McKinzey, R.K. (2011, April 21). The current accuracy rates of the Word Memory Test.*WebPsychEmpiricist*. Retrieved (date) from: http://www.wpe.info/papers_table.html

WPE WebPsychEmpiricist

The current accuracy rates of the Word Memory Test

4/21/11

R.K. McKinzey

Oakland, CA

Editor's Note: This article is expected to change as new studies become available. The date and citation is therefore also expected to change. Before using or citing this article, check for an update.

Abstract

The Word Memory Test (WMT) is a standalone measure of insufficient effort and malingering. This literature review specifies cross-validated accuracy rates, age ranges (7-adult), reading level (third grade), and a bibliography sufficient for clinical and forensic use of the test. The 10/1/03 & 10/20/03 updates change "in press" cites. The 11/1/03 update added two new cross-validations. The 4/30/05 update added one new cross-validation. The 4/4/07 update added two new cross-validations. The 10/15/07 update changed "in press" cites and corrects a N. The 5/7/10 update added two new cross-validations. The 5/17 update added two more cross-validations. The 11/17/10 update added two cites. This update adds another cite.

Correspondence concerning this article should be sent to: R.K. McKinzey, Ph.D., 389 30th St., Oakland, CA 94609. 510-655-3903, email: editor@wpe.info, or WPEList@yahoogroups.com

The current accuracy rates of the Word Memory Test

4/21/11

The Word Memory Test (WMT) is a verbal forced choice procedure designed to be a validity measure for tests of intellectual functioning.[1] It has proven quite sensitive to both deliberate efforts to malinger and mere lack of effort. However, the supporting research is moving rapidly, and any printed review will be out of date. This online review of accuracy rates can be kept current much more easily.

The WMT was first validated in proprietary studies described in a now out of print manual and two supplements (Allen & Green, 1999; Paul Green & Allen, 1999; Paul Green, Allen, & Astner, 1996). These studies found three scales to be useful: Immediate Recognition (IR), Delayed Recognition (DR), and Consistency. The same studies established specific cut scores, which are specified on the test forms.[2] Three separate decision rules have been used. The 2R rule is: The WMT is passed if both IR and DR are above the cut scores. The 3R rule is: The WMT is passed if IR, DR, and Consistency are above the same cut scores. The 3RM rule is: The WMT is passed if IR, DR, and Consistency are above slightly different cut scores. (In some studies, reports of the 3R rule will produce the same results for the 2R & 3RM rules.)

A sensitivity cross-validation (Paul Green, Lees-Haley, & Allen, 2002) used a convenience sample[3] of people referred for psychological or neuropsychological evaluations. 20 patients who had produced a normal WMT were asked to fake the test. All did so, producing a sensitivity of 100% (false negative rate of 0%), using the 3R rule (as noted above, the results would be the same for the 2R and 3RM rules).

An independent cross-validation (Tan, Slick, Strauss, & Hultsch, 2002) used an experimental sample of college students. 25 were asked to fake the WMT. The 3RM rule correctly identified all experimental malingerers. [4] Another independent cross-validation study (Shandera et al., 2010) asked 25 adult community volunteers with no more than 11 years of education to fake the Test of Malingered Memory (TOMM) and oral WMT. The WMT's false negative rate was 24%; the TOMM's was 56%. A third independent cross-validation study (Davis, Wall, Ramos, Whitney, & Barisa, 2010) asked 25 college students to fake the computerized WMT, resulting in a false negative rate of 21%.

The first published cross-validation of specificity (Iverson, Green, & Gervais, 1999) asked 38[5] normal volunteers to take the WMT. All passed (3R rule), yielding a specificity of 100% (false

---

[1] For a longer description of the WMT, see Allen & Green's paper at: http://wpe.info/papers_table.html. For definitions of test accuracy terms, see: http://wpe.info/2x2table.pdf.

[2] For these forms, write to: Paul Green, Ph.D., 17107-107 Ave., Ste. 201, Edmonton, Alberta, Canada T5S 1G3. 780-484-5550. paulgreen@shaw.ca http://www.wordmemorytest.com/

[3] Much of the WMT literature uses such convenience samples.

[4] In comparison, the TOMM had a 4% false positive rate and 20% false negative rate.

