Routledge
Taylor & Francis Group

Misleading conclusions about word memory test results in multiple sclerosis (MS) by Loring and Goldstein (2019)

Christopher Graver[a] and Paul Green[b]

[a]Madigan Army Medical Center, MCHJ-CLU-CP, Tacoma, WA, USA; [b]Green's Publishing, Kelowna, BC, Canada

ABSTRACT

Loring and Goldstein presented a case of a woman with Multiple Sclerosis (MS) who failed the traditional performance validity criteria of the WMT. Scoring lower than the mean from patients with Alzheimer's Disease on extremely easy subtests, the patient carried on to produce a WMT profile which is typical of someone with invalid test results, based on the usual interpretation, which is standardized within the Advanced Interpretation Program. Statements were made that are incorrect, including the claim there are no available data on the WMT in MS patients, that the minor tranquilizer Lorazepam can explain WMT failure even in healthy adults and that this patient produced a neuropsychological profile that is credible and typical of MS. We report data from MS patients given comprehensive neuropsychological assessment, including the WMT. Loring and Goldstein's interpretation of this case does not fit the facts.

Introduction

Accepting that Performance Validity Testing (PVT) is necessary to ensure that neuropsychological test results are valid, Loring and Goldstein (2019) tested a woman with a two year old diagnosis of MS, whom we shall label as the "the MS case" below. She was complaining of numbness and tingling in her fingers and toes, with persistent left sided paresthesia and right sided foot drop. There was no doubt about the diagnosis, which was based on multiple objective findings, such as lesions on CT scans of the brain and spine. Two assessments were performed.

PVT failure during the first assessment

In the first assessment, there was failure on the primary PVT measure of the Victoria Symptom Validity Test (Slick et al., 1997). The MS case's score on the "easy" items was reported as valid (20/24), whereas the "difficult" item score was at a chance level (12/24) and "raised concern" about the validity of the cognitive test results. The authors further stated " ... possible lack of engagement or disinterest could not be accurately determined" and considered psychiatric factors as contributing to the PVT results. Testing was discontinued, pending further psychiatric support.

There are two points to consider with such an interpretation. Loring and Goldstein raise the concern several times in their paper regarding the need to interpret PVTs in the light of data from relevant clinical comparison groups. Multiple studies have demonstrated that the MS case's

scores were far below neurologically impaired groups and consistent with invalid performance. For example, a study of profoundly amnesic patients, including those who had hippocampectomy, showed that none of the participants scored below 24 on the "easy" items or below 22 on the "difficult" items (Slick et al., 2003). In addition, her "easy" score was 1.7 SD below and her "difficult" score was 3.6 SD below patients with acute severe TBI who required extended multidisciplinary inpatient rehabilitation (Macciocchi et al., 2006).Whereas the MS case had a total VSVT score of 32, only 4% of the acute severe TBI patients in that study scored below 44 and none scored below 38, highlighting the very high probability of invalid data. Jones (2013) presented VSVT data from a mixed clinical sample that included multiple sclerosis. The MS case's "easy" score was 5.46 SD below the good effort clinical group, and that score was associated with 0.98 specificity for invalid performance. The "difficult" score was 6.1 SD below the good effort clinical group and was associated with 1.0 specificity for invalid performance (zero false positives). The conclusion from that first testing of the MS case then should be that her data were firmly invalid.

