Pergamon

Archives of Clinical Neuropsychology
18 (2003) 261–276

Archives of CLINICAL NEUROPSYCHOLOGY

# A validation of multiple malingering detection methods in a large clinical sample

John E. Meyers [a,*], Marie E. Volbrecht [b]

[a] *Mercy Rehabilitation Clinic, 500 Jackson Street, Ste 360, Sioux City, IA 51101, USA*
[b] *Sioux Valley Hospital, Sioux Falls, SD, USA*

Accepted 19 November 2001

**Abstract**

The purpose of this study is to further previous research that has shown that common neuropsychological tests can do "double duty" as test of motivation/malingering. Using a large clinical sample of 796 participants, it was found that the nine neuropsychological tests (when used together) were able to correctly identify litigant and nonlitigating groups. Failure on any two of the malingering tests suggested motivational/malingering issues. The groups consisted of mild, moderate, and severe traumatic brain-injured patients; chronic pain, depressed, community controls, and "malingering actors." Institutionalized and noninstitutionalized patient performance were also examined. This method showed 83% sensitivity and 100% specificity. A 0% false positive rate was found, suggesting good reliability especially in litigating settings. A group of patients for whom this method of motivational assessment might not be appropriate was also identified.


*Keywords:* Neuropsychology; Malingering; Token Test; Finger Tapping; Forced Choice Test; Rey Complex Figure; Dichotic Listening; Reliable Digits; Sentence Repetition; Rey Auditory Verbal Learning Test; Judgment of Line Orientation

## 1. Introduction

It has long been the authors' opinion that tests of malingering are (unfortunately) a necessary part of a neuropsychological assessment, and that the validity of the neuropsychological tests used in the profile need to be checked for validity. It has been well reported by other

* Corresponding author.
*E-mail address*: jmeyers@mstar2.net (J.E. Meyers).

0887-6177/02/$ – see front matter © 2002 National Academy of Neuropsychology.
PII: S0887-6177(02)00136-1

authors (Goebel, 1983; Greiffenstein, Gola, & Baker, 1995; Heaton, Smith, Lehman, & Vogt, 1978; Iverson & Binder, 2000; Meyers & Diep, 2000; Meyers & Volbrecht, 1998a; Oberg, Udessen, Thomsen, Gade, & Mortensen, 1985) that not all malingerers perform identically on neuropsychological tests. The assessment of malingering has been approached in many ways.

Slick, Sherman, and Iverson (1999) claimed that the pattern of performance method (PPM) is probably the most effective way to detect malingering with standard neuropsychological evaluations. PPM involves comparing several test scores with performance on a single test to assess validity. At least four procedures for detecting malingering have been found that can be considered PPMs. The first one was described by Slick et al. and involves inspecting performance on "floor" items for uncommon mistakes, such as forgetting one's own name. Similarly, scores for easy items are compared to scores on more difficult items, or performance curves across varying levels of difficulty are examined. A second variation of this method involves examination of scores within or across tests for consistency with established patterns of function or impairment within a certain area. Examples of this include unusual patterns of serial position effects in list learning and other memory tests (Bernard, 1991; Russell, Spector, & Kelly, 1993), comparison of recall to recognition (Beetar & Williams, 1995; Bernard, 1991; Binder, 1992), and comparison of tasks dealing with attention and memory indices (Mittenberg, Arzin, Millsaps, & Heilbronner, 1993). A third PPM is after-the-fact statistical evaluation of scores obtained and established contrast groups such as actor malingerers, probable malingerers, and nonlitigating patients. A fourth variation of PPM that is more recent is the evaluation of magnitude of errors, that is, errors that are more than would be expected given the reported injury.

Specific assessment tasks to detect malingering have been developed, such as the Forced Choice Test (FC; Hiscock & Hiscock, 1989) and its later variant, the Portland Digit Recognition Test (Binder, 1993). Although these specific malingering tools may be useful, they may also be inadequate for several reasons. First, these tests typically are employed solely for the detection of malingering. They are not useful for other neuropsychological interpretive purposes. If tests currently in common use for the purpose of neuropsychological assessment could do double duty and also detect malingering, this would be a more efficient procedure. In the current climate of managed health care and accountability, it is clear why this efficient method to conserve valuable resources could be advantageous. Second, some authors have reported that even in a forensic context where secondary gain for symptom production or exaggeration is obvious, specific tasks for malingering are not commonly utilized (Lees-Haley, Smith, Williams, & Dunn, 1996) despite evidence that without such instruments, "clinicians are often oblivious to malingering" (Binder & Rohling, 1996, p. 10). Third, individuals who malinger do not necessarily do so in a consistent manner, but rather, attempt to malinger different types of impairment (Goebel, 1983; Greiffenstein et al., 1995; Heaton et al., 1978; Meyers & Volbrecht, 1998a; Oberg et al., 1985). For example, while one individual may fake or exaggerate a visual perceptual problem, another may malinger reduced motor speed. These differential malingered impairments, just as actual impairments, would thus be better detected by techniques specific to the nature of the alleged dysfunction rather than one global measure of malingering. Fourth, recent findings indicate that individuals who are cognizant of the possibility of the evaluation for malingering during neuropsychological testing recognize the forced-choice format as an attempt to do so (Suhr & Gunstad, 2000). Finally, as Bernard, Houston, and Natoli (1993)

pointed out, this method does not allow for the clinician to examine the vulnerability to malingering of the other tests utilized in the interpretation. Therefore, developing validity checks in already existing and widely utilized neuropsychological instruments may not only be more efficient, but also potentially more valid (Meyers, Galinsky, & Volbrecht, 1999).

