0.891	0.531
6	0.848	0.653
7	0.848	0.694
8	0.761	0.735
9	0.696	0.857
10	0.587	0.918
11	0.544	0.959
12	0.435	0.959
13	0.304	0.980
14	0.217	1.000

Note. Cutoff scores are in error score units.

Table 3 Diagnostic classification rates from the receiver operating characteristic curve for Speech-Sounds Perception Test (errors)

Cutoff score	Sensitivity	Specificity
1	1.000	0.000
2	1.000	0.041
3	0.978	0.082
4	0.913	0.122
5	0.913	0.204
6	0.870	0.327
7	0.826	0.429
8	0.761	0.612
9	0.717	0.694
10	0.717	0.755
11	0.696	0.776
12	0.696	0.816
13	0.696	0.898
14	0.652	0.898
15	0.630	0.939
16	0.630	0.980
17	0.630	0.980
18	0.609	0.980
19	0.587	0.980
20	0.565	0.980
21	0.522	0.980
22	0.522	0.980
23	0.500	0.980
24	0.500	0.980
25	0.478	1.000

Note. Cutoff scores are in error score units.

Table 4 Logistic regression function with SSPT and SRT in predicting PM and HI groups

						95% CI for odds ratio	
Predictor	B	SE	Wald U	*p*	Odds ratio	Lower limit	Upper limit
SSPT	.117	.040	8.718	.003	1.125	1.040	1.216
SRT	.245	.080	9.264	.002	1.277	1.091	1.496
Constant	−3.321	.680	23.859	.000	.036		

Note. SSPT = Speech Sounds Perception Test. SRT = Seashore Rhythm Test.

insufficient effort. Logistic regression differs from ordinary least squares multiple regression in that the algorithm used to determine model fit is based on a logarithmic distribution in which the criterion variable is dichotomous. Because binary scaling of the criterion variable violates assumptions of linear regression (e.g., linearity and normality), ordinary least squares is inappropriate in estimating model parameters (Menard, 2002). A test of the full model with both SRT and SSPT against a constant-only model was statistically reliable, likelihood ratio χ^2 (2, $N = 95$) = 51.64, $p < .0001$, indicating that SRT and SSPT, as a set of predictor variables, reliably distinguished between persons with HI and persons exhibiting insufficient effort. The "variance" (or, more accurately, the proportional reduction in the absolute value of the log-likelihood measure) in group status accounted for by this model was moderate in magnitude, $R_L^2 = .39$. Table 4 shows the regression coefficients, odds ratios, and 95% confidence intervals for the odds ratios for both SRT and SSPT. According to the Wald criterion, both variables reliably predicted group status. Calibration refers to the extent to which the predicted probabilities agree with the observed probabilities. In this regard, the value of the Hosmer-Lemeshow goodness-of-fit statistic was 8.162 and the corresponding *p*-value was .32, which indicated that this model was reasonably well-calibrated.

In addition, ROC curve analysis for the logistic regression function produced classification rates slightly superior than either test alone ($AUC = .871$, $p < .001$). Differences between the logistic regression models containing both SRT and SSPT, and SRT and SSPT used individually, were further examined via the Bayesian information criterion (BIC) (Hardin & Hilbe, 2001). BIC is a measure of overall fit and can be used to compare nested and nonnested models. The formula is given by

$$\mathrm{BIC} = D(M_k) - (df)\ln(n)$$

where $D(M_k)$ is the model deviance, *df* is degrees of freedom, and *n* is the number of observations. The more negative the BIC, the better the fit. The BIC values for the logistic regression function, SRT alone, and SSPT alone were −338.989, −332.262, and −332.774, respectively. Raftery (1996) has provided guidelines for preferring one model over another based on the absolute difference in BIC. There is weak evidence for preferring one model over another when the absolute difference in BIC is 0 to 2; positive evidence, 2 to 6; strong evidence, 6 to 10; and very strong evidence when the difference is greater than 10. Hence, there is strong evidence to prefer the use of the logistic regression model containing both SRT and SSPT in lieu of the SRT or SSPT alone. In addition, use of the logistic regression function accounts

Table 5 Diagnostic classification rates from the receiver operating characteristic curve for predicted logistic regression probabilities