[5] Later expanded to 40.

positive rate 0%). In Tan's study, 27 students were asked to take the test normally[6]. All passed. In Davis' study, 25 control group students yielded the same results.

Another early study (Paul Green & Allen, 1999) asked 298 consecutive head injury patients to take the WMT. 64 had documented post-traumatic amnesia of at least one day, and 83% had abnormal MRI or CT results. 234 people had trivial or mild head trauma without post-traumatic amnesia (or less than a day). Their mean WMT score was significantly lower than the more severely injured sample.[7]

A second specificity cross-validation (Gervais et al., 2001) had 66 people undergoing evaluations for fibromyalgia or rheumatoid arthritis. All but 2 passed the WMT (specificity 97%, false positive rate 3%), using the 2R rule. The results were the same using the 3R rule (R.O. Gervais. personal communication, 3/6/03).

An interesting problem was exemplified when 29 adults took the WMT as part of a psychological evaluation (Paul Green & Flaro, 2003). All were trying to gain custody of their children, and so were presumed, a priori, to be motivated to do well. Two failed the WMT. When these 2 were confronted, they admitted they had little involvement in the evaluation and hadn't tried very hard. They weren't trying to malinger, they just hadn't exerted any effort during this very simple intellectual task. When the N was increased to 118, the false positive rate fell to 2% (Flaro, Green, & Robertson, 2007).

Further evidence for the importance of effort (Paul Green, Rohling, Lees-Haley, & Allen, 2001) came in a study which had 904 patients undergoing comprehensive neuropsychological evaluations take the WMT. All were in litigation. Not surprisingly, 28.5% failed (3R rule). However, of more importance was the finding that "effort explained 53% of the variance", far more than age (4%) or years of education (11%)! That is, the magnitude of the WMT scores predicted the magnitude of the neuropsychological test scores. Again, some of the patients weren't necessarily malingering (although some no doubt were), they just weren't trying very hard on either the WMT or the tests of intellectual ability. However, this finding has not replicated (Bowden, Shores, & Mathias, 2006), but see (Flaro et al., 2007), perhaps because of methodological failings (Rohling & Demakis, 2010).

The WMT has been compared to the TOMM and Computerized Assessment of Response Bias (CARB) (Gervais, Rohling, Green, & Ford, 2004). A convenience sample of 519 adults were given the WMT (using the 3R rule), TOMM, and CARB as part of their psychological evaluations. None had any head injury, but 326 (63%) were in litigation. The rest were being evaluated for potentially receiving vocational retraining benefits. Of the former group, 43% failed the WMT, 25% failed the CARB, and 17% failed the TOMM. Of the latter group, 12% failed the WMT, 4% failed the CARB, and 1% failed the TOMM. When the groups were combined, those who failed any of the tests had overall lower scores on ability tests. A clear trend was found, with those failing more tests getting

[6] One other student did badly on all the measures and was dropped from the analysis.
[7] Accuracy rates are not yet available.

progressively worse ability scores. The authors argue that the 12% are not false positives. Rather, they assert, the WMT is more sensitive than the TOMM or CARB.

The WMT has been extended to children (Paul Green et al., 2002). 135 children (ages 7-18) were evaluated for various reasons. None had any reason to malinger. Using the adult cutoffs and decision rule, 19 failed (specificity 86%, false positive rate 14%). There was no age or Verbal IQ effect, even for IQs less than 70. Since the computerized version of the WMT was used, there was a reading level effect, especially with those children with less than a third grade reading level. Six of the failing children were asked to take the WMT again, and pass it. Five did. The other got much lower scores. Taking the five children out of the analysis improved the false positive rate to 11%. The simplicity of the test was confirmed: "Children with pervasive developmental disorder, bipolar mood disorder, fetal alcohol syndrome and other serious disorders were very young, had a very low mean VIQ and displayed a low mean reading level. Nevertheless, they obtained nearly perfect scores on the WMT." A separate study (Courtney, Dinkins, Allen, & Kuroski, 2003) of 111 referred children confirmed the importance of reading level when using the computerized WMT, finding (using a subsample) that reading accounted for 34% of the WMT variance. Accuracy rates were not calculated.

