The Word Completion Memory Test (WCMT): a new test to detect malingered memory deficits

Robin C. Hilsabeck\textsuperscript{a,*}, Denny C. LeCompte\textsuperscript{b}, Allison R. Marks\textsuperscript{c}, Jordan Grafman\textsuperscript{c}

\textsuperscript{a}University of California, San Diego, CA, USA
\textsuperscript{b}BMC Software, Houston, TX, USA
\textsuperscript{c}National Institutes of Neurological Disorders and Stroke, National Institutes of Health, USA

Accepted 3 July 2000

Abstract

In recent years, much research has focused on developing tests to detect malingering. A drawback of existing tests is their poor ability to detect malingerers possessing more “sophisticated” knowledge of neuropsychological deficits. The current study presents preliminary validation data on a new measure, the Word Completion Memory Test (WCMT), which is the first malingering test to utilize a sophisticated coaching methodology in its development. The WCMT was administered to control participants, memory-impaired patients, and coached simulators. The coached simulators were provided with specific information about and examples of memory deficits commonly experienced following closed head injury (CHI; e.g., anterograde vs. retrograde amnesia). They also read a detailed scenario describing the lifestyle and motivations likely experienced by CHI litigants, and then practiced their roles by taking a quiz about their deficits. Results showed that 93\% of coached simulators and 100\% of control and memory-impaired participants were correctly classified by the WCMT. © 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

Keywords: Malingering; Neuropsychological; Analogue; Coaching; WCMT

Cognitive dysfunction is a common occurrence following closed head injury (CHI; Ellenberg, Levin, & Saydjari, 1996; Ruff et al., 1993). Given that cognitive impairments associated with CHI can be disabling (Levin, Benton, & Grossman, 1982; Lezak, 1995), it is
not surprising that CHI survivors often retain attorneys to aid in securing financial compensation. It also is not surprising that the possibility of receiving millions of dollars for lost cognitive abilities may tempt some litigants to exaggerate or fabricate their injury-related problems. In fact, according to Haines and Norris (1995), CHI is the most common neuropsychological syndrome feigned. As a result, neuropsychologists are often asked by the courts to determine the legitimacy of alleged cognitive dysfunction secondary to CHI. Thus, the detection of malingered neuropsychological impairments has become an important area of research.

The problem for neuropsychologists has been an inability to distinguish “real” from malingered deficits with an acceptable degree of certainty. Since there is currently no perfect measure of feigning and simulators are not likely to admit their deceit, researchers have been forced to rely on two primary research paradigms, known-groups and analogue designs, to study this phenomenon (Rogers, Harrell, & Liff, 1993). In known-groups designs, clinicians independent of the research project identify simulators via clinical judgment and then compare their performances on standardized measures with performances of truly impaired patients. The principal advantage of this design is its direct clinical applicability to “real-world” feigners. However, the inability of clinicians to accurately identify simulators using clinical judgment alone was the impetus for developing simulation detection measures in the first place (Faust, Hart, Guilmette, & Arkes, 1988; Heaton, Smith, Lehman, & Vogt, 1978). Therefore, the known-groups design is significantly limited by its use of clinical judgment as the criterion by which simulators are identified.

This limitation of the known-groups design has led most researchers to use the analogue design. In analogue designs, neurologically normal participants are instructed to feign impairment on standardized measures and their performances are then compared with one or more comparison groups, such as memory-disordered, brain-injured, or normal participants instructed to do their best (Rogers et al., 1993). The primary drawback to the analogue design is its unknown generalizability to the real world. In an effort to address this limitation, Rogers (1988) offered some methodological suggestions, including providing instructions to fake “believable” deficits. Franzen, Iverson, and McCracken (1990) further suggested offering differing kinds of instructions to help provide future directions for study when using analogue participants.

In general, research investigating the effects of task instructions has shown that the more information provided to help simulate impairment, the better analogue participants are at performing like truly impaired individuals. For example, Rose, Hall, and Szalda-Petree (1995) demonstrated that analogue simulators who were provided information about problems typically experienced by head-injured persons were able to avoid detection on a forced-choice simulation measure more often than analogue simulators not provided this information. Similarly, Martin, Gouvier, Todd, Bolter, and Niccolls (1992) found that analogue simulators specifically instructed to perform above chance levels and to miss more hard than easy items on a forced-choice recognition memory test performed more like brain-injured participants than analogue simulators instructed only to demonstrate memory impairment.