The current authors have undertaken a several-year project to identify different methods of detecting malingering using a battery of commonly used neuropsychological tests. This project has focused on several widely utilized tests. The first was presented in a series of studies using the Rey Complex Figure and Recognition Trial (RCFT; Meyers & Meyers, 1995). Initially memory error patterns (MEPs) were examined and five basic MEPs were identified: attention, encoding, storage and retrieval, and normal/other. Several studies were undertaken to identify the ecological validity and to examine expected real-world performance of persons that obtain different MEPs (Meyers, Bayless, & Meyers, 1996; Meyers & Volbrecht, 1998a; Meyers & Volbrecht, 1999).

The ecological data showed that attention, encoding, and storage MEPs were found in patients with Rancho Level 7 and below and the more impaired MEPs (i.e., attention) were found at Rancho Level 4, which occur in very impaired persons. Those patients with Rancho Level 8 and above obtained retrieval or normal/other MEPs.

With regard to malingering detection with the RCFT, it was found that persons attempting to malinger performance on the RCFT produced attention, encoding, and storage MEPs. However, these MEPs are only found in persons who are *extremely* impaired and functioning at very low levels with a need to be supervised. For instance, no one with one of these three MEPs has been able to drive and find their way to an appointment for a neuropsychological assessment. It was concluded that a person that produced one of these three MEPs is either malingering or under 24-h care.

The second test investigated for malingering detection was Reliable Digits (RD; Meyers & Volbrecht, 1998b). RD was first presented by Greiffenstein, Baker, and Gola (1994) and is defined as the longest number of digits repeated correctly on both trials of Digit Span forward plus the longest number of digits correctly repeated on both trials backwards. In their initial presentation of RD, Greiffenstein et al. recommended a cutoff score of 7 below which malingering and/or submaximal performance was indicated. In the validation of RD presented by Meyers and Volbrecht (1998b), they found that a cutoff score of 7 did adequately discriminate nonmalingerers (passed FC) and malingerers (failed FC). However, there was a 4% (2/49) false positive rate when using this cutoff. Although using a cutoff of 7 is statistically adequate, for the purposes of the current study a cutoff score of 6 on RD will be used to help eliminate any false positives. This is in keeping with the authors' conservative approach to the detection of malingering.

A third study examined a couple of common neuropsychological tests as markers for malingering (Meyers et al., 1999). They examined a combination of FC (Brandt, Rubinsky, & Lassen, 1985), Judgment of Line Orientation (JLO; Benton, Hamsher, Varney, & Spreen, 1983), Token Test (TT; Spreen & Strauss, 1998), and Dichotic Listening (DL; Roberts et al., 1994). In this study, several groups of participants were utilized. Group 1 consisted of participants with 1–7 days loss of consciousness (LOC) documented in their medical records. These individuals were followed in the hospital through their acute and postacute rehabilitation (inpatient and outpatient). All had identifiable injury on CT/MRI and were seen in the context of rehabilitation.

Group 2 participants consisted of individuals who had been seen as part of physician-referred neuropsychological evaluation following mild traumatic brain injury (TBI). Each participant's injury was due to motor vehicle accident, fall, or blow to the head. For those individuals who had a LOC, this was documented in their medical records by witnesses at the scene or by ambulance attendants. The longest LOC was 40 min. None of these individuals was in litigation at the time of the assessment. The cutoff for each test was established by inspection and defined as being one less than the lowest score obtained by the participants in Group 1 (1–7 days LOC). None of the mild brain-injured persons was expected to score at or below this cutoff score. Using this method it was recommended that failure on any of these measures (FC, JLO, TT, DL) may suggest exaggeration or inconsistent performance, but failure on two or more would suggest malingering unless the person had more than 7 days of LOC or was institutionalized. Using this method a 0% false positive rate was found (100% specificity) with 95% sensitivity for failure on at least one method and 60% sensitivity for failure on two or more tests. Each of the four tests and their respective cutoffs are discussed elsewhere in this paper.

Brandt et al. (1985) presented a two-item FC recognition memory test for detection of malingering memory impairment. This malingering test was also discussed in Lezak (1995). They initially utilized groups of normal controls, malingering simulators, and patients with Huntington's disease. They found that scores ranging from 6 to 14 would be at chance levels. The simplicity of this malingering test is one of the attractive features of the protocol. Testing time is increased only minimally, it can be administered with only a test protocol, and it can be administered bedside if necessary. In the study, a score of 10 or below was defined as indicative of malingering (Meyers et al., 1999).

The JLO task consists of matching two lines representing an angle to a multiple-choice response card. There are two forms of this test (Form H and Form V) and the reliability of both of these forms were reported to be .89 and above (Benton et al., 1983). It was reported that participants who took Form H and Form V demonstrated a test–retest reliability coefficient of .90. The JLO was found to be sensitive to brain injuries affecting the ability to judge lines and angles (Benton et al., 1983). In the original study by Meyers et al. (1999) the cutoff for malingered performance was defined as 12 or below.

TT (Spreen & Strauss, 1998) appears to participants to be a simple task of following basic directions. Participants are asked to follow simple directions such as "Move the green square away from the yellow square." The participants' accuracy in performing these tasks is assessed. Tasks range from very simple (e.g., "Show me a circle.") to more difficult (e.g., "Together with the yellow circle pick up the blue circle."). However, for this task the standard administration of easy items progressing to the presentation of more difficult items is reversed. That is, the most difficult section (F) is administered first and, if the participant makes any errors, the next "easier" section is given. This protocol is continued until the participant is able to follow all tasks in a given section or until Section A is completed. In this way, the most difficult tasks are presented first with easier tasks being presented later. Malingerers may expect the simpler tasks to be presented first and tasks that are more difficult to be presented later as this is a standard testing configuration. Scoring is the same as indicated in Spreen and Strauss (1998). In the original study by Meyers et al. (1999) the cutoff was defined as 150 or below for malingering.