Green (P. Green, Flaro, & Courtney, 2009) again reported on a sample of children. 350 children with a variety of psychiatric diagnoses described as having "developmental disorders" were given the WMT as part of a clinical evaluation. 10% failed. In an effort to further reduce the false positive rate, Green used an added decision rule, "profile analysis," which uses three additional subtests: Multiple choice recognition (MC), Paired associates (PA), and Free recall (FR). This decision rule looks for a "Genuine Memory Impairment Profile" (GMIP), which is defined as "at least a 30 point difference between the mean of the easy and hard subtests (IR, DR, and CNS versus MC, PA, and FR). Of the 10% of children failing the easy subtests, half had a GMIP, reducing the false positive rate to 5%. Using a subset of 42 children who had impaired California Verbal Learning Test scores, a 31% false positive rate using the 3RM rule was reduced to 5% using profile analysis. A subset of children with FSIQs below 71 yielded a 3RM false positive rate of 32%. Using profile analysis reduced the failure rate to 7%.

Normal adult experimental subjects (Loring et al., in press) were used to study the effects of lorazepam (Ativan), a benzodiazapine. The WMT was given to drugged subjects and then again when they were not drugged. When not drugged, the false positive rate was 4%. When drugged, the false positive rate increased to 21%. Using profile analysis, the rate dropped to 11%.

For the first time, a cross-validation study (Greve, Ord, Curtis, Bianchini, & Brennan, 2008) used a "known groups" design. In this design, a convenience sample of 43 brain injured and 42 chronic pain patients took the Portland Digit Recognition Test (PDRT), TOMM, and computerized WMT using standard instructions. The two patient groups were further separated into Suspected or

Unsuspected Malingering groups. The group criteria was a) known presence of malingering incentive and b) the failure of various combinations of within-test measures: Reliable Digit Span, WAIS-III, Meyers Index, and the Fake Bad Scale. This separation resulted in chronic pain patients yielding a false negative rate of 43%[8], and an false positive rate of 10%. The brain injured patients yielded a false negative rate of 15% and a false positive rate of 30%[9]. Whether these results replicate, or are found to be an artifact of their within-test vs. standalone test design remains to be seen. Also unknown is whether profile analysis will reduce the WMT false positive rate.

The WMT has been used with a sample of Germans, in another independent experimental cross-validation (Brockhaus & Merten, 2004). The control sample (n=27) and experimental simulators (n=100) were all correctly identified by the WMT. More importantly for U.S. psychologists (R K McKinzey, 2003, September 20), a group of 32 adults with mental retardation (MR) yielded a false positive rate of 3%.

All of the above studies used a computerized version of the WMT. A *WebPsychEmpiricist* study (Allen & Green, 2002, September 20) extended the literature to an oral, paper & pencil version. In another convenience sample, 52 adult patients undergoing neuropsychological evaluations were asked to take the WMT in both the computerized and oral versions. Only minute, nonsignificant differences were found. Practice effects were also not found. Failure on the WMT (either oral or computerized) was unrelated to age, sex, years of education, handedness, or English as a second language. The results also supported the effort studies: "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." (p. 9) Another study (Hoskins, Binder, Chaytor, Williamson, & Drane, 2010) also found no significant differences between versions.

The TOMM and oral WMT was given to an adult MR sample (N=24), this time in the US (Shandera et al., 2010). The TOMM had a false positive rate of 12%. The WMT had a false positive rate of 75% (Shandera, A.L., personal communication, May 5, 2010). The 3RM rule was used, not profile analysis.

A review article (Paul Green et al., 2002) discusses the issues involved in bringing the WMT into the courtroom, including admissibility standards such as Daubert. An independent review was laudatory (Hartman, 2002).

Unlike many other tests of effort/malingering, the WMT may not be affected by coaching (Dunn, Shear, Howe, & Ris, 2003). However, in Davis' study, coaching increased the false negative rate from 21% to 28%.

Elsewhere (R. K. McKinzey, 1997), I have suggested that the critical question for neuropsychological malingering tests is, "What are the cross-validated false negative and positive

---

[8] PDRT false negative rate =64%, false positive rate = 2%. TOMM false negative rate =60%, false positive rate =2%.

rates?" For the WMT, the answer is: "Unless used with severely demented or MR patients, the WMT's cross-validated false negative rate is 0-24%. The false positive rate in adults and children is 0-12%." The addition of profile analysis may lower the false positive rates in severely demented or MR patients.

