Allen, L.M., & Green, P. (9/20/02). Equivalence of the computerized and orally administered Word Memory Test effort measures. *WebPsychEmpiricist*. Retrieved (date) from: http://www.wpe.info/papers_table.html

WPE WebPsychEmpiricist

Equivalence of the computerized and orally administered Word Memory Test effort measures.

September 20, 2002

Lyle M. Allen III

CogniSyst, Inc., Durham, NC, USA

Paul Green

Neurobehavioral Associates, Edmonton, AB, Canada

Abstract

This study was designed to compare the oral and computerized versions of a widely used test containing measures of subject effort, the Word Memory Test (WMT). Both forms of the test were applied to a consecutive series of clinical cases. In half of the cases, the oral form of the WMT was given first and, in the remainder, the computerized test was given first. The results show that the effort subtests in the two forms of the test are closely equivalent. It is recommended that the same cut-offs for inadequate effort should be used for both the oral and computerized forms of the WMT.

Keywords: effort, symptom validity test, Word Memory Test

Please address correspondence to: Paul Green, Neurobehavioral Associates, 17107 – 107 Avenue, #201, Edmonton, Alberta, Canada T5S 1G3 or email: paulgreen@shaw.ca or lyleallen@nc.rr.com

Equivalence of the computerized and orally-administered Word Memory Test effort measures

9/20/02

The Word Memory Test (WMT; Green, Allen & Astner, 1996) is a word list learning and memory test designed with built-in measures of subject effort. Most of the research on the WMT has examined its utility as a symptom validity test, to detect feigning, symptom exaggeration or suboptimal effort during examination. Although the initial validation study employed an oral form, most of the published WMT research has involved the computerized form of the test (e.g., Green & Iverson, 2001; Green, Iverson & Allen, 1999; Iverson, Green & Gervais, 1999; Green, Rohling, Lees-Haley & Allen, 2001; Gervais et al, 2001). In the current study, we compared the oral and computerized WMT measures of effort and found them to be closely equivalent.

In the computerized WMT, the subject is presented with 20 semantically linked simple word pairs (e.g., “man – woman”), each pair appearing on the screen for six seconds. After the list is presented twice, there is an immediate recognition trial (IR), in which the person is shown new word pairs containing only one of the words from the original list and must select the words belonging to the original list (e.g., “man” from the pair “man – boy”). Without advance warning, a similar delayed recognition procedure (DR) is administered after 30 minutes, using different foil words paired with the original words (e.g., “man – teenager”). WMT scoring includes a calculation of consistency of responding from IR to DR (called “Consistency” in the tables below), and this measure is also sensitive to patient effort. The latter effort measures are followed by a series of tests of gradually increasing difficulty, which measure verbal memory. These include multiple choice (MC), in which the person is shown the first word of each pair and has to select the other word from eight options; paired associates (PA), in which the tester tells the person the first word and asks for the second word in each pair; delayed free recall (DFR), involving the person saying as many words as possible from the list in any order, while the tester records responses on the computer, and long delayed free recall (LDFR), which is the same as DFR but after another twenty minute delay.

The recognition subtests (IR and DR) serve a dual function as effort measures and also as additional learning trials because all words from the list are presented, feedback is given on correctness of responses and most people can score almost 100% correct if making a good effort. The MC and PA trials provide further exposure to the 20 first words from the pairs, again serving a dual purpose as tests of memory and as learning trials, prior to free recall. Not counting the delays between the three sections of the test in which other non-verbal tests or self-rating scales can be given, the WMT requires 12 to18 minutes of patient time to generate three effort scores and four memory scores. The computerized form is administered in a largely unattended manner and six to eight minutes of this time is required of the examiner for administration and simultaneous real-time computer scoring of the PA, DFR and LDFR subtests. All details of results are printed automatically, including comparisons between the person's scores on each subtest and scores from multiple standardization samples, such as bright normal controls, patients with severe brain injuries, neurological patients, and patients of average intelligence with either normal or impaired verbal memory.

