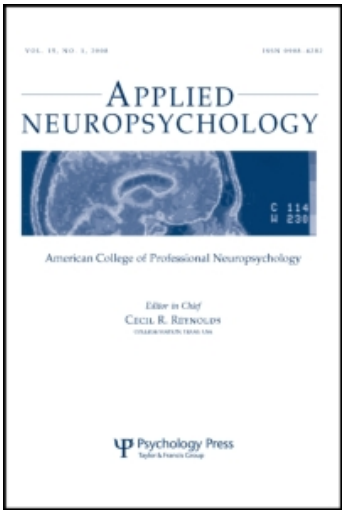

Applied Neuropsychology

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775648089

Memory Malingering: Evaluating WMT Criteria

Maurício Martins[a]; Isabel P. Martins[a]
[a] Language Research Laboratory, Neurological Clinical Research Unit, Institute of Molecular Medicine, Lisbon Faculty of Medicine, Lisboa, Portugal

Online publication date: 26 August 2010

To cite this Article Martins, Maurício and Martins, Isabel P.(2010) 'Memory Malingering: Evaluating WMT Criteria', Applied Neuropsychology, 17: 3, 177 — 182
To link to this Article: DOI: 10.1080/09084281003715709
URL: http://dx.doi.org/10.1080/09084281003715709

APPLIED NEUROPSYCHOLOGY, *17*: 177–182, 2010

ISSN: 0908-4282 print/1532-4826 online
DOI: 10.1080/09084281003715709

Psychology Press
Taylor & Francis Group

Memory Malingering: Evaluating WMT Criteria

Maurício Martins and Isabel P. Martins

Language Research Laboratory, Neurological Clinical Research Unit, Institute of Molecular Medicine, Lisbon Faculty of Medicine, Lisboa, Portugal

The Word Memory Test (WMT) has been considered one of the best assessment tools for assessing memory malingering in spite of some reported false-positive results. Our goal was to examine the sensitivity and specificity of Green's 2003 criteria for memory malingering using a Portuguese adaptation of the WMT. We applied the WMT to three groups of participants, including 20 healthy subjects asked to simulate memory impairment, 29 healthy adults asked to produce their best performance, and 21 patients with the diagnosis of mild cognitive impairment (MCI). Using Green's 2003 criteria, based on scores below cutoffs on the easy subtests, 67% of the MCI patients were classified as "poor effort." However, the sensitivity and specificity of the diagnosis could be raised to 95% using an alternative set of criteria, relying on comparisons between easy and hard subtest scores. We conclude that Green's original criteria based on easy subtest scores alone seem to be of low specificity for the diagnosis of memory impairment simulation, but the WMT is a good instrument for identifying simulation if the alternative criteria are applied.

Key words: criteria, malingering, memory, WMT

INTRODUCTION

Memory impairment is a frequent complaint of healthy people, and its frequency tends to increase with age. However, this symptom can often be found in people suffering from mild traumatic brain injury, in people making retirement claims, and in those involved in litigation where there are secondary gains (Donders & Boonstra, 2007; Flaro, Green, & Robertson, 2007; Green, Iverson, & Allen, 1999; Green, Rohling, Lees-Haley, & Allen, 2001). Neuropsychological evaluation is considered an objective way of measuring cognitive abilities and brain functional integrity. Yet, on the majority of neuropsychological tasks, it might be difficult to distinguish real memory impairment from motivation deficits or exaggeration of symptoms. For instance, although mild head injury patients present cognitive deficits during the early recovery phase, it is rare that these deficits persist for a long time without predisposing factors (social, psychiatric) or legal posttraumatic processes (Donders & Boonstra, 2007). For this reason, the inclusion of tasks able to measure effort and motivation has been recommended as part of the neuropsychological evaluation of those patients (Bush et al., 2005; Iverson, 2006).

Several tests have been developed to measure effort, and among them, the Word Memory Test (WMT) has been considered one of the most sensitive, not only in North America (Bauer, O'Bryant, Lynch, McCaffrey, & Fisher, 2007; Green, Lees-Haley, & Allen, 2002) but also in the German population (Brockhaus & Merten, 2004). When applied to samples of subjects claiming disability with mild head injury or with legal processes, WMT actually explains most of the variance found on psychometric tasks (Green, 2007; Green et al., 2001; Stevens, Friedel, Mehren, & Merten, 2008; Wynkoop & Denney, 2005). Also, it has been positively reviewed (Hartman, 2002; Wynkoop and Denney).