The DL test uses a simple audiotape player with stereo headphones and a cassette tape (Audiotec of St. Louis, 1991). The task involves identifying simultaneously presented words,

one in the right ear and one in the left ear. Impairments in Dichotic Listening have been found for patients with cerebral disease (Damasio & Damasio, 1979) for patients with demyelination disorders (Risse, Gates, Lund, Maxwell, & Rubens, 1989) and for patients with closed head brain injury (Levin et al., 1989). See Roberts et al. (1994) for a summary of neuropsychological deficits associated with dichotic impairment. A study by Meyers, Roberts, Bayless, Volkert, and Evitts (2002) increased the normative base for DL and also demonstrated the clinical utility of this task. The test protocol consists of a full 60-item version or the 60 items can be divided into two short form versions. From the short form, 30-item version, the score for both ears simultaneously correct was used. In the study by Roberts et al., it was found that the two versions correlated well suggesting that the two short forms of the test are equivalent. In the study by Meyers et al. (1999) the cutoff for malingering was defined as 9 or below and the short form was utilized.

The same process and data used by Meyers et al. (1999) was used by Meyers, Morrison, and Miller (2001) to identify two other tests with potential as malingering indicators. These two additional tests were Sentence Repetition (SR; Meyers, Volkert, & Diep, 2000; Spreen & Strauss, 1998) and Rey Auditory Verbal Learning Test—Recognition (AVLT-R; Spreen & Strauss, 1998).

SR is a tool for the assessment of auditory span in which sentences of increasing length are presented verbally. Sentence length increases from one syllable in the first sentence to 26 syllables in the final sentence (Spreen & Strauss, 1998). Recently, new norms for SR have been developed, increasing its clinical utility and affirming its validity for continued contemporary usage (Meyers et al., 2000). The new norms indicated that SR performance was not correlated with age, gender, or handedness, but was positively correlated with education.

The AVLT-R is an easily administered word list learning task. It provides measures of immediate memory, provides a learning curve, reveals learning strategies, as well as retroactive and proactive interference and confusion or confabulation on memory tasks. It measures both short-term and long-term retention following a time interval in which another activity takes place, and allows the examiner to compare retrieval efficiency with learning (Lezak, 1995). The AVLT consists of five verbal presentations of a 15-item word list, one presentation of a second 15-word list, and one final recall trial of the initial list. The recognition component of the AVLT-R consists of a list of 50 words in which the original 15 words are contained.

Another method of malingering assessment involves inspecting and comparing the performance on neuropsychological tests that use common functions (i.e., motor). Copying the RCFT (raw score), Digit Symbol (scale score) and Block Design (scale score) were used as variables to predict the expected performance on Finger Tapping—dominant hand (FT; Reitan & Wolfson, 1985; Shimoyama, Ninchoji, & Uemura, 1990). Using the scores from a large database of 650 varied neuropsychological patients, a linear regression was calculated. The resulting formula was (RCFT raw score $\times$ .185) + (Digit Symbol Scale score $\times$ .491) + (Block Design Scale score $\times$ .361) + 31.34. This formula was used to calculate the estimated FT score (EFT), which was then subtracted from the actual FT score and the difference (FT − EFT) was used as a band of error (database maintained by first author). It was found that persons with mild TBI or chronic pain were not expected to score less than a −10 difference in expected performance on FT.

Overall, nine individual methods of identifying malingered performance have been developed. These methods incorporated commonly used neuropsychological tests that are often

Table 1
Cutoff scores for each test/method of detection of malingering/submaximal effort

| Test/method | Cutoff |
|---|---|
| MEP | ≤3 (1 = attention, 2 = encoding, 3 = storage, 4 = retrieval/other) |
| RD | ≤6 |
| FC | ≤10 |
| JOL | ≤12 |
| TT | ≤150 |
| DL | ≤9 |
| SR | ≤9 |
| AVLT-R | ≤9 |
| EFT | ≤−10 |

part of a detailed neuropsychological evaluation. Therefore, incorporation of these methods of identifying malingered/submaximal performance does not add time or additional tests to a neuropsychological assessment that would normally use these tests. Table 1 provides a summary of the nine tests and their respective cutoff scores. The purpose of the current study was to investigate the use and integration of these nine methods of malingering assessment in a large clinical sample in order to identify how these methods may be integrated into clinical practice.

## 2. Methods

### *2.1. Participants*

A total of 796 people participated in the study; ages ranged from 16 to 86, with education ranging from 5 to 23 years. Tables 2 and 3 present means and standard deviations for demographic descriptors and the diagnoses as indicated in the participants medical records. An individual may have had more than one diagnoses (i.e., head injury and depression). Only the neurological and mental health diagnoses were used in this study, some participants had other health problems (i.e., hemorrhoids, diabetes) that were not listed. From this total pool of participants, subgroups were defined.

Groups 1 through 9 were defined based on diagnoses, as contained in medical records, so that a comparison of diagnostic groups and failure rates could be made. Group 1 ($n = 56$) consisted of patients seen for neuropsychological examination for TBI. None of these participants were involved in litigation at the time of the assessment. If at anytime prior to January 1, 2001 a request for records was received from an attorney or disability then the participant's records were not included in this group. All had LOC of less than an hour documented in their medical records by ambulance worker, witnesses, or medical staff. All were seen for neuropsychological assessment in the context of rehabilitation.

