ASSESSMENT
1998, Volume 5, Number 3, 303-307

VALIDATION OF RELIABLE DIGITS FOR DETECTION OF MALINGERING

John E. Meyers
Northwest Iowa Rehabilitation Consultants

Marie Volbrecht
University of South Dakota

This study is a cross validation of the previous work by Greiffenstein, Baker, and Gola (1994) and Greiffenstein, Gola, and Baker (1995) on the Reliable Digits (RD) method of detecting possible malingering. This study consisted of 47 mild brain-injured litigating and 49 mild brain-injured non-litigating participants. The result of this study was that only 4.1% of the non-litigating participants were classified as malingering by RD, while 48.9% of the litigating participants were classified as malingering by RD. Comparing litigating participants' performance on a forced choice task with their RD revealed that 77.8% (7/9) who failed the forced choice task also failed RD. RD classified more litigating patients as malingering than did the forced choice task and none of the non-litigating participants failed forced choice. These findings underscore the previous recommendations of Greiffenstein et al. (1994, 1995), that it is important to assess motivation on specific neuropsychological tests and that motivation on neuropsychological tests is not an all or none phenomenon.

Keywords: Malingering, neuropsychology, Reliable Digits, forced choice, brain injury litigation

The detection of malingering on neuropsychological evaluations is becoming increasingly important. This importance is recognized in the field of neuropsychology with many attempts to identify malingered performance on a variety of measures. The current study deals with two specific measures. However, those interested in a more comprehensive review of measures to identify possible malingering should refer to Lezak (1995).

Correspondence concerning this article and requests for offprints should be addressed to John E. Meyers, PsyD, Northwest Iowa Rehabilitation Consultants, Suite 340, 500 Jackson Street, Central Medical Building, Sioux City, IA 51101.

Greiffenstein, Baker, and Gola (1994) presented a new method for identification of malingering called Reliable Digits (RD). This method utilizes the longest number of digits repeated correctly on both trials of Digit Span (forward + backward). They report that a score of 7 or less suggests the possibility of malingering. In their study, they found that individuals defined as probable malingerers showed an average RD score of 6.7 (SD = 1.2) and non-malingering traumatic brain injury patients obtained an average score of 8.8 (SD = 1.2). They defined malingering based on the following classification system: (a) improbable poor performance on two or more measures, (b) total disability in a major social role, (c) contradiction

between collateral sources and symptom history, and (d) remote memory loss. If a patient met two or more of these criteria, the patient was classified as a possible malingerer.

In a second study, Greiffenstein, Gola, and Baker (1995) used the same classification system as in their 1994 study and found that severe brain-injured patients achieved a mean RD score of 8.75 (SD = 1.87) and probable malingerers obtained an average RD score of 6.60 (SD = 1.83). In the first study, a 33% to 60% rate of malingering among litigants was reported and, in the second study, a 55% malingering rate was reported. The authors again reported that a score of 7 or less suggests malingered performance. However, examination of the means and standard deviations reported, suggests that many of the non-litigating Traumatic Brain Injured (TBI) participants would be classified as malingerers. Greiffenstein et al. (1995) reported an 86% sensitivity and 57% specificity for the RD. This method of malingering detection has advantages over other special "tests" of malingering in that data that is normally used in neuropsychological test interpretation has a validity check. In addition, RD uses no special equipment and is available to most neuropsychologists in a battery of tests.

One important point in the Greiffenstein et al. (1995) study was that persons feigning severe TBI do not attempt to feign psychotic symptoms. They also suggest that malingering be treated as a dimension of behavior which should be considered separately from the presence or absence of neuropsychological deficits. They warn, "the clinician should not become trapped in false dichotomies ('brain damage versus malingering')" (p. 236). They also found that individuals may attempt to feign specific neuropsychological difficulties in a credible fashion rather than attempting to feign global cognitive, social, and emotional problems. This suggests that real world malingerers may attempt to feign specific difficulties that they attribute to brain injury, and less often attempt to feign global difficulties. Their findings suggest that some litigants may attempt to malinger on a particular neuropsychological test that they feel represents their idea of brain injury. If one is attempting to feign a visual perceptual difficulty, one might do well when given a task such as Digit Span, but do poorly when given a task such as the Rey Complex Figure (Meyers & Meyers, 1995).

These findings support the need for specific validity checks within the neuropsychological battery. The validity checks should be part of the neuropsychological tests that are being interpreted, rather than relying upon tasks designed specifically to detect malingering. Specific malingering tests may have usefulness, but the validity of the neuropsychological tests to be interpreted must also be checked. Therefore, it is important to develop validity checks that are already parts of the neuropsychological tests interpreted as part of the battery.

One often used "gold standard" for the detection of malingering is the use of forced choice (FC) techniques. This method uses performance below chance levels (i.e., 50%) to assess for malingering. Performance that is below the chance level is considered to be indicative of purposefully inhibiting responses. It is assumed that even random performance would be at least at the chance level. One of the earliest FC measures was provided by Brandt, Rubinsky, and Larson (1985). This task consists of an initial recall of 20 unrelated words, followed immediately with a recognition forced choice task. The participants are asked to identify which of two words were on the list. Scores of 10 out of 20 or less on the recognition task would be suggestive of malingered performance.