One of the problems associated with these types of tests is that they can hardly differentiate between

Address correspondence to Maurício Martins, Laboratório de Estudos da Linguagem, Faculdade de Medicina de Lisboa, Lisboa 1649-028. E-mail: labling@fm.ul.pt

motivation and intentionality. This means that a "poor effort" result can be either a consequence of a motivation deficit or of a voluntary intention to present low neurocognitive performance (Donders & Boonstra, 2007; Merten, Bossink, & Schmand, 2007; van Beilen, van Zomeren, van den Bosch, Withaar, & Bouma, 2005). In older subjects, or in people with psychiatric comorbidity as in depression or schizophrenia (Gorissen, Sanz, & Schmand, 2005; van Beilen et al.), "poor effort" can be explained due to motivation impairment related to the underlying condition and not necessarily by "memory malingering."

Because of these pitfalls, it becomes essential to understand how we can use these new tools in clinical practice. Our goals were: (1) to examine the internal consistency of the WMT Portuguese version and (2) to test the sensitivity and specificity of Green's 2003 and Genuine Memory Impairment Profile (GMIP) criteria in the Portuguese population.

METHODS

Participants

A computer version of the WMT 1.0.1 was applied to a sample of 70 subjects: 49 healthy controls and 21 patients with mild cognitive impairment (MCI).

Healthy volunteers were recruited among university students and families of patients attending a university hospital. These participants were randomly divided into two groups: (a) a "simulator" group, in which people were asked to behave as if they were trying to obtain anticipated retirement for medical reasons ($N = 20$) and (b) a "good effort" group, who were asked to give their best effort performing the WMT ($N = 29$). Participants received a closed envelope with instructions. The observer was "blind" to group allocation. Participants did not receive any monetary reward. None of the healthy participants had any identifiable secondary gain, history of alcohol or drug abuse, or systemic, psychiatric, or neurological disease that might have influenced test performance.

The clinical group was composed of patients with memory complaints, fulfilling the Petersen criteria for the diagnosis of amnesic MCI (Petersen et al., 2001): (1) memory complaints (corroborated by an informant); (2) objective memory impairment on the Logical Memory subtest of the Wechsler Memory Scales-III (Portuguese version validated by Garcia, 1984), scoring 1 standard deviation below normal for age and education on delayed recall; (3) maintained activities of daily living; the patient should be judged to be managing their professional, social, and familial activities, and have no or only mild impairment in the Instrumental Activities of Daily Living (IADL) scale (Botelho, 2000; Lawton & Brody, 1969). That is to say, no more than one item from the IADL scale suffered any changes (Pantoni et al., 2005). The patients had no other significant cognitive decline on clinical or neuropsychological assessment.

None of these patients had any identifiable secondary gain. All patients were retired and without ongoing legal processes. They did not have any other clinical deficit that might have influenced test performance, namely: other neurological disorder suggested by clinical history or imaging or laboratory tests and presence of psychiatric disorders, according to the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, criteria (American Psychiatric Association, 1994). Subjects with any other condition with possible impact on cognition, such as systemic disease or alcohol or drug abuse, were also excluded. Demographic characterization of the population is detailed in Table 1.

This project was approved by the local ethics committee, and all participants signed informed consent.

Instruments and Statistics

Translation and back translation. The WMT 1.0.1 was translated by the author and back translated by an external observer. The two versions were compared, and a consensus version was elaborated. Translation was performed with permission from the original author of the WMT and inserted in the original computer version by the Green's Publishing programmers. The translation remains copyright of the original owner (Green).

Neuropsychological assessment. Both MCI and healthy subjects had a detailed neuropsychological assessment using the Battery of Lisbon for the Assessment of Dementia (BLAD; Garcia, 1984). The BLAD includes tests within the following cognitive domains: Attention, Semantic Fluency, Motor and Graphomotor Initiatives,

TABLE 1
Sample Characterization

	Group		
	"Good Effort"	*"Simulator"*	*MCI*
N	29	20	21
Age ± *SD*	31.7 ± 2.8	24.0 ± 0.51	71.2 ± 2.0
Sex M:F	1:2.2	1:1.9	0.92:1
Education			
Less than 6 years	10.3%	.0%	71.4%
7–11 years	6.9%	.0%	19.0%
12 years	6.9%	.0%	.0%
University	72.3%	95.0%	9.5%
Masters	3.4%	5.0%	.0%

Note. N = Number of subjects; SD = standard deviation; M = Male; F = Female.

Object Naming, Verbal Comprehension, Orientation, Verbal and Nonverbal Abstraction, Visuoconstructional Abilities, Calculation, Short- and Long-Term Memory, and Learning.