Cutoff score (probability)	Sensitivity	Specificity
0.000	1.000	0.000
0.052	1.000	0.020
0.058	1.000	0.041
0.069	1.000	0.061
0.081	1.000	0.082
0.086	1.000	0.122
0.091	1.000	0.143
0.096	1.000	0.163
0.101	0.978	0.204
0.106	0.978	0.225
0.107	0.978	0.245
0.112	0.978	0.265
0.118	0.978	0.286
0.119	0.957	0.286
0.120	0.957	0.306
0.121	0.935	0.306
0.127	0.935	0.327
0.133	0.913	0.327
0.140	0.913	0.347
0.148	0.913	0.388
0.155	0.913	0.408
0.163	0.913	0.429
0.171	0.913	0.449
0.179	0.913	0.490
0.180	0.913	0.531
0.190	0.913	0.551
0.209	0.891	0.551
0.220	0.891	0.571
0.228	0.870	0.571
0.236	0.870	0.592
0.239	0.870	0.612
0.242	0.870	0.633
0.251	0.870	0.653
0.261	0.870	0.674
0.265	0.870	0.694
0.290	0.848	0.694
0.326	0.848	0.714
0.350	0.826	0.714
0.366	0.804	0.714
0.383	0.804	0.735
0.397	0.804	0.755
0.401	0.804	0.776
0.410	0.783	0.776
0.452	0.783	0.796
0.500	0.761	0.796
0.528	0.761	0.837
0.546	0.761	0.857
0.552	0.739	0.857

(*Continued*)

Table 5 Continued

Cutoff score (probability)	Sensitivity	Specificity
0.563	0.717	0.857
0.574	0.717	0.878
0.588	0.696	0.878
0.603	0.696	0.898
0.619	0.696	0.918
0.630	0.674	0.918
0.649	0.674	0.959
0.672	0.652	0.959
0.679	0.630	0.959
0.681	0.609	0.959
0.686	0.587	0.959
0.714	0.587	0.980
0.774	0.565	0.980
0.810	0.544	0.980
0.813	0.544	1.000

for the correlation or redundancy between SRT and SSPT, which is significant, $r = .70$ in this sample.

Clinicians may be unfamiliar with the use of logistic regression functions in obtaining cutoff scores. However, the calculations are quite straightforward. The raw scores from each predictor variable are entered into the linear function and the linear function is exponentiated in order to obtain the probability of group membership:

$$\text{Probability of response bias} = \frac{e^{[-3.321+.117(\text{SSPT Raw Error Score})+.245(\text{SRT Raw Error Score})]}}{1+e^{[-3.321+.117(\text{SSPT Raw Error Score})+.245(\text{SRT Raw Error Score})]}}$$

In this equation, HI participants were coded as "0" and PM participants as "1"; typically, cases whose probabilities exceed .50 are classified in the "1" group. However, a cutoff of greater than or equal to .45 produced optimal classification with a sensitivity of 78.3% and specificity of 79.6%, for an overall rate of 79.0% in the particular sample. Table 5 contains the diagnostic efficiency statistics for a range of probabilities associated with the logistic regression function.

It should be emphasized that a test score in isolation is of limited value unless it is combined with an estimate of the base rate or prevalence of the disorder of interest. First, the diagnostic test can be characterized in terms of a single number, known as the likelihood ratio (LR): sensitivity/(1 – specificity) (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000). The LR indicates how much more likely a positive test is to be found in a person with, as opposed to without, the disorder (Greenhalgh, 1997). The LR is then multiplied by the pre-test odds (base rate) to obtain the post-test odds, i.e., the probability that the person has the disorder given a positive test result.

Table 6 Positive predictive values for base rates using cutoff score guidelines

Test/base rates	.10	.25	.40
SRT	.24	.49	.65
SSPT	.25	.50	.67
LR function	.30	.56	.72

For illustration, let's assume that the base rate probability of malingering is 0.25 in a particular setting, which yields a pre-test odds of $0.25/(1 - 0.25) = 0.33$. Choosing a cutoff of 8 on SRT, produces a LR of $(.751)/(1 - .735) = 2.87$. Thus, the post-test odds would be $(0.33)(2.87) = .948$ in favor of a diagnosis of response bias. Converting odds to a probability, $.948/(1 + .948)$, there would be a 49% probability in support of a diagnosis of response bias. Table 6 contains the post-test probabilities (also known as the positive predictive values) for selected base rates for each test indicator, using the derived cutoff scores. However, clinicians need to consider the "cost" of making false positive or false negative errors in each situation, make use of estimated base rates, and adjust cutoff scores accordingly.