The oral WMT is an alternative form of the WMT, with several advantages, such as ease of administration without a computer, portability and the ability to test people who may be unable to read. It consists of identical stimulus words and subtest procedures to the computerized form, except that: words are read aloud to the subject in the learning phase; the person orally chooses from word pairs read out in the recognition subtests; the multiple choice options are written down and the person circles the second word from eight options; in the paired associates trial, the tester reads out the first word from each pair and the person is asked to say the second word; free recall is tested orally and feedback is not given on correctness of responses on any oral subtest. Because of the absence of separate norms for the oral form of the memory subtests, it is recommended that only the computer version should be used for memory testing, until oral norms for the memory subtests become available.

The original WMT manual described oral WMT results from 58 patients with heterogeneous diagnoses and computerized results from 159 patients, as well as 11 sophisticated simulators given the computer version. For cross-validation purposes, all subjects were also given a second symptom validity test, the Computerized Assessment of Response Bias (CARB; Allen, Conder, Green & Cox, 1997). In a study of 298 consecutive head injury patients, 64 cases with moderate to severe brain injury scored a mean greater than 90% correct on all of the computerized WMT effort measures (Green et al., 1999). In supplements to the CARB and WMT manuals, the results on these computerized tests were described from 57 patients with moderate to severe traumatic brain injuries (STBI) and 40 patients with various neurological disorders (Allen & Green, 1999; Green & Allen, 1999). Means and standard deviations for CARB and the three primary WMT effort measures were greater than 90% correct in each of these samples. The criteria for failure on the WMT effort measures defined in the test manual were used in the current study. They were based on the results from normal volunteers, patients with impaired memory, patients with severe traumatic brain injuries and neurological disorders. When response bias is evident in failure on WMT effort measures, it is likely that the person's scores on some or all of the other cognitive tests given in the same assessment will significantly underestimate the person's true abilities. Green et al. (2001) found in a series of 904 consecutive clinical cases that poor effort on testing had a considerably greater effect in suppressing neuropsychological test scores than did severe brain injuries, brain tumors, aneurysms or strokes. Failure on any of the WMT effort measures, therefore, raises doubts about the validity of the patient's test results and claimed cognitive impairment.

Clinical experience suggests that the effort measures of the two forms of the WMT are roughly equivalent, but no formal research had previously compared the alternate forms directly. The symptom validity cutoff scores for the oral form were originally based on a small sample of patients, and out of caution, owing to the relatively small severe brain injury sample size in 1996, the cutoffs were set very conservatively and slightly lower than those later derived from the larger computer-administered samples. The present research was undertaken to assess directly the equivalence of the oral and computer-administered WMT effort measures by administering both tests to a series of 52 clinical cases in a counterbalanced order of administration.

Method

Subjects

A total of 52 clinical cases referred for compensation-related neuropsychological evaluations were studied. The sample was 65% male with an average of 40.9 years of age (SD = 12.5) and 12.3 years of education (SD = 2.5). All cases were fluent in English and 95% were right handed. The sample consisted of 30 patients assessed after head injury (24 mild head injuries and six moderate to severe brain injuries); seven patients with neurological diseases (including ruptured cerebral aneurysm, stroke and multiple sclerosis); eight patients with a primary psychiatric diagnosis (six major depression, two bipolar disorder); two patients with orthopedic injuries; one patient with chronic pain disorder and four patients with other miscellaneous conditions.

Procedure

All patients were assessed as part of a consecutive series of referrals. All patients were given a full neuropsychological evaluation that also included the Computerized Assessment of Response Bias (CARB; Conder, Allen & Cox, 1992; Allen, et al., 1997) as an independent means of assessing subject effort. Assignment to one of two conditions was alternated based strictly on examination date. A total of 28 patients (54%) was given the oral WMT on the first day of a two-day evaluation and then given the computerized version on day two. The order was reversed for the other 24 patients. On the first day of testing, patients given either form of the WMT received all of the memory subtests in that form of the test, but on the second administration, patients were only given the WMT effort subtests. Three patients began the assessment with the oral WMT but were dropped from the study because of failing to attend for the second day or for not completing both forms of tests for reasons unrelated to their test scores (e.g., the computer accidentally turned off by patient's foot on power bar during test or patient became too sick to carry on). Only one patient had been given both forms of the test prior to this study and, although he was not in the consecutive series, his scores were used as the first subject in the group given the computer form first. In accordance with standard administration procedures, once it was clear that the task instructions were being followed, the patients were left alone to complete the computer-administered immediate and delayed recognition trials of the WMT or the CARB. The oral WMT effort measures were administered face to face, according to the instructions contained in the manual.