WMT application. The WMT is divided into five subtests: The first two subtests (Immediate Recognition [IR] and Delayed Recognition [DR]) measure effort, and the last three tasks measure verbal memory (Multiple Choice [MC], Paired Associate Recall [PA] and Free Recall [FR]). All subtests are based on a list of 20 pairs of words, presented twice on a computer screen at the beginning of the test. In the IR subtest, the subject is presented with 40 pairs of words and is asked to choose (by pointing or saying out loud) which word of the pair he has seen previously on the screen. For each pair, there is only one correct answer. This task is repeated 30 minutes later, maintaining the correct words of each pair with different foils. This latter subtest DR. Between DR and IR, a Consistency parameter (CNS) is calculated automatically by the software.

These first two subtests—IR and DR—appear to measure verbal memory when, in fact, they are relatively insensitive to impairment of memory. They are considered mainly to measure "effort." They are very easy subtests, based on a forced-choice principle with a baseline 50% chance of scoring correctly when choosing at random. Besides "effort" subtests, the WMT also contains three verbal memory subtests presented sequentially and in an increasing order of difficulty. All these subtests are based on the 20 word pairs presented at the beginning of the WMT (the "original list"): (1) In the MC task, one word of the "original list" is presented on the screen. The subject is then asked to select among eight given possibilities, which word was paired with it in the "original list." (2) On the PA subtest, the person is told one word from the "original list" and is asked to recall its pair without external cues. (3) The last subtest requires the FR of all words from the original list.

Taken together, "effort" and "verbal memory" subtests, by their sequentially increasing difficulty, allow us to build a response profile that may be more or less suggestive of a "simulator" pattern. The original criteria proposed by Green (2003) suggested that subjects scoring below 82.5% on any of the "effort" subtests (IR, DR, CNS) should be classified as "poor effort" and therefore considered possible simulators of memory impairment.

However, in recent years, the WMT computer program has incorporated an analysis of the profile of results, featuring an explicit comparison between the mean score on the easy versus the hard subtests. These new criteria classify as "simulators" subjects who meet Green's (2003) criteria and whose difference between the mean of the easy (IR, DR, CNS) and hard (MC, PA, FR) subtests scores below 30.

Statistics. In this work, we calculated the sensitivity and specificity of Green's criteria for the diagnosis of "memory malingering" both between the two control groups and between "simulators" and "MCI" (in a sample of the Portuguese population). We also calculated the internal consistency of the Portuguese version. Because of gender distribution differences between different groups, we also tested for gender influence on WMT performance using an independent sample *t*-test ($p < .05$). All statistics were performed using Statistical Package for the Social Sciences 14.0 for Windows.

RESULTS

WMT Portuguese Version

The internal consistency of the Portuguese version was .949 (Cronbach's alpha) considering all the subjects. Because word usage frequency could influence word recall, we tested whether the translation maintained a similar distribution of word frequency among correct responses in IR and DR subtests. We collected data about word frequency on the "Corpus of Contemporary American English" for American English words and on "Corpus de Referência do Português Contemporâneo" for Portuguese words. There were no significant differences ($x^2 = .02$, *ns*) in the distribution of higher frequency words among correct choices in the two versions (Portuguese and American English) of WMT (Bacelar do Nascimento, 2007; Davies, 1990).

Using an independent samples t-test with a significance criterion of $p < .05$, we found no difference between male and female genders in the performance of any of the WMT subtests.

Sensitivity and Specificity

The average results of the different groups in the several WMT subtests are presented in Table 2.

When we used the original "poor effort" Green (2003) criteria to distinguish between "simulators" and "good effort" (i.e., IR, DR, or CNS $\leq$82.5), we found that the test had 95% sensitivity and 100% specificity. However, when we applied the same criteria to the "MCI" group, 67% of the patients scored below the "poor effort" cutoff. This means that, if we were using a "poor effort" classification based only on the easy subtest scores to diagnose "memory malingering," the WMT would be unsuitable for clinical practice because of the high rate of false positives. However, when we analyzed the data, we could clearly see that there were empirical ways to distinguish between "MCI" and "simulator" groups based on the response profile. Therefore, we explored other criteria to maximize the specificity and sensitivity when contrasting "simulators"

TABLE 2
Average Results of the Different Groups in the Several WMT Subtests

	WMT Subtest					
Group	*IR*	*DR*	*CNS*	*MC*	*PA*	*FR*
"Good effort" (mean ± *SD*)	97.3 ± 4.3	99.1 ± 3.4	97.1 ± 5.0	90.5 ± 14.0	85.3 ± 21.8	72.3 ± 16.8
"Simulator" (mean ± *SD*)	45.9 ± 23.0	47.4 ± 21.8	60.0 ± 14.8	37.0 ± 18.6	32.8 ± 21.3	33.4 ± 17.2
MCI (mean ± *SD*)	75.8 ± 14.4	74.6 ± 17.3	71.8 ± 16.3	36.0 ± 17.4	22.2 ± 15.0	18.5 ± 15.7