DISCUSSION

The results of this study further support the use of forced-choice tests such as the SRT and SSPT in the detection of malingering in litigating MHI. In a sample of probable malingerers and patients with suspected head injury, a cutoff score of 10 on the SSPT correctly classified 73% of participants. These findings are consistent with those of Trueblood and Schmidt (1993) in the identification of malingered head injury in which the same cutoff score was used. In addition, a cutoff score of 8 on the SRT correctly classified 74% of participants. When cutoff scores suggested by previous studies for the SRT were applied to the current sample, diagnostic efficiency decreased. Although participants performing at chance levels on the RMT also performed poorly on other neuropsychological tests, probable malingerers obtained scores on the SRT and SSPT that were better than chance, but still much lower than patients with MHI from representative samples (Dikmen, Machamer, Temkin, & Winn, 1995), and even lower than patients with moderate to severe head injury (Millis, 2002).

Logistic regression analyses further supported the joint use of the SRT and SSPT, where each contributed unique variance to the prediction of probable malingering. Although no test should be used in isolation to determine malingering (Putnam, Millis, & Adams, 1996), performance on the SRT and SSPT appears to be a red flag suggesting further assessment. Estimates of the base-rate of malingering vary widely, depending on the setting. However, if the base rate is around 50% as some reports have suggested (Mittenberg et al., 2002; Schmand et al., 1998), the positive predictive value of the logistic regression model based on both tests is .72, indicating 7 true positive detections for every 10 patients with a positive test sign. Even when the base rate approaches a more conservative level (25%; Binder, 1993), the SRT/SSPT logistic model has a positive predictive value (.56) that well exceeds chance and is commensurate with even published measures that specifically assess response invalidity (e.g., the Validity Indicator Profile; Ross & Adams, 1999). In light of previous studies demonstrating the utility of forced-choice neuropsychological tests in the detection of response bias, these findings clearly support the SRT and SSPT as partial indicators of probable malingering. Nonetheless, a positive test sign on the SRT and SSPT is only one index and needs to be considered along with other tests, contextual factors, extrinsic motivation, etc., in order to fully determine the presence or absence of malingering (Millis & Volinsky, 2001; Rohling et al., 2003).

Unlike most previous studies, multiple test-based measures were used to further increase our confidence in participant selection. Not only did participants in the probable malingering group perform within chance levels on the RMT, but demonstrated a similar pattern of responding on other measures (e.g., the MMPI-2 and WAIS-R), consistent with our characterization of group participants as probable malingerers. Additionally, these participants also performed significantly lower on WAIS-R FSIQ compared to controls, and obtained WAIS-R FSIQ scores significantly lower than premorbid FSIQ estimates. In contrast, groups did not differ on demographic- (i.e., Barona) or test-based (e.g., NAART) indicators of premorbid intellectual ability. These findings bode well for the internal validity of the study, and suggest we were largely successful in identifying patients aptly characterized as probable malingerers.

Although the results support the use of these tests in the identification of probable malingering, some persons who malinger may not perform so poorly that they are detected using an "insufficient effort" paradigm. Those who are more sophisticated and savvy in their approach to feigning head injury may have escaped detection using our method for inclusion. This weakness, inherent to our approach, may limit generalizability. For example, Rapport et al. (1998) found that those who are more intelligent tend to be more effective in simulating the effects of head injury on neuropsychological tests. Consequently, our paradigm represents a relatively conservative method for identifying persons engaging in highly suspect responding indicative of malingering.

REFERENCES

Axelrod, B. N. & Rawlings, D. B. (1999). Clinical utility of incomplete effort WAIS-R formulas: A longitudinal examination of individuals with traumatic brain injury. *Journal of Forensic Neuropsychology*, *1*, 15–27.

Baker, R., Donders, J., & Thompson, E. (2000) Assessment of incomplete effort with the California Verbal Learning Test. *Applied Neuropsychology*, *7*, 111–114.

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

Beetar, J. T. & Williams, J. M. (1995). Malingering response styles on the Memory Assessment Scales and symptom validity tests. *Archives of Clinical Neuropsychology*, *10*, 57–72.

Bernard, L. C., McGrath, M. J., & Houston, W. (1993). Discriminating between simulated malingering and closed head injury on the Wechsler Memory Scale—Revised. *Archives of Clinical Neuropsychology*, *8*, 539–551.

Bigler, E. D. (1990). Neuropathology of traumatic brain injury. In E. D. Bigler (Ed.), *Traumatic brain injury*. Austin, TX: Pro-ed.