Results

The mean subtest scores for the 52 patients given both forms of the WMT are contained in Table 1. There were no significant differences between scores on the two forms on any of the WMT effort subtests ($p > .08$ for all comparisons). Because the data are not normally distributed, the nonparametric Wilcoxon signed-ranks test for paired comparisons was used to compare oral and computerized WMT test scores. A total of 12 within-subject comparisons was undertaken, and a Bonferroni alpha level correction (.05/12 = 0.0042) was adopted which produced a significance criterion of $p <= .004$.

To further assess equivalence of the two forms, patients were separated into two groups. The first group consisted of all patients who passed all three oral WMT effort measures and also passed all three computerized WMT effort measures (N=34). The second group included those who failed any of the effort measures in either form of the WMT (N=18). These results are also shown in table 1. Separate analyses were then undertaken comparing the oral and computerized scores within each group. There were no significant differences between scores on the two forms of the WMT in patients who passed all subtests ($p > 0.67$ for all comparisons). Similarly, there were no significant differences between scores on the two forms of the WMT in the patients who failed any WMT measure (*p* >=.02 for all comparisons).

Passing CARB was also used as an external criterion to establish patient effort, as in the study of Green, et al. (1999). In the 43 patients who passed CARB, there were no significant differences between their performances on any of the oral or computerized WMT effort subtests ($p > 0.37$ for all comparisons). Means and standard deviations for each of these subgroups are contained in Table 1.

Failure rates on each form of the WMT are shown in Table 2. Failures on each of the WMT oral and computerized effort measures and on CARB were defined according to the criteria defined in the test manuals. The criteria will not be specified here in order to safeguard the integrity of the tests. The presence or absence of poor effort was calculated for each form of the WMT, and two categorical variables were created based on whether or not patients passed all three effort measures in each form of the test. A total of 16 patients failed the computerized form, 14 patients failed the oral form, and a total of 18 patients failed the WMT effort measures on either form of the test. A discriminant function analysis was performed utilizing failure of any computerized WMT effort measure to predict passing or failing any oral WMT effort measure. This analysis was highly significant ($Chi = 36.6$, $p < 0.0001$) and correctly identified pass or fail on the oral form in 88.5% of the 52 patients in the study.

Pearson correlations using all patients are displayed in Table 3. Correlations between the effort scores on the oral and computerized forms range between 0.80 and 0.90 ($p < .001$ for values in the table). Spearman rank order correlations using the same data were somewhat lower, with correlations between the oral and computerized forms of the WMT effort measures also highly significant ($p<.001$ for all). The Spearman rank correlation between WMTIR and IRORAL was $r = 0.62$, between WMTDR and DRORAL was $r = 0.73$, and between CONS and O-CONS was $r = 0.66$. The marked ceiling effect that characterizes normal effortful performance on these measures appears to be responsible for the difference in correlation values between these two statistical methods. There was a strong ceiling effect in the WMT results from these 52 patients. For example, there was a median score of 92.5% or greater for immediate recognition, delayed recognition and consistency on both forms of the WMT. Similarly, the median CARB score was extremely high at 97.5%. As expected from severely restricted variance, when Pearson correlations were undertaken using only the 34 patients passing all subtests in both WMT forms, the relationship between the two forms was far weaker and the highest correlation was found to be 0.39 ($p = 0.02$) between the oral and computer versions of DR (delayed recognition). The Spearman rank order correlation between these two measures was slightly higher ($r = 0.42$, $p = 0.013$).

The effects of practice were examined by calculating a new variable for the average of IR and DR subtests on first or second testing, irrespective of the form of the test given first. Combining the two forms of the WMT, the second administration did not result in significantly improved scores for the average of IR and DR compared with the first administration. This was true when all patients were examined ($Z = -1.35$, $p = 0.18$) and it was also true for only patients demonstrating good effort on both forms of the WMT ($Z = -1.68$; $p = 0.09$).