Group 2 participants ($n = 10$) were also seen for neuropsychological assessment for TBI. None were involved in litigation at the time of the assessment. If at any time a request for records was received from an attorney or disability then the participant's records were not

Table 2
Descriptive statistics of demographics for study participants

| Group | *n* | Age, mean (S.D.) | Ed, mean (S.D.) | Gender (*n*) | | Handed (*n*) | | FSIQ, mean (S.D.) | LOC, mean (S.D.) | Months since injury |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Female | Male | Right | Left | | | |
| 1 | 56 | 32.11 (14.0) | 13.09 (2.4) | 16 | 40 | 52 | 4 | 97.82 (11.3) | <1 h | 12.23 (20.7) |
| 2 | 10 | 29.00 (17.2) | 12.80 (1.8) | 2 | 8 | 8 | 2 | 98.70 (15.2) | .36 (.25) | 9.40 (15.7) |
| 3 | 29 | 30.03 (11.5) | 12.62 (2.0) | 7 | 22 | 27 | 2 | 93.45 (13.8) | 4.65 (2.6) | 40.85 (82.5) |
| 4 | 11 | 32.73 (14.7) | 11.73 (2.6) | 6 | 5 | 10 | 1 | 89.45 (9.2) | 24.72 (23.4) | 78.00 (88.1) |
| 5 | 38 | 37.95 (14.3) | 13.08 (1.8) | 27 | 11 | 35 | 3 | 99.18 (12.3) | | 21.14 (19.0) |
| 6 | 25 | 45.28 (12.4) | 13.56 (2.6) | 12 | 13 | 25 | 0 | 104.08 (10.5) | | |
| 7 | 84 | 38.61 (12.1) | 12.71 (2.4) | 32 | 52 | 76 | 8 | 90.19 (11.7) | <1 h | 21.94 (41.6) |
| 8 | 19 | 37.00 (15.7) | 12.05 (2.5) | 9 | 10 | 16 | 3 | 88.47 (14.1) | 9.36 (1.5) | 62.84 (115.3) |
| 9 | 64 | 41.42 (8.3) | 11.77 (2.2) | 30 | 34 | 56 | 8 | 89.67 (15.3) | | 25.00 (24.8) |
| 10 | 32 | 32.97 (17.7) | 13.72 (3.1) | 15 | 17 | 32 | 0 | 111.31 (12.5) | | |
| 11 | 160 | 40.07 (19.4) | 11.77 (2.3) | 74 | 86 | 137 | 23 | 76.39 (11.9) | 10.28 (17.3) | 24.00 (46.0) |
| 12 | 211 | 39.47 (17.4) | 12.62 (2.8) | 99 | 112 | 179 | 32 | 92.88 (12.3) | .36 (2.8) | 35.15 (66.8) |
| 13 | 19 | 45.47 (16.1) | 12.16 (2.7) | 15 | 4 | 17 | 2 | 79.58 (11.8) | | 16.71 (16.4) |
| 14 | 17 | 35.29 (14.5) | 11.59 (2.2) | 9 | 8 | 16 | 1 | 76.53 (11.3) | | 7.00 (5.2) |
| 15 | 21 | 30.29 (6.7) | 18.00 (1.3) | 15 | 6 | 20 | 1 | 76.52 (15.0) | | |
| Total | 796 | 38.17 (16.1) | 12.61 (2.6) | 368 | 428 | 706 | 90 | 89.61 (15.4) | 4.17 (11.1) | 27.53 (55.6) |

*Note*. Here 1 = traumatic brain injury (TBI) with loss of consciousness (LOC) of less than 1 h, not in litigation; 2 = TBI with LOC greater than 1 h and less than 24 h, not in litigation; 3 = TBI with LOC greater than or equal to 1 day and less than or equal to 8 days, not in litigation; 4 = TBI with LOC 9 days or greater, not in litigation; 5 = chronic pain patients not in litigation; 6 = depressed patients; 7 = TBI with LOC less than an hour and in litigation; 8 = TBI with LOC greater than an hour, in litigation; 9 = chronic pain patients in litigation; 10 = normal controls; 11 = institutionalized patients; 12 = noninstitutionalized and failed no more than one validity check; 13 = noninstitutionalized and failed two or more validity checks and not in litigation; 14 = noninstitutionalized and failed two or more validity checks and was involved in litigation; and 15 = informed actors (portraying role of a malingerer).

Table 3
Diagnostic classifications for Groups 11 through 14

| | Groups (*n*'s) | | | | |
|---|---|---|---|---|---|
| Diagnosis | 11 | 12 | 13 | 14 | Sample total |
| MVA/TBI | 36 | 0 | 0 | 0 | 253 |
| Bilateral CVA | 3 | 11 | 2 | 0 | 19 |
| Brain tumor | 4 | 7 | 0 | 0 | 11 |
| Anoxia | 9 | 4 | 0 | 0 | 13 |
| Blow to head/fall | 8 | 0 | 0 | 0 | 68 |
| Gun shot to head | 1 | 1 | 0 | 0 | 2 |
| Encephalitis | 4 | 3 | 1 | 0 | 9 |
| Other/multiple DX | 9 | 34 | 0 | 1 | 51 |
| Multiple sclerosis | 8 | 3 | 2 | 0 | 13 |
| Primary generalized epilepsy | 3 | 2 | 1 | 0 | 7 |
| Electrical | 1 | 0 | 0 | 0 | 3 |
| Hydrocephalus | 3 | 3 | 0 | 0 | 6 |
| Alzheimer's | 5 | 6 | 2 | 0 | 13 |
| Subcortical dementia | 7 | 5 | 0 | 0 | 12 |
| Multi-infarct | 3 | 2 | 0 | 0 | 5 |
| Parkinson's | 1 | 3 | 0 | 0 | 4 |
| Mental retardation | 15 | 0 | 4 | 1 | 20 |
| Developmental delay | 8 | 6 | 0 | 0 | 14 |
| Cerebral palsy | 0 | 3 | 0 | 0 | 3 |
| Learning disability | 0 | 18 | 0 | 0 | 20 |
| Neurologic normal | 0 | 2 | 0 | 0 | 64 |
| Left CVA | 10 | 11 | 4 | 1 | 26 |
| Right CVA | 5 | 3 | 0 | 0 | 8 |
| Brain stem CVA | 2 | 0 | 0 | 0 | 2 |
| Celebellar CVA | 1 | 0 | 0 | 0 | 1 |
| Carbon monoxide | 0 | 6 | 0 | 0 | 6 |
| ADD/ADHD | 0 | 9 | 0 | 2 | 12 |
| Amnestic disorder | 0 | 2 | 0 | 0 | 2 |
| Substance abuse | 4 | 9 | 0 | 2 | 15 |
| Mental health | 7 | 34 | 0 | 8 | 163 |
| Toxic chemicals | 0 | 0 | 0 | 1 | 1 |
| Huntington's | 0 | 1 | 0 | 0 | 1 |
| Epilepsy left | 0 | 8 | 0 | 0 | 9 |
| Epilepsy right | 0 | 4 | 0 | 0 | 4 |
| Nonverbal learning disability | 1 | 2 | 0 | 0 | 4 |
| Pick's disease | 2 | 0 | 2 | 1 | 5 |
| Lupus | 0 | 4 | 0 | 0 | 6 |
| Right temporal lobectomy | 0 | 3 | 1 | 0 | 4 |
| Fibromialgia | 0 | 1 | 0 | 0 | 4 |
| Right aneurysm | 1 | 0 | 0 | 0 | 1 |
| Total (*n*) | 160 | 211 | 19 | 17 | 796 |