References

Some of these references can be obtained via: http://wpe.info/reprints_available.html

Allen, L., & Green, P. (1999). *Severe TBI sample performance on the Word Memory Test and Computerized Assessment of Response Bias*. Durham, NC: CogniSyst.

Allen, L., & Green, P. (2002, September 20). Equivalence of the computerized and orally administered Word Memory Test effort measures. *WebPsychEmpiricist*, *2002*. Retrieved September 20, 2002 from http://wpe.info/papers_table.html.

Bowden, S. C., Shores, E. A., & Mathias, J. L. (2006). Does effort suppress cognition after traumatic brain injury? A re-examination of the evidence for the Word Memory Test. *Clinical Neuropsychologist, 20*(4), 858.

Brockhaus, R., & Merten, T. (2004). Neuropsychologische Diagnostik suboptimalen Leistungsverhaltens mit dem Word Memory Test [Neuropsychological Assessment of Suboptimal Performance: The Word Memory Test]. *Nervenarzt, 75*(9), 882-887.

Courtney, J., Dinkins, J., Allen, L. M., & Kuroski, K. (2003). Age related effects in children taking the Computerized Assessment of Response Bias and Word Memory Test. *Child Neuropsychology, 9*(2), 109-116.

Davis, J., Wall, J., Ramos, C., Whitney, K., & Barisa, M. (2010). Using Grip Strength Force Curves to detect simulation: A preliminary investigation. *Archives of Clinical Neuropsychology, 25*(3), 204-211.

Dunn, T. M., Shear, P. K., Howe, S., & Ris, M. D. (2003). Detecting neuropsychological malingering: Effects of coaching and information. *Archives of Clinical Neuropsychology, 18*(2), 121-134.

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury, 21*(4), 373-383.

Gervais, R., Rohling, M., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19*(4), 475-487.

Gervais, R., Russell, A., Green, P., Allen, L., Ferrari, R., & Pieschl, S. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology, 28*(8), 1892-1899.

---

[9] PDRT false negative rate =48%, false positive rate = 3%. TOMM false negative rate =44%, false positive rate =2%.

Green, P., & Allen, L. (1999). *Performance of neurological patients on the Word Memory Test and Computerized Assessment of Response Bias*. Durham, NC: CogniSyst.

Green, P., Allen, L., & Astner, K. (1996). *Manual for Computerized Word Memory Test*. Durham, NC: CogniSyst.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9*(3), 189-207.

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury, 23*(9), 741-750.

Green, P., Lees-Haley, P. R., & Allen, L. M. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 2*, 97-124.

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

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist, 22*(5), 896-918.

Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing. Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology, 17*(7), 709-714.

Hoskins, L., Binder, L., Chaytor, N., Williamson, D. J., & Drane, D. (2010). Comparison of Oral and Computerized versions of the Word Memory Test. *Archives of Clinical Neuropsychology, 25*(7), 591-600.

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.

Loring, D., Marino, S., Drane, D., Parfitt, D., Finney, G., & Measor, K. (in press). Lorazepam effects on Word Memory Test performance: A randomized, double-blind, placebo controlled, cross-over trial. *The Clinical Neuropsychologist*.

McKinzey, R. K. (1997). The Cross-examination of neuropsychologists: Countering the claim of brain damage. *Prosecutor's Brief, 19*(2), 13-20.

McKinzey, R. K. (2003, September 20). Too dumb to die: Mental retardation meets the death penalty. *WebPsychEmpiricist*, *2003*. Retrieved September 20, 2003 from http://wpe.info/papers_table.html.

Rohling, M. L., & Demakis, G. J. (2010). Bowden, Shores, & Mathias (2006): Failure to replicate or just failure to notice. Does effort still account for more variance in neuropsychological test scores than TBI severity? *The Clinical Neuropsychologist, 24*(1), 119-136.

Shandera, A. L., Berry, D. T. R., Clark, J. A., Schipper, L. J., Graue, L. O., & Harp, J. P. (2010). Detection of malingered mental retardation. *Psychological Assessment, 22*(1), 50-56.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *Clinical Neuropsychologist, 16*(4), 495-505.