The effect of WMT failure on neuropsychological tests was also examined. The 34 patients passing all six oral and computerized WMT effort measures were compared with the 18 patients who failed any of these measures on a number of neuropsychological measures. These results are summarized in Table 4, which also includes effect size calculations using pooled variance (Hodge's *g*). The WMT memory measures in Table 4 are the combined results from the oral and computerized forms. These data illustrate the widespread impact of poor effort, inferred from symptom validity test failure, in suppressing scores on a variety of cognitive tests. Mean and median effect sizes for the ability measures (all measures except CARB) equaled 0.48, which indicates that poor effort had a moderate overall effect on test scores.

Post-hoc ANOVA showed that failure on any WMT effort subtest in either the oral or computerized version (N=18) was unrelated to age, sex, handedness, years of education, or being foreign-born ($p > .08$ for all comparisons). A single variable reflecting WMT performance was calculated by summing all six effort measures in the oral and computerized forms. Spearman rank order correlations showed that this overall WMT effort score was not significantly related to age ($p > 0.30$), sex ($p > 0.41$), years of education ($p > 0.14$), handedness ($p > 0.09$) or English as a second language ($p > 0.24$).

Discussion

The results of this study demonstrate that the effort subtests of the oral and computerized forms of the WMT are of equivalent difficulty. When data from all 52 patients were used, the differences in mean scores between the oral and computerized forms failed to reach significance for any of the three WMT effort measures. When patients were selected for good effort based on CARB performance or for passing all six effort measures of the oral and computerized WMT, no differences were observed between mean scores on these two forms of the test on the forced choice recognition measures (WMT-IR & WMT-DR) or on consistency of performance between these subtests. In these six comparisons, the largest difference in mean effort test scores between the oral and computer-administered forms was only 1.1% of the maximum score. There were also no significant differences between the mean scores on the alternate forms of the WMT effort measures in the 18 patients who showed any evidence of poor WMT effort.

The results from the discriminant function analysis support the conclusion that the WMT oral and computerized effort measures are equivalent. There was 88.5% agreement between the oral and computer forms of the WMT with regard to the classification of effort as satisfactory or suboptimal, based on failing any subtest of the WMT. This compares very favorably with comparison rates between different symptom validity tests in large clinical samples (see, for example, Allen, et al., 1997; Frederick, 1997). The 11.5% discrepancy between the two forms of the test may, in part, be a function of the high variability in test scores which characterizes symptom exaggerators. Such high variability is illustrated by the fact that the standard deviations in the WMT effort measures were three to four times larger in the group failing any WMT effort measure compared with the performance of patients who passed all of these measures. One well-known symptom validity test, the Validity Indicator Profile (VIP; Frederick, 1997) exclusively utilizes performance inconsistency to identify invalid test results. Similarly, the WMT effort measures include a measure of consistency of performance across equivalent tasks (e.g. "Consistency" in Table 1).

WMT effort subtest scores do not behave like measures of ability. Previous research has shown that, paradoxically, patients with moderate to severe brain injuries scored higher than patients with mild or unverified head injury on the WMT effort measures (Green, et al., 1999). This finding was interpreted to mean that failure is a function of suboptimal effort during testing. A similar finding was reported when using CARB (Green & Iverson, 2001). Also, unlike tests of ability, WMT effort test performance is almost entirely independent of demographic and other variables usually associated with differences in cognitive ability, such as age and years of education (Green, et al., 1996). An absence of significant correlations between WMT effort scores and the latter variables was confirmed in the current study. Despite the lack of effect of such demographic variables on WMT effort tests, the present study shows that poor effort on the WMT is associated with significantly reduced scores on other tests, which has already been reported by Green et al.

(2001). The measures presented in table 4 were not exhaustive. However, it is clear that, in comparison with cases passing all the WMT effort subtests, those who failed any oral or computerized WMT effort subtest scored significantly lower on a number of ability tests, including tests of memory and executive abilities. Thus, failing the WMT effort measures implies a generalized tendency to produce test results which underestimate the person's true abilities. The largest effect sizes were associated with verbal memory measures.