The FC method uses an indication of purposefully inhibiting performance. The RD method for detecting possible malingering uses performance that is improbably low, meaning that performance is below that expected given a suspected mild brain injury. In this respect, RD may be a measure of reduced motivation and FC may be a measure of purposeful malingering. Submaximal effort may be an indication of purposeful malingering.

Hypothesis

The purpose of this study is to provide a cross validation of the usefulness of RD in identifying

malingered performance on neuropsychological tests. It is hypothesized that similar results to those obtained by Greiffenstein et al. (1994, 1995) will be found. That is, we predict that RD will discriminate patients with suspected malingering from non-malingering patients.

Methods

Participants

Forty-seven mild TBI patients with less than an hour of reported loss of consciousness (LOC) and who were involved in litigation (i.e., third party law suits or disability) were classified as Group 1. Group 1 participants had a mean age of 39.77 years (SD = 9.69 years) with an average of 12.26 years of education (SD = 2.44 years). The average time since reported injury was 16.81 months (SD = 41.57 months). Performance on the WAIS-R (seven subtest short form; Ward, 1990) FSIQ showed a mean score of 89.00 (SD = 11.34). Twelve participants were female and 35 were male. Forty-two participants were right handed and 5 were left handed, and all but 1 were Caucasian.

Group 2 participants consisted of 49 referrals for neuropsychological evaluation by treating physicians. All participants were seen in the context of evaluation and treatment. None of these Group 2 participants were involved in litigation. The mean age for Group 2 participants was 36.00 years (SD = 10.76 years) with an average of 13.94 years of education (SD = 2.51 years). They were an average of 16.37 months (SD = 34.09 months) post injury. The mean WAIS-R FSIQ was 100.65 (SD = 12.06). All had less than an hour of reported LOC. Twenty-six participants were female and 23 were male. All participants were right handed and all but 1 were Caucasian.

For all participants (Groups 1 and 2), medical records and school records were available. All had sustained an injury from a motor vehicle accident or blow to the head. All complained not only of cognitive difficulties, but also chronic pain in the head, neck, and/or back. All were assessed with a comprehensive neuropsychological battery that included RD.

Materials

The battery consisted of the WAIS-R, Trail Making Test (Reitan & Wolfson, 1985), Judgment of Line Orientation (Benton, Hamsher, Varney, & Spreen, 1983), Finger Tapping (Reitan & Wolfson, 1985), Finger Localization (Benton et al., 1983), Token Test (Spreen & Strauss, 1991), Sentence Repetition (Spreen & Strauss, 1991), COWAT (FAS) (Spreen & Strauss, 1991), Animal Naming (Spreen & Strauss, 1991), Boston Naming Test (short form; Mack, Freed, Williams, & Henderson, 1992), Dichotic Listening (Roberts et al., 1994), AVLT (Spreen & Strauss, 1991), Rey Complex Figure and Recognition Trial (Meyers & Meyers, 1995), and the Booklet Category Test (Victoria Revision; Spreen & Strauss, 1991). In addition, 40 of the Group 1 participants and 39 of the Group 2 participants were administered a forced choice (FC) task.

Procedure

All participants were administered the neuropsychological measures by a trained master's level or graduate student technician. All measures were administered according to standardized procedures. Scores for RD were obtained by summing the digits of the longest string done without error on both trials forward and on both trials backward on the Digit Span subtest of the WAIS-R.

Results

Independent samples t tests were used to compare the demographics of Group 1 and Group 2 participants. The groups were not significantly different on age, LOC, or months since injury at the time of evaluation (all $ps > .05$). Significant differences were found in years of education $t(1, 94) = -3.326$, $p = .001$, FSIQ $t(1, 94) = -4.872$, $p = .000$, and RD $t(1, 94) = -3.915$, $p = .000$. Group 1 had a mean RD score of 8.02 (SD = 2.07) and Group 2 had a mean RD score of 9.59 (SD = 1.86). A Pearson's Correlation between FSIQ and RD revealed an overall correlation of .52 ($p = .000$). The correlation for Group 2 was .35 ($p = .01$) and for Group 1 was .51 ($p = .000$).

A Pearson's Correlation between education and RD showed a correlation of .18 ($p > .05$) and the correlation between RD and FC was .43 ($p = .000$) for all participants. For Group 1, the correlation between RD and FC was .38 ($p = .013$), and for Group 2 the correlation was .26 ($p >, 05$). Examination of Table 1 shows the frequency of scores for the each group. Using the cutoff of 7 or less as an indication of possible malingering, only 2 of the 49 (4.1%) participants from the Group 2 (non-litigating) were classified as malingering and 23 of the 47 (48.9%) participants from the Group 1 (litigating) were classified as malingering. Examination of the records for the two non-litigating (Group 2) members that scored below the cutoff showed that they had 12 (RD = 6) and 14 (RD = 7) years of education. Therefore, although there was a difference between the groups in education, this difference did not appear to affect the score on the RD.