PVT failure during the second assessment

A second assessment employed the WMT and her results are shown in Figure 1, using a chart created by the AI program (Green, 2008). The MS case failed the traditional
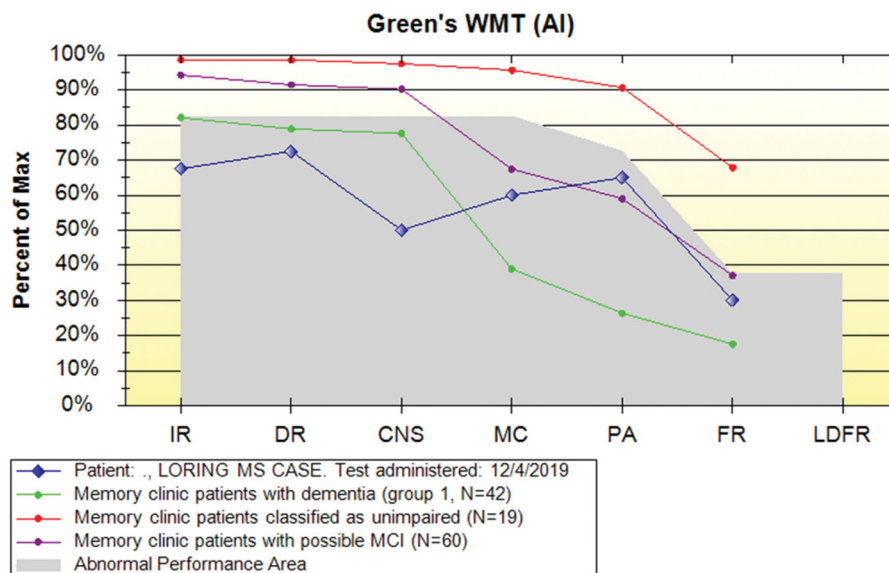
Figure 1. WMT results from the MS case contrasted with mean group data from people with dementia from Green et al. (2011).

performance validity criteria of the WMT (Green & Astner, 1995; Green, Allen, & Astner, 1996; Green, 2003).

The first two subtests were incorrectly labeled "immediate recall" and "delayed recall" by the authors (page 3). The correct terms would be immediate and delayed "recognition" (IR and DR). The distinction is important because recognition is an ability which is very resistant to impairment, whereas recall of verbal information is sensitive to actual impairment. For example, see the summary of the basic validity studies of the WMT in Green et al. (2003) and Erdodi et al. (2019). They summarize results from many diagnostic groups who have no trouble passing the WMT recognition subtests even though many of them do have impairment of verbal recall, including patients with bilateral hippocampal damage (Goodrich-Hunsaker & Hopkins, 2009) and left temporal lobectomy (Carone, Green, & Drane, 2014).

Within the Advanced Interpretation program, when the standard rules are applied to the MS case, we see failure on very easy recognition measures (failure on Criterion A), which would be unlikely in valid data, except in cases of dementia. There was also a failure on Criterion B, which involves an inability to produce a credible WMT profile typical of cases with actual severe impairment from dementia, severe dyslexia or another condition that could, in principle, explain failure (e.g. active temporal lobe seizure at the time of testing). Instead of showing progressively lower scores as the subtests became harder, in keeping with objective task difficulty, the MS case produced a paradoxical profile, in which her scores steadily became relatively higher as the subtests became harder. As a result, we see in Figure 1 that the MS case scored lower than the dementia group mean on the very easy recognition subtests, IR, DR, and consistency (CNS), but scored higher than the dementia group mean on the harder subtests, Multiple Choice (MC), Paired Associate (PA) and Free Recall (FR), which are actually much more sensitive to true impairment than the recognition subtests.

A mean easy-hard difference score of at least 30 points is very frequently seen in people with valid data who truly

cannot pass the easy subtests, such as the dementia patients in the study by Green et al. (2011). In contrast, the mean easy-hard difference score in the MS case was only 12 percentage points (mean of IR, DR & CNS minus the mean of MC, PA & FR), indicating little difference in scores between very easy and much harder subtests. Thus, the MS case would be classified as producing invalid test results both at the first assessment, based on the VSVT, and on the second assessment, based on failure of both criteria A and B of the WMT. Such failure on validity tests is known to be linked with a generalized suppression of scores across the test battery (Green et al., 2001, Green, 2007, Green & Flaro, 2019).

Data from studies of people with MS have found that only 12-23% of these patients have two or more tests at least 1.5 SD below the mean (Uher et al., 2014; Viterbo et al., 2013). In contrast, Loring and Goldstein (2019) presented the patient's neuropsychological test results in Table 1 where it can be seen that the patient scored in the bottom tenth percentile or lower on 17 different tests, with a notably extreme score of 300 seconds taken to complete Trail Making B. Longitudinal studies of multiple sclerosis patients involving a 5-year follow-up found mean baseline z scores across cognitive domains ranging from - 0.55 to -1.29 and showing an annualized change of -0.16 (Eijlers et al., 2018). Thus, the production of 17 impaired scores two years after diagnosis is well outside the usual profile seen in these patients, even when considering progression over several years. The usual interpretation would be that these data from the MS case markedly underestimate true ability and that the data are unreliable.