Note. *N* = Number of subjects; *SD* = Standard Deviation; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; MC = Multiple Choice; PA = Pair-Associate; FR = Free Recall.

with "MCI" subjects. When we used the following criteria for the diagnosis of "simulator":

IR–FR < 40 or (IR < 60 and DR < 60)

We obtained a sensitivity of 95% in identifying the simulators and a specificity of 95.2% in identifying the MCI cases, who were presumed to be performing at an optimal level. This means that a subject with suboptimal performance (according to Green [2003] criteria) is not a probable simulator if he satisfied both of the following conditions: (1) difference between IR and FR scores higher than 40 (IR–FR > 40) and (2) at least one of the "effort" subtest scores—IR and DR—is higher than 60 (IR > 60 and DR > 60).

GMIP Criteria

Recently, after we had finished the first draft of this article, we became aware of a newer version of the WMT (1.08). This version also proposed new criteria to distinguish between "simulation" and "genuine memory impairment." These criteria, called the GMIP take into consideration the difference between the "easy" subtests and the "hard" subtests. The algorithm classifies as "simulators" subjects who satisfy both conditions:

A. IR, DR, or CNS <82.5% (Green 2003 criteria);

B. [(IR+DR+CNS)/3]−[(MC+PA+FR)/3] < 30 (difference between the mean of the easy and hard subtests below 30).

Applying these criteria retrospectively to our samples, we found that the new criteria proposed by Green (as presented in the newer WMT version) had a sensitivity of 85% and specificity of 95.2% in the diagnosis of "simulator."

Comparing our criteria with GMIP criteria, both have similar specificity although our criteria seem to have higher sensitivity.

DISCUSSION

The development and application of tests measuring "effort" in neuropsychological evaluation are fundamental because "effort" has an important influence on neuropsychological performance and therefore may bias the correct interpretation of testing results. The main problem with current symptom validity tests is that they do not discriminate well between motivation/attention deficits and intentional fabrication of symptoms. In spite of being proposed as a sensitive test for memory malingering, the WMT has the same problem if only the easy subtests are considered. This issue of specificity (Merten et al., 2007) has also been raised when the WMT is applied to subjects with psychiatric illness (Gorissen et al., 2005; Morel, 2007; Sullivan, May, & Galbally, 2007; van Beilen et al., 2005) and to patients with severe memory impairment (Merten et al., 2007). The rationale behind Green's original criteria for "poor effort" is that IR and DR are tasks which are so easy that most people with cognitive impairment could perform above 82.5%. (Green, 2003). However, as pointed out in a functional magnetic resonance imaging study (Allen, Bigler, Larsen, Goodrich-Hunsaker, & Hopkins, 2007), brain regions associated with cognitive processing and motivation (like dorso-lateral prefrontal cortex, anterior cingulated cortex, anterior insula, and superior parietal cortex) are activated during IR and DR execution. This raises issues about the loss of integrity of these networks in patients submitted to WMT. If these neuronal circuits are necessary to execute WMT, then their dysfunction could explain "poor effort" results in groups of people with schizophrenia or Alzheimer's disease and possibly some head injuries.

In our study, we found the same problem as other studies (Merten et al., 2007). When we relied on the cutoffs based only on the easy subtests, many of the patients with MCI were classified as "poor effort." Although cultural and translation factors could partially explain our findings, results similar to ours have been described in other studies. Therefore, it is probable that using the original Green "poor effort" criteria alone is not enough to identify poor effort when testing subjects with real memory impairment.

Some authors (Merten et al., 2007) have suggested lower cutoff values when testing subjects with severe memory impairment, and after analyzing our own data,

we share that same opinion. However, what we propose for the Portuguese population goes a little further: Besides analyzing IR and DR results, we think it is useful if, as suggested by Green (2003), an analysis is performed of the response profile, comparing the easy subtest scores with those on the more difficult subtests (MC, PA, FR). Therefore, we support the inclusion of an additional quantitative criterion that measures the *horizontality* of the profiles' curve. The rationale behind this proposal is that individuals voluntarily producing symptoms tend to approach both easy and hard subtests with the same strategy. As a result, the difference between the easiest subtest (IR) and the hardest subtest (FR) tends to be lower in subjects who are intentionally producing their symptoms.