Bigler, E. D., Johnson, S. C., Anderson, C. V., Blatter, D. D., Gale, S. D., Russo, A. A., et al. (1996). Traumatic brain injury and memory: The role of hippocampal atrophy. *Neuropsychology*, *10*, 333–342.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology*, *15*, 170–183.
Binder, L. M. & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment*, *3*, 175–181.
Blair, J. R. & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, *3*, 129–136.
Bornstein, R. A. (1982). Reliability of the speech sounds perception test. *Perceptual and Motor Skills*, *55*, 203–210.
Bornstein, R. A. (1983). Reliability and item analysis of the Seashore Rhythm Test. *Perceptual and Motor Skills*, *57*, 571–574.
Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota multiphasic personality inventory (MMPI-2)*. Minneapolis: University of Minnesota Press.
Charter, R. A. (1994). Determining random responding for the Category, Speech-Sounds Perception, and Seashore Rhythm Test. *Journal of Clinical and Experimental Neuropsychology*, *16*, 744–748.
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.
Forrest, T. J., Allen, D. N., & Goldstein, G. (2004). Malingering indexes for the Halstead Category Test. *The Clinical Neuropsychologist*, *18*, 334–347.
Frederick, R. I. (1997). *The validity indicator profile*. Minneapolis, MN: NCS.
Gallop, R. J., Crits-Christoph, P., Muenz, L. R., & Tu, X. M. (2003). Determination and interpretation of the optimal operating point for ROC curves derived through generalized linear models. *Understanding Statistics*, *2*, 219–242.
Gfeller, J. D. & Cradock, M. M. (1998). Detecting feigned neuropsychological impairment with the Seashore Rhythm Test. *Journal of Clinical Psychology*, *54*, 431–438.
Greenhalgh, T. (1997). How to read a paper: Papers that report diagnostic or screening tests. *British Medical Journal*, *315*, 540–543.
Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and postconcussion syndrome. *Journal of the International Neuropsychological Society*, *2*, 477–485.
Greve, K. W., Bianchini, K. J., Mathias, C. W., Houston, R. J., & Crouch, J. A. (2003). Detecting malingered performance on the Wechsler Adult Intelligence Scale: Validation of Mittenberg's approach in traumatic brain injury. *Archives of Clinical Neuropsychology*, *18*, 245–260.
Halstead, W. (1947). *Brain and intelligence. A quantitative study of the frontal lobes*. Chicago: University of Chicago Press.
Hardin, J. & Hilbe, J. (2001). *Generalized linear models and extensions*. College Station, TX: Stata Press.
Heaton, R. K., Smith, H. H., Lehman, R. A.W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*, 892–900.
Hiscock, C. K., Branham, J. D., & Hiscock, M. (1994). Detection of feigned cognitive impairment: The two-alternative forced-choice method pared with selected conventional tests. *Journal of Psychopathology & Behavioral Assessment*, *16*, 95–110.
Iverson, G. L. & Franzen, M. D. (1994). The Recognition Memory Test, Digit Span, and Knox Cube Test as markers of malingered memory impairment. *Assessment*, *1*, 323–334.
Iverson, G. L., Slick, D. J., & Franzen, M. D. (2000). Evaluation of a WMS-R Malingering Index in a non-litigating clinical sample. *Journal of Clinical and Experimental Neuropsychology*, *22*, 191–197.
Kneebom, A. C., Chelune, G. J., & Luders, H. O. (1997). Individual patient prediction of seizure lateralization in temporal lobe epilepsy: A comparison between neuropsychological

memory measures and the intracarotid amobarbitol procedure. *Journal of the International Neuropsychological Society*, *3*, 159–168.

Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *The Clinical Neuropsychologist*, *12*, 179–188.

Larrabee, G. J. (2003). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist*, *17*, 54–68.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, *68*, 203–210.

Menard, S. (2002). *Applied logistic regression* (2nd ed.). Thousand Oaks, CA: Sage.

Miller, L. J. & Donders, J. (2001). Subjective symptomatology after traumatic head injury. *Brain Injury*, *15*, 297–304.

Millis, S. R. (1992). Recognition Memory Test in the detection of malingered and exaggerated memory deficits. *The Clinical Neuropsychologist*, *7*, 406–414.

Millis, S. R. (1994). Assessment of motivation and memory with the Recognition Memory Test after financially compensable mild head injury. *Journal of Clinical Psychology*, *50*, 601–605.

Millis, S. R. (2002). Warrington's Recognition Memory Test in the detection of response bias. *Journal of Forensic Neuropsychology*, *2*, 147–166.

Millis, S. R. & Djikers, M. (1993). Use of the Recognition Memory Test in traumatic brain injury. *Brain Injury*, *7*, 53–58.