Given the patient's failure on the two main standalone PVTs, we would expect her test scores to vary from time to time within and between assessments because of non-credible performance on PVTs, although the reasons for the invalid results might not be known. It would be anticipated that some scores would conflict with others, as in the case of scoring relatively higher on harder subtests of the WMT than on the easy subtests (Figure 1). An outstanding

Table 1. MS cases from the current series who passed the WMT contrasted with the Loring MS case.

MS Cases passing WMT	WMTIR	WMTDR	WMT CONS	WMTMC	WMTPA	WMTFR	Easy minusHard	AGE	SEX	Years of Education
1	97.5	92.5	90.0	75.0	45.0	37.5	NA	43	F	14
2	100.0	97.5	97.5	90.0	95.0	57.5	NA	55	F	12
3	100.0	100.0	100.0	95.0	90.0	48.0	NA	50	M	14
4	98.0	98.0	95.0	80.0	70.0	68.0	NA	55	F	16
5	95.0	95.0	95.0	65.0	70.0	30.0	NA	53	F	16
6	97.5	97.5	95.0	80.0	95.0	42.5	NA	37	M	16
7	95.0	100.0	95.0	100.0	100.0	70.0	NA	54	F	15
8	97.5	95.0	92.5	90.0	90.0	70.0	NA	36	F	16
9	100.0	97.5	97.5	90.0	70.0	52.5	NA	44	F	11
10	95.0	92.5	87.5	70.0	70.0	47.5	NA	39	F	8
11	100.0	100.0	100.0	100.0	95.0	72.5	NA	46	F	11
12	92.5	90.0	87.5	80.0	75.0	40.0	NA	50	M	6
13	100.0	100.0	100.0	100.0	100.0	47.5	NA	51	M	17
14	97.5	87.5	85.0	80.0	75.0	40.0	NA	30	F	14
15	100.0	92.5	92.5	95.0	90.0	50.0	NA	52	M	17
16	100.0	97.5	97.5	95.0	95.0	47.5	NA	43	F	16
17	100.0	100.0	100.0	90.0	100.0	72.5	NA	41	M	18
18	100.0	100.0	100.0	90.0	90.0	42.5	NA	26	F	12
19	100.0	93.0	93.0	95.0	95.0	48.0	NA	38	F	17
20	90.0	95.0	90.0	70.0	65.0	30.0	NA	46	M	16
21	92.5	92.5	90.0	80.0	85.0	45.0	NA	22	F	12
22	100.0	100.0	100.0	100.0	100.0	50.0	NA	37	F	16
23	97.5	92.5	90.0	60.0	40.0	25.0	NA	38	M	18
24	93.0	93.0	85.0	90.0	90.0	63.0	NA	38	F	14
Mean	97.4	95.8	93.9	85.8	82.9	49.8		42.6		14.3
Std. Dev.	3.0	3.6	4.9	11.6	16.8	13.6		9.0		3.0
Loring MS case	**67.5**	**72.5**	**50**	**60**	**65**	**30**	**11.6**	**Early 50s**	**F**	**Unknown**

IR: immediate recognition; DR: delayed recognition; MC: multiple choice; FR: free recall. All are expressed as percent correct. CONS: consistency between IR & DR responses; It indicates consistency of responses, not accuracy. For example, two incorrect responses from IR to DR would get a score of 1. "Easy minus hard" is $(IR + DR + CONS)/3 - (MC + PA + FR)/3$.
Main comparison data are highlighted in bold.

anomaly is that the MS case produced normal results on the Auditory Verbal Learning Test, proving that she was capable of passing the much easier WMT verbal recognition subtests and yet she failed them with a score not significantly above chance.

Because of such discrepancies, her results would usually not be interpreted as reflecting her actual neuropsychological abilities, but Loring et al. (2019) argued that her WMT test results reflected actual impairment typical of MS patients. They dismissed the idea that the data were invalid, arguing that "low PVT scores reflect disease-related effects of MS." They also ignored the small easy-hard difference on the WMT subtests, a pattern inconsistent with neurologic compromise, but rather indicating non-neurologic variability. They identified "decreased processing speed and impaired working memory" as factors explaining PVT failure on the VSVT and WMT, even though there are no empirical data to support such an interpretation, except perhaps in dementia (Green et al., 2011). The authors were essentially adopting the myth that there is a high level of false positives on the WMT, which has recently been refuted (Alverson et al., 2019; Erdodi et al., 2019).