In patients suspected of putting forth suboptimal effort, there was a trend towards higher scores on the oral WMT administration, with mean scores up to seven percent higher than on the computerized administration (see Tables 1 and 2) but the differences were not significant. It is possible that, on face-to-face oral administration, the exaggerating patients moderated the degree to which they performed poorly on the WMT compared with their performance when they were unobserved on the computerized version. The modest sample sizes and high degree of variability observed in those failing any WMT measure in this research render this conclusion tentative. Further exploration of a potential "observer effect" for face-to-face versus computerized administration of the WMT will require additional research. For example, a comparison could be made between the scores from repeated administration of the computerized test, in which the tester is present throughout the test in one administration but absent from the other. If it were found that the presence of an observer on the WMT effort measures leads to higher scores than unattended computer administration in patients exaggerating their impairment, this would indicate the need to develop separate cutoffs for the oral and computer forms of the test, and for observed versus unobserved computer administration.

The equivalence of the oral and computerized forms of the WMT effort measures in the present study leads us to conclude that users of the oral form should employ the computerized cutoffs when evaluating subject effort with the oral form. Sample sizes in published research utilizing the computerized WMT far exceed those involving the oral form. A recent manual supplement and the current WMT software provide extensive data for the WMT effort and ability measures drawn from various patient and non-patient samples, such as patients with severe traumatic brain injury or neurological diseases, normal controls and patients with normal memory on other verbal memory tests (Allen & Green, 1999; Green & Allen, 1999). Owing to time constraints, the current research focused only on the question of equivalence in the WMT effort measures, and could not fully assess the WMT measures of verbal memory ability. Whereas the oral and computerized WMT effort measures appear to be equivalent, the extrapolation from normative data based on the computerized WMT measures of memory ability to the oral form has not yet been empirically supported, and is the subject of ongoing research.

References

Allen, L.M., & Green, P. (1999). Severe TBI Sample Performance on CARB and the WMT. In *1999 Supplement to CARB '97 and Word Memory Test Manuals*. CogniSyst, Inc., Durham, NC.

Allen, L.M., Conder, R.L., Green, P., & Cox, D.R. (1997). *CARB '97 Manual for the Computerized Assessment of Response Bias*. CogniSyst, Inc., Durham, NC.

Conder, R., Allen, L.M., & Cox, D. (1992). *Manual for the Computerized Assessment of Response Bias*, CogniSyst, Inc., Durham, N.C.

Frederick, R.I. (1997). *Validity Indicator Profile Manual*, National Computer Systems, Minneapolis, MN.

Green, P., Iverson, G.L., & Allen, L.M. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, *13*, (10), 813-819.

Green, P., & Iverson, G.L. (2001). Validation of the Computerized Assessment of Response Bias in litigating patients with head injuries. *The Clinical Neuropsychologist. 15*(4), 492-497.

Green, P., Allen, L.M., & Astner, K. (1996). *The Word Memory Test (Manual): A User's Guide to the Oral and Computer-Administered Forms, US Version 1.1*. CogniSyst, Inc., Durham, NC.

Green, P., & Allen, L.M., (1999). Performance of Neurological Patients on the Word Memory Test (WMT) and Computerized Assessment of Response Bias (CARB). In *1999 Supplement to CARB '97 and Word Memory Test Manuals*. CogniSyst, Inc., Durham, N.C.

Green, P., Rohling, M..L., Lees-Haley, P.R., & Allen, L.M. (2001) Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15* (12) 1045-1060

Gervais, R.O., Russell, A.S., Green, P., Allen, L.M., Ferrari, R. and Pieschl, S D. (2001) Effort Testing in Patients with Fibromyalgia and Disability Incentives. *Journal of Rheumatology, 28*, 1892-1899.

Iverson, G., Green P. & Gervais, R. (1999). Using the Word Memory Test to detect biased responding in head injury litigation. *Journal of Cognitive Rehabilitation, 17*(2), 4-8.

Jackson, D. N. (1984). *Multidimensional Aptitude Battery (MAB) Manual*. Port Huron, MI Sigma Assessment Systems.

Table 1.