*Note*. Here 11 = institutionalized patients; 12 = noninstitutionalized and failed one or less of the validity checks; 13 = noninstitutionalized and failed two or more validity checks and were not involved in litigation; and 14 = noninstitutionalized and failed two or more validity checks and were involved in litigation.

included in this group. All had LOC of more than an hour but less than a day. All were seen in the context of rehabilitation. Group 3 ($n = 29$) participants were similar to Groups 1 and 2, with the exception that all had LOC of 1–8 days. Group 4 ($n = 11$) participants were also similar to the other groups but had LOC of greater than or equal to 9 days.

Group 5 ($n = 38$) participants were patients seen in the context of chronic pain treatment, none were involved in litigation or disability proceedings at the time of the assessment. Those who had been involved in litigation had already had their cases settled. All were seen in the context of rehabilitation. Group 6 ($n = 25$) participants were individuals who were seen in a mental health context. All were diagnosed with depression sufficient to warrant hospitalization in either a psychiatric inpatient setting or a partial hospitalization setting and none were involved in litigation. If at anytime a request for records was received from an attorney or disability then the participant's records were not included in this group. Therefore, Groups 1 through 6 are all nonlitigating groups.

Group 7 ($n = 84$) participants had LOC of less than 1 h documented in their medical records and were seen in the context of rehabilitation or assessment for litigation/disability. Group 8 ($n = 19$) individuals were seen in a similar context as Group 7 but had LOC of greater than 1 h. Group 9 ($n = 64$) participants were patients seen in the context of chronic pain treatment and were involved in litigation or disability proceedings. All were seen in the context of rehabilitation or disability evaluation. Therefore, Groups 7 through 9 were all involved in litigation or disability proceedings. Demographic descriptors for these groups are in Tables 2 and 3.

Next, a group of normal controls (Group 10; $n = 32$) were selected from community volunteers. These persons were recruited by newspaper and radio spots asking for volunteers to participate in this study. Some community volunteers were also recruited by word of mouth. Table 2 also contains descriptive data for this group. None of these individuals had any history of developmental delay, learning disability, neurological condition, or mental health disorder. All were independent community-dwelling individuals.

Groups 11 through 14 consisted of individuals who did not fit in the diagnostic groups (Groups 1–9). These individuals represent participants from a general neuropsychological practice and were not previously classified as members of Groups 1 through 9. Tables 2 and 3 also provide the descriptive statistics for these groups.

Group 11 ($n = 160$) consisted of persons who were institutionalized at the time of the assessment. Institutionalized was defined as under professional nursing care 24 h a day, or in a nursing home environment where nursing care was available 24 h a day. Individuals in Group 12 ($n = 211$) were independent in functioning but failed no more than one of the validity checks. Group 13 ($n = 19$) consisted of participants who were independent and failed two or more of the validity indicators and were not involved in litigation. Group 14 ($n = 17$) consisted of independent participants who failed two or more validity checks and were involved in litigation or disability proceedings.

Group 15 ($n = 21$) participants were graduate students taking a neuropsychology course and individuals who worked in brain-injury treatment facilities. They had greater than average knowledge of neuropsychological testing and brain injury. They were asked to portray the role of an individual who was attempting to take the neuropsychological tests to "fake" having a brain injury. They were given the following paragraph of instructions.

> Assume the role of someone who has been involved in a motor vehicle accident. After the accident you experienced some headaches, mild confusion and reduced memory. This cleared up after a couple of weeks. However, you feel you deserve compensation from the insurance company. You have decided to "fake" having a brain injury by presenting symptoms so as to make it appear that you have a significant brain injury, and have hired an attorney. You have been referred for a neuropsychological assessment. You have been informed by your attorney that the neuropsychological assessment is an important piece in "proving your case." (Take the tests accordingly.)

### *2.2. Procedure*

All participants were administered the same battery of tests that contained the nine different methods of detection of malingering. This battery was administered by a neuropsychologist or master's level student or technician. All administrations followed standard protocols unless otherwise noted (i.e., TT). This battery of tests took approximately 2.5–3 h to administer.

For each of the 796 participants, a total failure score on the nine malingering tests summarized in Table 1 was calculated. The total number of failures for each participant was identified by summing the total number of failures across all nine tests using the cutoff scores in Table 1. Then each participant's total score was evaluated by diagnostic group (Groups 1–15).

### *2.3. Materials*

Participants all completed a similar battery of neuropsychological tests (Volbrecht, Meyers, & Kaster-Bundgaard, 2000) that included Wechsler Adult Intelligence Scale—Revised (WAIS-R) or Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997), short form (Pilgrim, Meyers, Bayless, & Whetstone, 1999; Spreen & Strauss, 1998; Ward, 1990); Trail Making Test (Reitan & Wolfson, 1985); JLO (Benton et al., 1983); SR (Meyers et al., 2000; Spreen & Strauss, 1998); Controlled Oral Word Association Test (FAS; Spreen & Strauss, 1998); Animal Naming (Spreen & Strauss, 1998); Boston Naming (Spreen & Strauss, 1998); DL (Meyers et al., 2002; Roberts et al., 1994); Auditory Verbal Learning Test (Spreen & Strauss, 1998); Rey Complex Figure Test and Recognition Trial (Meyers & Meyers, 1995); the Booklet Category Test (Victoria Revision; Kozel & Meyers, 1998); and the FC test (Brandt et al., 1985).