Loring et al. (2019) interpreted a Reliable Digit Span score of 7 as suggesting valid data, although 7 is often defined as a PVT failure in adults (Greiffenstein, Baker and Gola, 1994; Larrabee, 2003). Many would consider a Digit Span Age Scaled Score of 5 (Table 1, page 4) also to reflect invalid performance (Kirkwood et al., 2011; Webber & Soble, 2018). They also treated two losses of set on the Wisconsin Card Sorting Test as suggesting valid test results,

although two or more errors has been used to define failure in other studies (Larrabee, 2003).

Inaccurate portrayal of drug effects on WMT

Loring et al. (2019) noted that the drug, Zonasamide, taken by the patient causes cognitive impairment and assumed this was why she failed the PVTs. It was stated that Lorazepam, a minor tranquillizer supplied to millions of people annually has "robust effects on WMT performance, decreasing WMT validity performance by 8% in a double blind crossover trial." This assumption must be challenged. If it were true that Lorazepam causes failure on the WMT in healthy adults given this drug, it would mean that the drug is capable of creating a dementia level of impairment on extremely easy recognition memory subtests, which are very resistant to actual organic impairment.

In fact, the original Lorazepam study by Loring et al. (2011) did not show support for the drug causing WMT failure because there was a basic flaw in the design of the trial. In a poster with the pointed title "When Are Your Trial Data Real?" Rohling (2013) challenged the conclusions of Loring et al. (2011) about Lorazepam effects on the WMT. Using Loring et al. (2011) own raw data to refute their major claim, Rohling (2013) pointed out that the WMT had not been given at all in the no-drug, baseline phase. Thus there was no proof that the volunteers had been trying to produce valid results in the baseline phase on no drug. On the contrary, on reanalyzing the original data, Rohling (2013) found that nearly all those who later failed
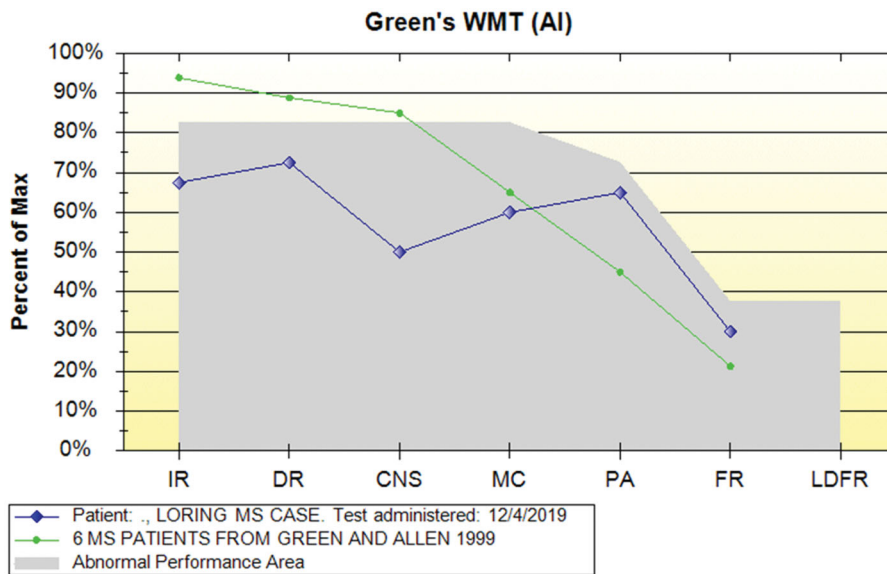
Figure 2. Contrast between previously published WMT mean scores from MS patients and case under discussion.