Millis, S. R. & Putnam, S. H. (1994). The Recognition Memory Test in the assessment of memory impairment after financially compensable mild head injury: A replication. *Perceptual and Motor Skills*, *79*, 384–386.

Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment*, *7*, 463–471.

Millis, S. R., Ross, S. R., & Ricker, J. H. (1998). Detection of incomplete effort on the Wechsler Adult Intelligence Scale—Revised: A cross-validation. *Journal of Clinical and Experimental Neuropsychology*, *20*, 167–173.

Millis, S. R. & Volinsky, C. T. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology*, *23*, 809–828.

Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale—Revised. *Psychological Assessment*, *5*, 34–40.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base-rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094–1102.

Mittenberg, W., Rotholc, A., Russell, E., & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead-Reitan Battery. *Archives of Clinical Neuropsychology*, *11*, 271–281.

Mittenberg, W., Theroux-Fichera, S., Zielinski, R., & Heilbronner, R. L. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale—Revised. *Professional Psychology: Research and Practice*, *26*, 491–498.

Mutchnick, M. G., Ross, L. K., & Long, C. J. (1991). Decision strategies for cerebral dysfunction IV: Determination of cerebral dysfunction. *Archives of Clinical Neuropsychology*, *6*, 259–270.

Putnam, S. H., Millis, S. R., & Adams, K. M. (1996). Mild traumatic brain injury: Beyond cognitive assessment. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (2nd ed., pp. 529–551). New York: Oxford University Press.

Raftery, A. E. (1996). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Social methodology* (Vol. 25, pp. 111–163). Oxford: Basil Blackwell.

Rapport, L. J., Farchione, T. J., Coleman, R. D., & Axelrod, B. N. (1998). Effects of coaching on malingered motor function profiles. *Journal of Clinical & Experimental Neuropsychology*, *20*, 89–97.

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, *10*, 10–20.

Reitan, R. M. & Wolfson, D. (1990). The significance of the Speech-Sounds Perception Test for cerebral functions. *Archives of Clinical Neuropsychology*, *5*, 70–78.

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

Rogers, R., Gillis, J. R., & Bagby, R. M. (1990). The SIRS as a measure of malingering: A validation study with a correctional sample. *Behavioral Sciences & the Law*, *8*, 85–92.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingers: A prototypical analysis of explanatory models. *Law and Human Behavior*, *22*, 353–367.

Rohling, M. L., Langhinrichsen-Rohling, J., & Miller, L. S. (2003). Actuarial assessment of malingering: Rohling's interpretive method. In R. D. Franklin (Ed.), *Prediction in Forensic and Neuropsychology: Sound statistical practices* (pp. 171–207). Mahwah, NJ: LEA.

Ross, S. R. & Adams, K. M. (1999). Test review of the Validity Indicator Profile (VIP): One more test of malingering? *The Clinical Neuropsychologist*, *13*, 112–116.

Ross, S. R., Krukowski, R. A., Putnam, S. H., & Adams, K. M. (2003). The Memory Assessment Scales in the detection of incomplete effort in mild head injury. *The Clinical Neuropsychologist*, *17*, 581–591.

Ross, S. R., Millis, S. R., Krukowksi, R. A., Putnam, S. H., & Adams, K. M. (2004). Detecting incomplete effort on the MMPI-2: An examination of the Fake-Bad Scale in mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *26*, 115–124.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine*. New York: Churchill Livingstone.

Schagen, S., Schmand, B., de Sterke, S., & Lindeboom, J. (1997). Amsterdam Short-Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical & Experimental Neuropsychology*, *19*, 43–51.

Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, H. L. (1998). Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery, and Psychiatry*, *64*, 339–343.

Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology*, *18*, 911–922.

Sweet, J. J., Demakis, G. J., Ricker, J. H., & Millis, S. R. (2000). Diagnostic efficiency and material specificity of the Warrington Recognition Memory Test: A collaborative multisided investigation. *Archives of Clinical Neuropsychology*, *15*, 301–309.

Sweet, J. J., Wolfe, P., Sattlberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S., et al. (2000). Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology*, *15*, 105–113.

Tenhula, W. N. & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist*, *10*, 104–116.

Trueblood, W. & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *15*, 578–590.

Tsushima, W. T. & Tsushima, V. G. (2001). Comparison of the Fake Bad Scale and other MMPI-2 validity scales with personal injury litigants. *Assessment*, *8*, 205–212.

Warrington, E. K. (1984). *Recognition Memory Test: Manual*. Berkshire, United Kingdom: NFER-Nelson.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.