## 3. Results

An analysis was undertaken to identify the relationships between the total score and the demographic variables of the groups. A Pearson correlation was calculated using participants in Groups 1 through 6 to identify if a relationship existed between the total failure score and age ($r = .01$, $P = .88$) or education ($r = -.007$, $P = .92$); however, no significant relationship was found. Additionally, a $\chi^2$ test was performed comparing the total failure score with gender and handedness. Again, no significant relationships were found between these variables ($P$'s $> .05$).

Groups 1 and 7 were similar in make-up with the exception that the Group 7 participants were involved in litigation. An independent samples *t* test showed that education and months since injury were not significantly different between the groups, $t(1, 138) = .881$, $P = .38$ and $t(1, 133) = -1.61$, $P = .11$, respectively. However, differences were found for mean age, $t(1, 138) = -2.915$, $P = .004$, and the total number of failures on the malingering tests, $t(1, 138) = 3.854$, $P < .001$). However, as previously shown, age was not a significant factor in the total score. This suggests that although the Groups 1 and 7 have a difference in age, this difference does not account for the difference found in total score.

Groups 5 and 9 both involved individuals in treatment for chronic pain. Group 9 participants were involved in litigation, whereas Group 5 participants were not. They were also compared using a *t* test for independent samples. In this comparison, age, $t(1, 100) = -1.552$, $P = .124$ and months since injury, $t(1, 30) = -.378$, $P = .708$, were not significantly different. However, education, $t(1, 100) = 3.043$, $P = .003$, and the total failure score, $t(1, 100) = -3.848$, $P < .001$, were significantly different between the groups. As has already been explored, age, education, gender, and handedness do not appear to have a significant effect on the total failure score. Therefore, although the groups are different on some demographic variables, these differences do not account for the differences found in total failure score.

None of the nonlitigants or control participants failed more than one of the validity measures (Table 4). All participants in the nonlitigating groups (Groups 1–6) were correctly classified as not malingering. For the litigating groups (Groups 7–9) there were some individuals who fell within the malingering range. In Group 7, 20% of the participants were classified as

Table 4
Frequency count by groups for total failures on validity checks

| | Groups | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Total |
| 0 | 49 | 7 | 21 | 4 | 33 | 24 | 49 | 10 | 31 | 31 | 23 | 134 | 0 | 0 | 0 | 416 |
| 1 | 7 | 3 | 8 | 7 | 5 | 1 | 16 | 6 | 16 | 1 | 36 | 77 | 0 | 0 | 4 | 187 |
| 2 | | | | | | | 2 | 1 | 3 | | 35 | | 13 | 14 | 0 | 70 |
| 3 | | | | | | | 5 | 1 | 7 | | 25 | | 4 | 1 | 4 | 47 |
| 4 | | | | | | | 1 | 0 | 3 | | 23 | | 1 | 1 | 1 | 30 |
| 5 | | | | | | | 4 | 0 | 2 | | 10 | | 0 | 0 | 3 | 19 |
| 6 | | | | | | | 2 | 1 | 1 | | 6 | | 1 | 0 | 1 | 12 |
| 7 | | | | | | | 0 | | 0 | | 1 | | | 1 | 4 | 6 |
| 8 | | | | | | | 3 | | 1 | | 1 | | | | 2 | 7 |
| 9 | | | | | | | | | | | | | | | 2 | 2 |
| % Failed 2 or more | | | | | | | 20 | 15 | 26 | | 63 | | 100 | 100 | 83 | 24 |

*Note*. Here 1 = traumatic brain injury (TBI) with loss of consciousness (LOC) of less than 1 h, not in litigation; 2 = TBI with LOC greater than 1 h and less than 24 h, not in litigation; 3 = TBI with LOC greater than or equal to 1 day and less than or equal to 8 days, not in litigation; 4 = TBI with LOC 9 days or greater, not in litigation; 5 = chronic pain patients not in litigation; 6 = depressed patients; 7 = TBI with LOC less than an hour and in litigation; 8 = TBI with LOC greater than an hour, in litigation; 9 = chronic pain patients in litigation; 10 = normal controls; 11 = institutionalized patients; 12 = noninstitutionalized and failed no more than one validity check; 13 = noninstitutionalized and failed two or more validity checks and not in litigation; 14 = noninstitutionalized and failed two or more validity checks and was involved in litigation; and 15 = informed actors (portraying role of malingerer).

Table 5
Diagnoses of 19 noninstitutionalized, nonlitigating patients who failed two or more malingering tests

| Participant | Diagnoses/explanation |
|---|---|
| 1 | Large bilateral CVA |
| 2 | Wheelchair bound MS patient (unable to be reliably tested) |
| 3 | Advanced dementia, at home under 24-h spouse supervision |
| 4 | Left CVA, receptive language deficit |
| 5 | Severe mental retardation |
| 6 | Dementia, marked atrophy on CT |
| 7 | Mental retardation |
| 8 | Seizure four to five times a week, poor seizure control |
| 9 | Left CVA, receptive language impairment |
| 10 | Burst aneurysm, tested within 2 months of injury, prolonged rehabilitation, global language impairment |
| 11 | Cambodian, mental retardation (very poor English) |
| 12 | Alzheimer's disease, under 24-h spouse supervision |
| 13 | Left CVA, global aphasic |
| 14 | Advanced Alzheimer's disease |
| 15 | Encephalitis, under 24-h care at home |
| 16 | Homebound MS patient, 24-h home and family care |
| 17 | Moderate to severe mental retardation |
| 18 | Left CVA, severe receptive language deficit |
| 19 | Postright temporal lobectomy, continued seizures (several times a week), history of continued drug and alcohol abuse |

malingerers; Group 8 had 15% similarly classified. For the chronic pain group (Group 9) 26% were classified as malingering. None of the community volunteers (Group 10) were classified as malingering. For participants who were institutionalized (Group 11), 63% were classified in the malingering range.