the WMT when on Lorazepam had already actually failed embedded measures of the Vital Signs Test Battery in the no-drug baseline phase. Several studies have found the base rate of invalid data in research studies to range from 9% in veterans (Clark et al., 2014) to 38% in undergraduates, depending on the criteria used (An et al., 2017; DeRight & Jorgensen, 2015). Similar to the pattern of validty test performance in the study by Loring et al. (2011), DeRight & Jorgensen found that participants who failed validity indicators in the baseline condition were more likely to fail validity indicators during repeat administration. It is also worthwhile to note that participants who failed validity indicators had an average test battery mean at the 15th percentile as compared to an overall score at the 48th percentile for those who passed validity indicators. In short, those failing the WMT were already not producing valid test results in the no-drug baseline phase and, consistent with that, their data were still not valid when on Lorazepam. The study did not show that WMT failure can be explained by a drug effect. Rohling (2013) wrote "These data are an example of how randomized clinical trial results can be distorted by subjects who are poorly motivated, despite being paid."

Performance of MS patients on WMT

There are test manual supplements available from the publisher of the WMT and one of them covers WMT results from people with a variety of neurologic diseases, such as strokes, seizure disorders, brain tumors and ruptured aneurysms. Contrary to Loring et al.'s (2019) claim that there are no data on the WMT in MS patients, one manual supplement describes a group of six cases of MS (Green & Allen, 1999).

In all six MS cases presented (Green & Allen, 1999), the mean WMT scores were as follows: Immediate Recognition = 94% (SD 5), Delayed Recognition = 89% (5), Consistency = 85% (7), Multiple Choice = 65% (7), Paired Associate Recall = 45% (7), Free Recall = 21% (5). The mean of the

easy scores was 89.3% and the mean of the harder scores was 37.3%, a difference of almost 52 points, illustrating the major difference in objective difficulty level between the easy and hard subtests, contrasting markedly with the profile produced by the MS case, as seen in Figure 2. There was only one failure in the original MS sample (GK).

In Figure 2, we see that the MS case scored dramatically lower than the mean of the above MS group on the easiest subtests based on recognition memory (IR and DR), which are unaffected by FSIQ and age and which are not sensitive to most diagnostic conditions (Green & Flaro, 2019). However, she then scored higher than the same group on all three memory subtests, which actually are sensitive to memory impairment. This profile shows intrinsically contradictory data, which cannot be reliable or valid.

Figure 3 shows how extremely impaired the MS case scored on very easy subtests relative to children with intellectual disability (Green et al., 2012, Green & Flaro, 2015, 2016, 2019). Her profile reveals the absence of a meaningful pattern of progressively lower scores as the subtests get harder. Consistent with the notion that these verbal memory scores are not valid, the client went on to show normal verbal learning on certain measures (e.g. the Auditory Verbal Learning Test, AVLT), thereby proving that her scores on verbal memory tests were unreliable. Eichstaedt et al. (2014) commented on a sample of temporal lobe epilepsy patients that examining the easy-hard difference on the WMT "is valuable to identify individuals with severe memory loss who score below criterion on WMT primary effort subtests." People with normal range verbal memory do not show deficits on much easier recognition memory tasks.

In this paper, we will provide new data on MS cases gathered after the 1999 test manual supplement and prior to 2018 for further comparison with the Loring and Goldstein (2019) MS case. The data presented below comprise a retrospective data review from the second author's practice.
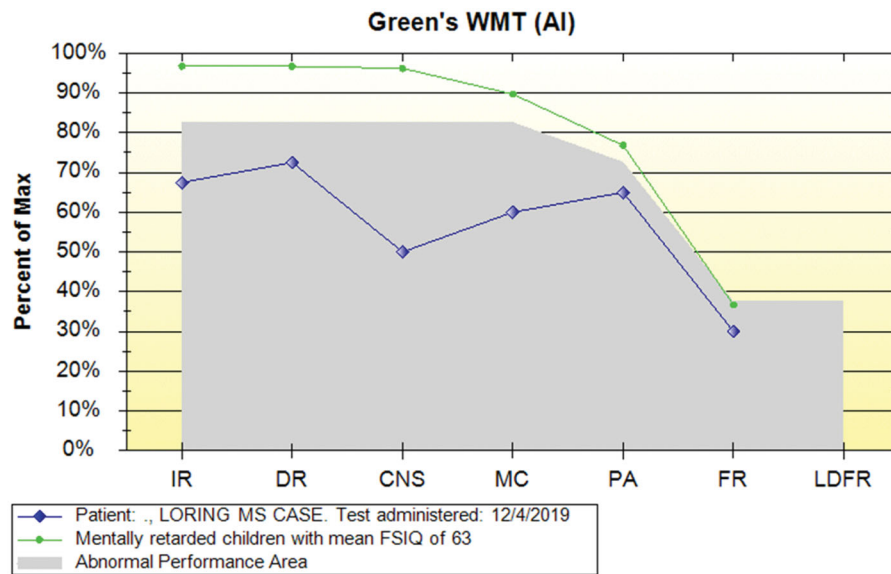
Figure 3. Contrast between the MS case and group of children with mean FSIQ of 63.