None of the Group 2 participants failed the FC task, although 9 of the Group 1 participants did fail FC. Of the 9 participants in Group 1 that failed FC, 7 of them also failed RD. However, 16 participants in Group 1 were classified as malingering by the RD who were not classified as malingerers by FC. Examining the correlations of the WAIS-R subtests shows variability between the FC and RD depending on the group membership (litigating or non-litigating). Table 2 shows the correlation results and the significance levels.

Table 1
Frequency Distribution of Reliable Digits (RD) Scores for Groups 1 and 2

Obtained RD Score	Number with score	
	Group 1[a]	Group 2[b]
3	1	0
4	0	0
5	4	0
6	3	1
7	15	1
8	6	13
9	7	11
10	4	13
11	5	4
12	1	3
13	1	1
14	0	0
15	0	1
16	0	1

Note. [a]$n = 47$. [b]$n = 49$.

Table 2
Correlation Matrix of Wechsler Adult Intelligence Scale–Revised (WAIS-R) Subtests With Forced Choice (FC) and Reliable Digits (RD)

WAIS-R Subtests	Group 1 Litigants		Group 2 Non Litigants	
	FC	RD	FC	RD
Information	.17	.41	.17	.19
Digit Span	.33*	.91**	.46**	.85**
Arithmetic	.26	.36*	.52**	.24
Similarities	.36*	.39*	.48**	–.01
Picture Completion	.61**	.48**	.38*	.27
Block Design	.30	.40*	.07	.30*
Digit Symbol	.68**	.45**	.44**	.17
Forced Choice		.38*		.26

Note. $*p < .05$. $**p < .01$.

Discussion

The two groups used in this study were very similar in age, LOC, and time since injury. Although a difference was found in the education level of the groups, education level was not an influencing factor for performance on RD. Litigating participants demonstrated much poorer performance on RD and on their general neuropsychological performance than did the non-litigating participants as demonstrated by their FSIQ. The correlations for both groups indicate that RD may be a useful validity marker for the WAIS-R and WAIS-III, given that the Digit Span subtest (from which the RD is taken) is essentially unchanged in the newer version.

If one uses FC as a "gold standard" for detection of malingering, then RD shows a 77.8% (specificity)

agreement with failure on FC and a 95% (sensitivity) correct classification of non-malingering participants. These results would then suggest an acceptable level of sensitivity and specificity for RD in identifying possible malingering patients and are similar to those obtained by Greiffenstein et al. (1995) who obtained an 86% sensitivity and a 57% specificity for detection of malingering. However, the RD does demonstrate some overall difference from FC. The overall classification rate for litigants by RD was 49%. If one assumes that litigation status is the "gold standard" for detection of malingering then nearly half of the litigants would be classified as malingering. This is within the range described by Greiffenstein et al. (1994, 1995).

However, taking into consideration the correlation of RD with the WAIS-R subtests and FSIQ, RD may be more a measure of submaximal performance on the WAIS-R. This submaximal performance may be an indication of possible malingering. Failure on more than one malingering test would be needed to support a conclusion of malingering. However, failure on one malingering test may suggest submaximal effort. The significant correlations of FSIQ with RD suggest that overall motivation on the WAIS-R (III) may be evident by performance on RD. However, the findings that the RD and FC classified some different litigants as malingering underscores the need to use multiple measures of motivation in the neuropsychological assessment. FC is less correlated with performance on the WAIS-R (III) than is the RD. These results also support the previous findings by Greiffenstein et al. (1995) that litigating patients did not "malinger" on all tests equally. This further supports Greiffenstein et al. (1995) recommendations that the validity of each of the specific neuropsychological tests used to identify brain injury should be assessed.

Individual markers of validity, within individual neuropsychological tests used to assess brain injury in a battery, need to be developed. Specific and separate tests of motivation may be helpful, but it is important that the validity of those specific neuropsychological tests that are used in the neuropsychological interpretation be assessed, as part of the interpretation process. RD appears to be able to perform this function for the WAIS-R (III). However, additional measures of motivation also need to be inspected before a decision that malingering is occurring can be made. Given that the participants in this study were TBI referrals, the clinician will need to be cautious in applying these findings to other populations. Further research may help to determine if this method for detecting possible malingering is effective with other patient groups.

References

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

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

Greiffenstein, M., Baker, J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*, 218-224.

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

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

Mack, W., Freed, D., Williams, B., & Henderson, V. (1992). Boston Naming Test: Short versions for use in Alzheimer's disease. *Journal of Gerontology: Psychological Sciences, 47*, 154-158.

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

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

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.

Spreen, S., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms and commentary*. New York: Oxford University Press.

Ward, L. (1990). Prediction of Verbal, Performance and Full Scale IQs from seven subtests of the WAIS-R. *Journal of Clinical Psychology, 46*, 436-440.