An examination of the persons who were not institutionalized and not in litigation who failed the validity checks (i.e., a total score of 2 or more) was made. There were 19 participants who fit this category (Group 13). In examining the data from all the noninstitutionalized (Groups 12–14), it was found that there was a significant effect of litigating status on test performance. Of the total 247 participants, 66 were involved in litigation, 181 were not. Again finding of litigating status with the total validity performance was significant, $\chi^2(6) = 13.934$, $P = .03$. A $t$ test showed that when divided as litigants and nonlitigants there was a significant differences between groups, $t(1, 245) = 2.436$, $P < .015$. Table 5 provides a description of each participant, based on a review of their medical chart. The data from Group 14 indicate that 17 participants also were classified in the malingering range. Group 15 participants were persons asked to "malinger" on testing 83% of these persons were classified as malingering.

## 4. Discussion

The results of this study indicate that in a general neuropsychological population, the use of the nine selected neuropsychological tests in doing "double duty" not only as clinically

useful neuropsychological tests, but also as malingering tests, seems appropriate. In using this method, the validity of performance across the neuropsychological test battery can be assessed. The results show that some participants may fail one of the nine malingering tests, but failure on more than one is rare in noninstitutionalized or nonlitigating groups. Using failure of two or more malingering tests as a criterion, none of the control, depressed, or nonlitigating mild TBI or chronic pain patients were defined as malingering. Those who failed one may have had variable or inconsistent motivation, but it was not sufficient to invalidate performance on the battery of tests. It was found that in litigating groups (mild TBI and chronic pain) some litigants do fail two or more tests, raising question of malingering in those participants.

Even those with lengthy LOC do not appear to fail two or more of these validity checks. This is significant even with those who have lengthy LOC (i.e., 9 days or more of LOC). Those with much milder injuries (i.e., LOC $<1$ h) would also not be expected to fail these validity indicators unless motivational issues are present.

Litigating status, although significant in the number of individuals who are identified as malingering, should not be used as a sole criteria for malingering. That is, poor motivation/malingering can be found in individuals who are not involved in litigation as well as those that are involved in litigation. This speaks to the need to use tests of motivation in all neuropsychological settings.

In addition when participants were instructed to convincingly malinger, 83% failed two or more of the malingering test. However, all (100%) failed at least one malingering test, which at least raises the question of motivation on the testing. The data show an overall base rate of 15–26% malingering, which is consistent with other published data (Goebel, 1983; Greiffenstein et al., 1995; Heaton et al., 1978; Meyers & Volbrecht, 1998b; Oberg et al., 1985; Spreen & Strauss, 1998). In a forensic-based practice these rates may be higher.

Some persons who are institutionalized (under 24-h nursing care) may also fail two or more of the malingering tests. In these cases, the use of the malingering test may not be appropriate. Persons for whom this method may not be appropriate are those described in Table 5. They were not in litigation and were not institutionalized, but had obvious cognitive impairment and easily identifiable brain injury. From this examination of the records of the patients who failed, it can be concluded that there is a group of patients for whom this method of assessing validity is probably not appropriate. Those for whom this method may not be appropriate are:

1. Persons who are untestable neuropsychologically or under 24-h institutional care.
2. Patients with large (easily identified) CVAs that affect the ability to understand even *simple* directions.
3. Patients with advanced dementia or mental retardation (moderate to severe).

In particular, this method is appropriate for assessing the validity of performance by brain-injured persons, and persons complaining of chronic pain. Diagnoses of depression and chronic pain do not appear to significantly effect performance on the malingering tests. Therefore, failure on these malingering tests cannot be attributed to depression and chronic pain, but to motivation/malingering. The use of this method of detection of malingering is appropriate and clinically valid. This method is also appropriate for persons with known or suspected TBI.

This method of detecting malingering is unique in that it uses already established neuropsychological tests that are commonly given as part of a neuropsychological battery to assess

for malingering. Using this method, the tests that are used to assess cognitive function also assess for malingering. This allows the tests that are interpreted as part of the battery to be checked individually and collectively for validity. It has been widely shown that persons do not all malinger with the same approach or symptoms (Goebel, 1983; Greiffenstein et al., 1995; Heaton et al., 1978; Meyers & Volbrecht, 1998b; Oberg et al., 1985) and therefore a wider sampling of a variety of malingering methods is appropriate and necessary. Specific tests of malingering may be useful, but the validity of the individual neuropsychological tests needs also to be checked. In this way, the validity of the interpreted data can be assessed. Another advantage is that these tests are given as part of a general neuropsychological assessment and, when used as validity checks, do not add any additional time to the assessment. In this time of managed care and limited funding, tests that can do "double duty" have an obvious advantage to the clinician.

This study is a compilation and second validation of nine different neuropsychological tests that have the additional capacity to assess for malingering. These tests are commonly used neuropsychological tests that have the advantage of performing cognitive assessment and malingering, thus allowing for motivation to be assessed throughout the neuropsychological battery. This method showed 100% specificity and an 83% sensitivity. This method has a minimal false positive rate (0%); therefore in a medical–legal setting a malingerer may be missed (e.g., may not be caught by this method, 17% false negative). However, if failure is obtained on two or more of these measures the clinician may have confidence that motivational issues are present in performance on the neuropsychological battery.

## References

Audiotec of St. Louis. (1991). *Dichotic Word Listening Test (DWLT)*. St. Louis, MO: Audiotec.