Method

Participants

The MS cases shown in Tables 1 and 2 were drawn from a database of 2,173 adults seen consecutively as outpatients by the second author, nearly all in the context of disability assessments. Most were receiving income for disability at least on a temporary and some on a permanent basis, such that there were incentives to exaggerate impairment and disability. Out of this database, all cases with a primary diagnosis of MS or dementia were extracted. A total of 29 MS cases were found, including the six described above. There were 35 cases with some form of dementia diagnosis. The diagnoses of MS or dementia had been made in each case by at least one neurologist usually before referral, occasionally afterwards. Those with any neurologic diagnosis whatsoever (including MS and dementia) are shown in Table 2, broken down into those who passed or failed the WMT. All clients agreed in a signed consent form that their data would later be used anonymously to do retrospective studies.

Results

Table 1 shows that 24 cases of MS out of 29 passed the WMT. Their mean scores on IR (97%) and DR (96%) are very similar to what has been found previously with healthy adults (e.g. Green et al., 2003) and with children (Green & Flaro, 2019). In the latter paper, children with a mean FSIQ of 59 had mean scores of 96% correct on WMT IR and DR, consistent with the fact that these subtests are insensitive to actual differences in abilities in most neurologic illnesses and in children with developmental disabilities. Thus, failure suggests invalid test results or it is a sign of extremely severe cognitive impairment, similar to that seen in Alzheimer's disease and such failure is typically seen in people requiring 24 hours a day care and supervision (Green, Montijo, & Brockhaus, 2011).

In the Advanced Interpretation program, there is a function called "best fit, weighted averages" whereby the computer selects the groups with the most similar WMT profiles to the single case under examination. When we choose available groups from published papers, the single best match to the MS case was a group labeled "Sophisticated simulators, mainly psychologists and physicians ($N = 25$)" (Green et al., 2003). Please see Figure 4. This group had a weighted average value of 0.78, where a value of 1.00 or lower means the profile is very similar to the current case. The simulators were sophisticated in that they all knew that the WMT was designed to identify invalid data. They were asked to act on the test as if they had dementia but to do so in such a way that they were not detected as producing invalid data. All but one case failed the WMT (Green et al., 2003).

Another group very similar to the MS case included 20 patients who passed the WMT in the morning and were asked to take part in a simulator study in the afternoon (Weighted average value 1.23). Thus, we know they were able to pass the WMT but when they took the WMT the second time, we asked them to act as if they had impaired brain function (simulators). Also very similar was a group of 197 people with mild head injury, about 40% of whom failed the WMT. Therefore, the most similar groups on the WMT to the MS case were those known to be producing invalid test results.

MS does not appear to be a candidate for explaining WMT failure if a full effort is applied to doing well. Only 5 cases of MS out of 29 failed the WMT (20%), which is much lower than the failure rate on the WMT in the whole compensation seeking sample (30%). It is lower than the WMT failure rate of 41% in 635 adults with mild TBI ($z = 2.15$, $p = .01$) and about the same as the 21% failure rate in 214 cases with moderate to severe TBI in the same sample ($z = 0.14$, $p > .05$). Note that the excess of failures in mild versus severe TBI constitutes a reverse dose-response effect in these data, proving that it is not severity of brain dysfunction that causes failure on the WMT (Hill, 1965).

Table 2. MS cases who failed the WMT contrasted with the Loring MS case, all dementia patients and neurologic cases who either passed or failed the WMT.