Means and standard deviations for effort measures of the oral and computerized WMT

		% correct on IR	% correct on DR	% Consistency
All 52 cases	*Oral form*	88.1 (16.3)	89.5 (15.0)	87.7 (14.2)
	Computer form	87.0 (17.3)	87.2 (16.6)	84.9 (16.5)
Cases Passing WMT	*Oral form*	96.4 (4.4)	97.1 (3.6)	95.3 (4.9)
	Computer form	96.8 (4.5)	96.8 (4.0)	95.0 (5.7)
Cases Failing WMT	*Oral form*	72.5 (19.2)	75.3 (17.9)	73.3 (15.1)
	Computer form	68.6 (17.6)	69.0 (16.2)	65.7 (12.5)
Cases Passing CARB	*Oral form*	90.3 (14.9)	91.8 (14.1)	90.2 (12.6)
	Computer form	90.9 (14.7)	90.8 (14.5)	89.1 (14.1)

Note. Means and (standard deviations) for effort measures of the oral and computerized WMT are shown in percent correct. The table shows results from 34 clinical compensation cases passing all of the WMT effort measures, 18 cases who failed any WMT effort subtest, and 43 patients who passed an independent effort test, the CARB. IR = Immediate Recognition trial, DR = Delayed Recognition trial.

Table 2.

Number of cases (and percentage of group) failing IR, DR or Consistency measures in the oral and computerized forms of the WMT

	IR	DR	Consistency
No. of cases failing oral WMT:	12	11	13
(% of group failing)	(23%)	(21%)	(25%)
No. failing computerized WMT:	15	15	15
(% of group failing)	(28%)	(28%)	(28%)

Note. IR = Immediate Recognition trial, DR = Delayed Recognition trial, Consistency = consistency of responding from IR to DR.

Table 3.

Pearson product moment correlations between subtest scores on the computer and orally administered WMT.

	Computerized IR	Computerized DR	Consistency	Oral IR	Oral DR
Computerized DR	0.94				
Comp. Consistency	0.87	0.88			
Oral IR	0.83	0.86	0.70		
Oral DR	0.84	0.90	0.75	0.92	
Oral Consistency	0.78	0.81	0.80	0.88	0.89

Note. IR = Immediate Recognition trial, DR = Delayed Recognition trial. $p < 0.001$ for all values; Comp.= computerized

Table 4.

Means and SD on CARB and ability tests in those who passed or failed the WMT effort subtests

	Pass	Fail	Effect Size
CARB	97.7%(4.4)	82.8% (22.7)	0.48**
WMT-MC	85.7% (12.8)	43.6% (23.5)	0.76**
WMT-PA	78.7% (18.5)	43.1% (21.1)	0.67**
WMT-DFR	50.6% (15.4)	30.3% (19.1)	0.51**
WMT-LFDR	51.0% (18.3)	28.8% (18.7)	0.51**
CVLT trials 1 to 5	48.3 (11.2)	35.8 (9.6)	0.47**
CVLT SD Free	9.5 (3.3)	6.5 (2.9)	0.39*
CVLT LD Free	9.8 (3.3)	5.6 (3.9)	0.48**
CVLT Recog. Hits	13.6 (2.6)	10.4 (3.8)	0.44*
FSIQ	100.3 (14.4)	83.1 (13.4)	0.48**
VIQ	99.8 (15.0)	86.0 (13.6)	0.39*
PIQ	101.0 (14.9)	82.5 (13.4)	0.50**
CSRT Immediate	47.8 (9.3)	35.1 (10.9)	0.51**
CSRT Delayed	36.2 (13.6)	24.3 (10.6)	0.39*
Warrington Words	45.7 (4.7)	32.8 (8.1)	0.71***
Warrington Faces	41.9 (4.3)	33.1 (11.4)	0.50**
WCST Categories	4.8 (1.9)	3.6 (2.2)	0.27
Trails B	86.9 (92.1)	139.5 (74.2)	0.25

Note. Means, (standard deviations) and effect sizes for the Computerized Assessment of Response Bias (CARB), average oral and computerized Word Memory measures (MC = Multiple Choice, PA = Paired Associates, DFR = Delayed Free Recall, LDFR = Long-delayed Free Recall), California Verbal Learning Test (CVLT), IQ measures using the Multidimensional Aptitude Battery (MAB; Jackson, 1984) (FSIQ = Full Scale Intelligence Quotient, VIQ = Verbal Intelligence Quotient, PIQ = Performance Intelligence Quotient), Immediate and 30 minute delayed recall on CogniSyst's Story Recall Test (CSRT), Warrington's Recognition Memory Test for Words and Faces, Wisconsin Card Sorting Test (WCST) total categories achieved, and Trails B in 34 clinical compensation cases passing all of the WMT measures and 18 cases who failed any WMT effort measure. *$p < 0.005$; ** $p < 0.001$; ***$p < 0.0001$