In spite of the above evidence that coaching analogue participants has a significant effect on test performance, some researchers continue to simply instruct analogue participants to fake “believable” cognitive deficits (Beetar & Williams, 1995; Chouinard & Rouleau, 1997), without providing them with specific information about how to fake believable deficits.
Apparently, these researchers are assuming that real-world simulators have the same knowledge base as the average layperson. However, many real-world simulators may have suffered legitimate injuries resulting in some real deficits on which to build more credible presentations (Greiffenstein, Gola, & Baker, 1995; Heaton et al., 1978). In addition, contact with physicians and other patients in the course of their lawsuits may provide information for refining their presentations (Franzen et al., 1990; Trueblood & Schmidt, 1993). Moreover, “education” about psychological testing and simulation measures provided by attorneys may influence evaluation results (Wetter & Corrigan, 1995; Youngjohn, 1995). Therefore, one could argue that providing analogue simulators with specific knowledge about how deficits are experienced by impaired individuals is critical to enhancing generalizability to real-world settings. Furthermore, without specific information about the deficit to be faked, analogue simulators may be unable to feign deficits as believably as their real-world counterparts, thereby resulting in obscured research findings regarding the classification accuracy of simulation measures.

Although many new tests have been developed to detect malingering (e.g., Binder & Willis, 1991; Chouinard & Rouleau, 1997; Frederick, 1997; Hiscock & Hiscock, 1989; Inman et al., 1998; Iverson, Franzen, & McCracken, 1991; Slick, Hopp, Strauss, & Spellacy, 1996; Tombaugh, 1997), none has been developed with these methodological concerns in mind. The Word Completion Memory Test (WCMT; Hilsabeck & LeCompte, 1997) is the first measure to use a sophisticated coaching methodology in its development. The WCMT is a priming task based on a dissociation framework proposed by Jacoby (1991). Jacoby posits that the contributions of automatic and intentional uses of memory can be separated by implementing a methodology called process dissociation. The rationale behind this approach is that “conscious control can be measured as the difference between performance when a person is trying to as compared with trying not to use information from some particular source” (p. 527). For example, on the WCMT, the examinee is first instructed to complete word stems with words from a previously studied list (i.e., inclusion task), but on a second task, is instructed to complete word stems with words that were not from a previously studied list (i.e., exclusion task). The difference score is hypothesized to be a measure of the examinee’s conscious control over his or her memory for the previously studied words.

The purpose of the present study was to test the ability of the WCMT to discriminate well-coached simulators from control and memory-impaired participants. It was hypothesized that simulating participants would perform significantly worse than controls and memory-impaired participants on the WCMT, and that the WCMT would correctly classify 80% or more of well-coached simulators while also correctly classifying 100% of control and memory-impaired participants.

1. Method

1.1. Participants

A total of 147 undergraduates participated in this experiment in exchange for extra credit in their psychology courses. Undergraduates were randomly assigned to one of two
groups, control or simulation. Thus, there were 72 and 75 participants in the control and simulation groups, respectively. Data from three control participants were excluded due to incomplete test performances, resulting in 69 control participants retained for inclusion in the analyses. The control group was predominantly female (87%), and approximately 74% were Caucasian, 15% were African American, and 11% were of other ethnicities. Mean age of control participants was 19.93 (S.D. = 2.37) years, and average educational level was 13.18 (S.D. = 1.09) years.

Of the 75 simulation participants, data from 12 were excluded due to admittance of failure to fake memory problems as instructed, and data from five participants were excluded due to failure to demonstrate adequate knowledge of the simulation role (see below). Thus, 58 simulation participants were retained for analyses. Due to an oversight, demographic information was not collected for the simulation group. However, data from multiple experiments in our laboratory have failed to reveal significant differences on demographic variables among groups of undergraduates. Therefore, there is no reason to expect that our simulation group differed significantly from the control group on any of these variables.

Because the process dissociation procedure was initially validated in patients with amnesia, inclusion in the memory-impaired group was based on the following three criteria: (1) average IQ or above, (2) intact attention, and (3) severely impaired learning and/or delayed recall abilities compared to the normative sample. Thus, only patients with an amnesic-like presentation were considered for this initial validation study. Fourteen outpatients met these criteria. None were involved in litigation or presented with evidence of other secondary gain, and none endorsed significant depressive or anxious symptoms. Mean age was 41.71 (S.D. = 20.04) years, and mean years of education was 14.29 (S.D. = 3.15). Eleven participants were male. Primary ethnicities represented were Caucasian (79%) and African American (14%). Etiologies for memory impairment were as follows: CHI (N = 7), hypoxia (N = 2), dementia (N = 2), penetrating head wound (N = 1), and brain tumor (N = 1).

1.2. Materials

1.2.1. Word Completion Memory Test

The development of the WCMT arose from observations that existing simulation measures were able to detect only very obvious attempts at faking bad and from calls by researchers that tests designed to detect sophisticated simulators were needed (Guilmette, Sparadeo, Whelihan, & Buongiorno, 1994). With this in mind, the WCMT was developed using well-coached analogue participants as the criterion group rather than analogue simulators simply instructed to fake “believable” memory impairments.