Although the generalization and validity of our criteria is still very restricted, they could be a useful guide for a broader search of specific ''memory malingering'' criteria using WMT, for Portuguese and for other populations. More studies are needed in different countries with different languages to assess which criteria would best suit each population in the goal of overcoming the specificity problem which arises if only the IR and DR subtests are employed. It may be noted finally that the current WMT program in the report section issues a warning that it is not advisable to use only the IR and DR subtests and that all subtests need to be administered so that the profile of results can be analyzed.

Limitations

We acknowledge several limitations in our study, mainly concerning small sample size and external validity. There are demographic differences between our samples, which were not matched for sex, age, or education. However, gender had no impact on WMT performance, and with regard to age and education, there are several reasons to justify the specific features of our sample selection. The ''MCI'' group was composed of retired subjects with memory complaints, providing the best guarantee of minimizing ''retirement claim'' as a secondary gain. This group was necessarily composed of subjects around the age of 65. If we had opted for a similar sample in the ''simulator'' group, we would be taking the risk of testing individuals with some degree of cognitive impairment, making data interpretation more ambiguous. Therefore, we opted to test younger subjects, capable of fully understanding the task and capable of executing it (something we found particularly hard in older subjects).

In spite of all these problems, we believe that our results are ecologically valid, because the studied samples represent the best trade-off between choosing the best possible ''simulators'' on one hand and testing people with memory impairment without secondary gains on the other.

ACKNOWLEDGEMENTS

The authors acknowledge all volunteers who participated in this study, the research team of the Language Research Laboratory, and Gabinete de Apoio à Investigação Científica of Medicine Faculty of Lisbon, which funded this project.

REFERENCES

Allen, M. D., Bigler, E. D., Larsen, J., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2007). Functional neuroimaging evidence for high cognitive effort on the Word Memory Test in the absence of external incentives. *Brain Injury*, *21*, 1425–1428.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bacelar do Nascimento, M. F., et al. (2007). *CRPC-Corpus de Referência do Português Contemporâneo*. Retrieved September 2007, from http://www.clul.ul.pt/sectores/linguistica_de_corpus_projecto_crpc.php

Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2007). Examining the Test of Memory Malingering Trial 1 and Word Memory Test Immediate Recognition as screening tools for insufficient effort. *Assessment*, *14*, 215–222.

Botelho, M. A. (2000). *Autonomia funcional em idosos* [Functional autonomy in the elderly]. Porto, Portugal: Laboratórios Bial.

Brockhaus, R., & Merten, T. (2004). Neuropsychological assessment of suboptimal performance: The Word Memory Test. *Nervenarzt*, *75*, 882–887.

Bush, S., Ruff, R., Troster, A., Barth, J., Koffler, S., Pliskin, N., et al. (2005). Symptom validity assessment: Practice issues and medical necessity (NAN Policy and Planning Committee). *Archives of Clinical Neuropsychology*, *20*, 419–426.

Davies, Mark. (1990). *Corpus of contemporary American English*. Retrieved September 2007, from http://www.americancorpus.org

Donders, J., & Boonstra, T. (2007). Correlates of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, *21*, 319–326.

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*, 373–383.

Garcia, C. (1984). *Alzheimer's disease: Difficulties in clinical diagnosis*. Unpublished doctoral dissertation, University of Lisbon.

Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research*, *78*, 199–208.

Green, P. (2003). *Manual for the Word Memory Test for Windows*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America*, *18*, 43–68.

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

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. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*, 1045–1060.

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

Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology*, *13*, 77–90.

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist*, *9*, 179–186.

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, *29*, 308–318.

Morel, K. R. (2008). Comparison of the Morel Emotional Numbing Test for posttraumatic stress disorder to the Word Memory Test in neuropsychological evaluations. *The Clinical Neuropsychologist*, *22*, 350–362.

Pantoni, L., Basile, A. M., Pracucci, J., Asplund, K., Bogousslavsky, J., Chabriat, H., et al. on behalf of the Ladis study group. (2005). Impact of age-related cerebral white matter changes on the transition to disability—The Ladis Study: Rationale, design, and methodology. *Neuroepidemiologia*, *24*, 51–62.

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., et al. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, *58*, 1985–1992.

Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, *157*, 191–200.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit/hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology*, *14*, 189–207.

van Beilen, M., van Zomeren, E. H., van den Bosch, R. J., Withaar, F. K., & Bouma, A. (2005). Measuring the executive functions in schizophrenia: The voluntary allocation of effort. *Journal of Psychiatric Research*, *39*, 585–593.

Wynkoop, T. F., & Denney, R. L. (2005). Test review: Green's Word Memory Test (WMT) for Windows. *Journal of Forensic Neuropsychology*, *4*, 101–105.