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

Benton, A., Hamsher, K., Varney, N., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual*. New York: Oxford University Press.

Bernard, L. (1991). The detection of faked deficits on the Rey Auditory Verbal Learning Test: The effect of serial position. *Archives of Clinical Neuropsychology*, *6*, 81–88.

Bernard, L., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. *Journal of Clinical Psychology*, *49*, 45–53.

Binder, L. M. (1992). Malingering detected by forced choice testing of memory and tactile sensation: A case report. *Archives of Clinical Neuropsychology*, *7*, 155–163.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology*, *15*, 170–182.

Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry*, *153*, 7–10.

Brandt, J., Rubinsky, E., & Lassen, G. (1985). Uncovering malingered amnesia. *Annals of the New York Academy of Science*, *44*, 502–503.

Damasio, H., & Damasio, A. (1979). "Paradoxic" ear extinction in dichotic listening: Possible anatomic significance. *Neurology*, *29*, 644–653.

Goebel, R. (1983). Detection of faking on the Halstead Reitan Test Battery. *Journal of Clinical Psychology*, *39*, 731–742.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*(3), 218–240.

Greiffenstein, M. F., Gola, T., & Baker, W. J. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist*, *9*(3), 230–240.

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

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, *11*, 967–974.

Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessments. *Journal of Trauma Rehabilitation*, *15*(2), 829–858.

Kozel, J., & Meyers, J. (1998). A cross-validation study of the Victoria Revision of the Category Test. *Archives of Clinical Neuropsychology*, *13*, 327–332.

Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology*, *11*(1), 45–51.

Levin, H., High, W., Williams, D., Eisenberg, H., Amparo, E., Guinto, F., & Ewert, J. (1989). Dichotic listening and manual performance in relation to magnetic resonance imaging after closed head injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, *52*, 1162–1169.

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

Meyers, J. E., Bayless, J. D., & Meyers, K. R. (1996). The Rey Complex Figure: Memory error patterns and functional abilities. *Applied Neuropsychology*, *3*, 89–92.

Meyers, J. E., & Diep, A. (2000). Assessment of malingering in chronic pain patients using neuropsychological tests. *Applied Neuropsychology*, *7*, 133–139.

Meyers, J. E., Galinsky, A. M., & Volbrecht, M. (1999). Malingering and mild brain injury: How low is too low. *Applied Neuropsychology*, *6*(4), 208–216.

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

Meyers, J. E., Morrison, A. L., & Miller, J. C. (2001). How low is too low, revisited: Sentence repetition and AVLT recognition in the detection of malingering. *Applied Neuropsychology*, *8*, 234–241.

Meyers, J. E., Roberts, R. J., Bayless, J. D., Volkert, K. T., & Evitts, P. E. (2002). Dichotic listening: Expanded norms and clinical application. *Archives of Clinical Neuropsychology*, *17*, 79–90.

Meyers, J. E., & Volbrecht, M. (1998a). Validation of memory error patterns on the Rey Complex Figure and Recognition Trial. *Applied Neuropsychology*, *5*, 120–131.

Meyers, J. E., & Volbrecht, M. (1998b). Validation of Reliable Digits for detection of malingering. *Assessment*, *5*, 301–305.

Meyers, J. E., & Volbrecht, M. (1999). Detection of malingerers using the Rey Complex Figure and Recognition Trial. *Applied Neuropsychology*, *6*, 201–207.

Meyers, J. E., Volkert, K. T., & Diep, A. (2000). Sentence Repetition Test: Updated norms and clinical utility. *Applied Neuropsychology*, *7*, 154–159.

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

Oberg, R., Udessen, H., Thomsen, A., Gade, A., & Mortensen, E. (1985). Psychogenic behavioral impairment in patients exposed to neurotoxins: Neuropsychological assessment in differential diagnosis. In *Neurobehavioral methods in occupational and environmental health* (Environmental Health Document 3, pp. 130–135). Copenhagen, Denmark: World Health Organization.

Pilgrim, B., Meyers, J. E., Bayless, J. D., & Whetstone, M. (1999). Validity of the Ward Seven-Subtest WAIS-III Short Form in a neuropsychological population. *Applied Neuropsychology*, *6*, 243–246.

Reitan, R., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsychological Test Battery: Theory and clinical interpretation*. Tuscon, AZ: Neuropsychology Press.

Risse, G., Gates, J., Lund, G., Maxwell, R., & Rubens, A. (1989). Interhemispheric transfer in patients with incomplete section of the corpus callosum: Anatomic verification with magnetic resonance imaging. *Archives of Neurology*, *46*, 437–443.

Roberts, M., Persinger, M., Grote, C., Evertowski, L., Springer, J., Tuten, T., Moulden, D., Franzen, K., Roberts, R., & Baglio, C. (1994). The Dichotic Word Listening Test: Preliminary observations in American and Canadian samples. *Applied Neuropsychology*, *1*, 45–56.

Russell, M. L., Spector, J., & Kelly, M. (1993, February). *Primary and recency effects in the detection of malingering using the WMS-R Logical Memory Subtests*. Poster presented at the 21st annual meeting of the International Neuropsychological Society, Galveston, TX, USA.

Shimoyama, I., Ninchoji, T., & Uemura, K. (1990). The Finger Tapping Test: A quantitative analysis. *Archives of Neurology*, *47*, 681–684.

Slick, D. J., Sherman, E., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms and commentary* (2nd ed.). New York: Oxford University Press.

Suhr, J. A., & Gunstad, J. (2000). The effect of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, *15*(5), 415–424.

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

Ward, L. (1990). Prediction of verbal, performance and full scale IQ from seven subtests of the WAIS-R. *Journal of Clinical Psychology*, *46*, 436–440.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale—III and the Wechsler Memory Scale—III* (3rd ed.). San Antonio, TX: The Psychological Corporation.