As alluded to above, the WCMT consists of two subtests, Inclusion and Exclusion. On the Inclusion subtest, each of 30 words is read aloud to the examinee. After each word is read, the participant copies the word and rates it for pleasantness. Requiring the examinee to hear, copy, and rate each word was implemented to facilitate attention to and memory for each word (cf. Craik & Lockhart, 1972). After copying and rating each word, the examinee is instructed to complete 30 word stems with words from the previously studied
list and is asked to demonstrate understanding of a “good” memory performance via an example (i.e., identifying that the person who completed more word stems with words from the sample list has the best memory). On the Exclusion subtest, the examinee copies and rates another 30 words but then is instructed to complete 30 word stems with words that were not from the previously studied list. As before, the examinee is required to demonstrate understanding of a good memory performance on this task via an example (i.e., identifying that the person who completed more word stems with words that were not from the sample list has the best memory). The examinee is not told of the Exclusion task until after the Inclusion task is completed.

The WCMT yields three scores: (1) an $I$ score, which is the number of stems completed with words from the previously studied Inclusion subtest word list; (2) an $E$ score, which is the number of stems completed with words from the previously studied Exclusion subtest word list; and 3) an $R$ score, which is the difference between the $I$ and $E$ scores. For example, a person with no memory impairment should obtain a relatively high $I$ score (e.g., 24 out of 30), a relatively low $E$ score (e.g., 4 out of 30), and a relatively high positive $R$ score (e.g., $24 - 4 = 20$). Hypothetically, a pure amnesic would perform equally on both tasks due to the facilitative effects of implicit memory (i.e., the amnesic would frequently complete stems with list words on both the Inclusion and Exclusion tasks), resulting in an $R$ score near zero; however, since a pure amnestic syndrome is rare, memory-impaired examinees also would be expected to obtain positive $R$ scores, although lower than $R$ scores obtained by control participants. In contrast, persons simulating memory impairment would be expected to obtain a negative $R$ score because an intentionally poor memory performance on the Inclusion subtest would result in few stems completed with words from the previously presented list (e.g., 5 out of 30) and an intentionally poor memory performance on the Exclusion subtest would result in many stems completed with words from the list (e.g., 20 out of 30). Thus, the $R$ score would be a negative number (e.g., $5 - 20 = -15$).

1.3. Procedure

Undergraduates assigned to the control condition and memory-impaired participants were instructed to do their best on all tasks. Control participants were administered the WCMT in groups of 10 or less. Memory-impaired participants were administered the WCMT as part of their neuropsychological evaluations.

Participants assigned to the simulation group were instructed to take on the role of a person who had suffered a mild CHI in an automobile accident for which they were seeking compensation. Participants were provided with a detailed scenario of the mild CHI survivor’s memory problems and the motivations that might lead the person to feign or exaggerate memory difficulties. The scenario was read aloud to simulating participants to ensure exposure to the information, and they were instructed to refer back to the scenario throughout the testing session to aid their abilities to undertake the role of the CHI litigant.

Next, participants in the simulation group were asked to complete a questionnaire designed to assess their understanding of the role they were being asked to play (that of a simulator)
and to give them an opportunity to practice that role. Because real-world simulators are very likely to have a clear understanding of their roles and are well-practiced at playing them, only data from simulating participants achieving 80% correct on selected items of the questionnaire were included in the analyses; we felt that this exclusion more closely approximated the level of understanding of a real-world simulator (as noted above, five participants were excluded based on this criterion).

The simulation group, then, was administered the WCMT followed by a post-experiment questionnaire to assess compliance with the procedure and understanding of instructions. As previously noted, data from 12 participants were eliminated based on their responses to the post-experiment questionnaire. A closer examination of their responses showed that five participants indicated they did not try to act like the person in the scenario, six said they did not pretend to have memory problems while taking the test, and one failed to understand the instructions on the Exclusion subtest.

2. Results

Because demographic information was not collected for the simulation group but was not expected to differ significantly from the control group (both are composed of undergraduates), Mann–Whitney U tests were conducted to examine differences between the control and memory-impaired groups. There was a significant difference between the groups for age (\(U = 206.0, P = .001\)) but not for education (\(U = 458.5, P = .823\)). Thus, the memory-impaired group was significantly older than the control and simulation groups, but education levels among the groups were not significantly different.

Due to unequal variances, differences among the groups on the three WCMT scores (i.e., \(I\), \(E\), and \(R\) scores) were analyzed using Kruskal–Wallis one-way ANOVAs. To help control for Type I errors, a Bonferroni correction was applied resulting in a significance level of .01 or less. All three Kruskal–Wallis ANOVAs were significant (\(I\) score \(x^2 = 98.637, P < .001\), \(E\) score \(x^2 = 61.713, P < .001\), and \(R\) score \(x^2 = 98.637, P < .001\)). Mann–Whitney U tests revealed significant differences among the groups for all three scores, except the \(E\) scores of the control and memory-impaired groups, which did not significantly differ from each other (see Table 1).

A visual analysis of the data revealed that an \(I\) score of \(< 15\) correctly classified 100% of control and memory-impaired participants and 86% of simulating participants. Of the remaining simulators not correctly classified with the \(I\) score criterion, another 7% are

<table>
<thead>
<tr>
<th>Group</th>
<th>(N)</th>
<th>(I) score</th>
<th>(E) score</th>
<th>(R) score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>69</td>
<td>24.01(^a) (S.D. = 3.35)</td>
<td>2.13(^d) (S.D. = 3.77)</td>
<td>21.88(^c) (S.D. = 5.42)</td>
</tr>
<tr>
<td>Simulators</td>
<td>58</td>
<td>7.93(^b) (S.D. = 5.57)</td>
<td>13.86(^e) (S.D. = 8.81)</td>
<td>– 5.93(^g) (S.D. = 11.19)</td>
</tr>
<tr>
<td>Memory-impaired</td>
<td>14</td>
<td>20.00(^c) (S.D. = 3.64)</td>
<td>3.07(^d) (S.D. = 2.79)</td>
<td>16.93(^b) (S.D. = 4.30)</td>
</tr>
</tbody>
</table>

Superscript letters indicate means that differ significantly from other means designated with a different letter.
correctly classified using an $R$ score of < 9 as an additional indicator, while still correctly classifying 100% of control and memory-impaired participants. Thus, the correct classification rate of simulators is raised to 93%, resulting in an overall correct classification rate of 97% (see Table 2).

### 3. Discussion

Results of this study indicate the WCMT is a valid measure of simulation of memory deficits. Significant differences were found among the control, simulating, and memory-impaired groups on all three WCMT scores, except for the $E$ score, which did not differ significantly between the control and memory-impaired groups. Applying cut-off scores of < 15 for the $I$ score and < 9 for the $R$ score resulted in correct classification of 93% of simulators and 100% of control and memory-impaired participants. Thus, 97% of all participants were correctly classified by the WCMT. Most importantly, no controls or truly impaired individuals were misclassified as malingerers. Patients with significant memory impairments were chosen for this study to help establish that individuals with severe memory problems are able to obtain positive (as opposed to negative) $R$ scores on the WCMT. Persons with milder memory impairments and adequate motivation would be expected to obtain even higher positive $R$ scores than these patients with more severe memory deficits. Litigants with poor motivation, on the other hand, would be expected to obtain a low positive $R$ score or a negative $R$ score due to their efforts to appear more memory-disordered than they truly are.

In addition to its promising classification accuracy, there are four primary advantages of the WCMT. First, the WCMT is the first test developed utilizing a sophisticated coaching methodology. It was developed and validated on well-coached simulators who were given detailed information about typical memory problems experienced following CHI and motivations which might lead CHI survivors to feign or exaggerate memory difficulties. Well-coached simulators were allowed to refer to the information as often as necessary during testing to aid their abilities to play the role of a simulator, and they were tested over their understanding of the role and practiced it. This methodology is similar to that used by Lamb, Berry, Wetter, and Baer (1994), and is one of the most extensive coaching methodologies detailed in the literature on malingering. While our coaching methodology focused on providing information about the type of memory impairment typically experienced following CHI, future studies should examine the resiliency of the WCMT to coaching when simulators are provided with information about how to fake deficits on the WCMT specifically.
Second, the WCMT uses a graduated scoring approach. If suspected simulators do not meet the first cut-off criterion, a second criterion can be applied and used in combination with the first. Third, the WCMT is easy to administer and score. Administration requires approximately 10–20 min, and scoring requires less than 5 min. Finally, the format of the WCMT affords the simulator more variety in choosing a simulation strategy (e.g., random responding, incorrect responding, bizarre responding, near misses, failure to respond) as opposed to symptom validity tests, which force one of two strategies, random responding and incorrect responding.

While these initial data appear promising, additional research is needed before more widespread use of the WCMT is appropriate. Limitations of the present study include the lack of control for possible effects of age and the small sample of memory-impaired participants. Although it is difficult to amass large samples of amnesic-like patients, this population was chosen for this initial validation study for two reasons: (1) to replicate the utility of the process dissociation procedure within the framework of the WCMT, and (2) to establish the performance pattern of patients with severe memory impairments. Clearly, further validation of the WCMT is warranted using larger sample sizes and a variety of populations. In particular, comparisons of truly impaired and litigating groups would be most desirable. Also, the effects of demographic variables, such as age and education, should be more thoroughly investigated. A project comparing the WCMT with other measures of simulation in larger populations is near completion.

References