	IR	DR	CONS	MC	PA	FR	Easy minus hard	Age	Sex	Years of education
MS cases failing WMT										
1	75.0	80.0	85.0	40.0	40.0	45.0	38.0	40	F	17
2	55.0	52.5	62.5	30.0	30.0	40.0	23.3	38	F	18
3	77.5	55.0	52.5	45.0	45.0	17.5	25.8	35	F	12
4	90.0	85.0	80.0	70.0	50.0	17.5	45.8	47	M	17
5	75.0	52.5	57.5	35.0	40.0	12.5	16.9	31	M	12
Mean	**74.5**	**65.0**	**67.5**	**44.0**	**41.0**	**26.5**	**31.8**	**38.2**	**–**	**15.2**
Std. Dev.	12.5	16.1	14.2	15.5	7.4	14.8		5.9		2.9
All 189 neurologic cases passing the WMT	96.5 (3.8)	96.4 (3.6)	94.4 (4.6)	86.4 (12.9)	83.4 (15.6)	48.2 (15.5)	NA	47.9 (9.5)	50% F	13.4 (2.8)
All 59 neurologic cases failing the WMT	73.6 (17.2)	71.7 (17.5)	69.3 (11.3)	40.6 (20.5)	44.5 (20)	24.8 (15.5)	34.8	48.8 (10.2)	45% F	12.4 (3.3)
All 35 dementia cases	87.1 (10.8)	85.5 (14.1)	82.5 (13.1)	60.7 (27.2)	58.0 (24.1)	28.7 (14.6)	11.6	60 (9)	66% F	13.2 (3.5)
Loring MS case	**67.5**	**72.5**	**50**	**60**	**65**	**30**		**Early 50s**	**F**	**Unknown**

Main comparison data are highlighted in bold.

In MS cases, there were brain abnormalities on CT or MRI in all cases, whether they passed or failed the WMT. In the whole neurologic sample, for whom such data were available ($n = 150$), brain abnormalities were evident in 90% of the subgroup who passed the WMT ($n = 117$) and 91% of those who failed the WMT ($n = 33$). Thus the presence of abnormal brain imaging did not predict WMT failure. Such an absence of a dose-response relationship between objective brain abnormality and WMT failure contraindicates brain abnormality causing WMT failure in MS and most other neurologic conditions.

In our whole sample, data were available on CT or MRI brain imaging in 961 cases. In those who passed the WMT, 54% of cases had an abnormal brain image. However, in those who failed the WMT, only 36% had an abnormal image. This is yet further evidence of a reversed dose-response relationship (Hill, 1965), meaning that WMT failure is inversely related to the presence of visualized structural brain abnormality. Rather than brain lesions leading to WMT failure, those with brain lesions are significantly less likely to fail the WMT than those with no brain lesions visible. This is a highly significant and clinically important finding ($F = 29$, df 1, 959, $p < .0001$).

The WMT scores of the MS case in Table 1 were not just below established cut-offs for valid data, but were many standard deviations below the means from the 24 MS cases who passed the WMT. For example, her score on IR was 10 SD lower than the mean from the MS cases passing the WMT. Even in non-normal data, such an extreme separation suggests membership of very different groups, in this case, people with valid versus invalid data.

Table 2 shows that the MS case had a mean of 63% correct on IR, DR and CNS, which is no better than chance. In doing so, she scored even lower than the mean of 69% for the five MS cases who failed the WMT. On consistency of responses, she scored 50% which, if valid, means that her memory on these very easy recognition subtests was zero and no better than chance, as if she had not seen the word list at all. If that were the case, her score on the last two subtests (PA and FR) would be zero. Yet, on these subtests, her scores were 65 and 30%, which are far above zero. This is another important inconsistency within her data.

The MS case scored dramatically lower on WMT IR, DR and CNS than the neurologic cases as a whole who passed the WMT and even lower than the means from those who failed the WMT, which would include those with a dementia diagnosis (Table 2). Her scores on IR, DR and CNS were as much as two standard deviations lower than the means from the dementia group, some of whom were genuinely unable to pass the WMT. For example, the mean IR score in the dementia group was 87.1% correct (SD 10.8) compared with 67.5% correct in the MS case, a difference of almost two standard deviations.

The mean Trail Making B time taken by the MS cases passing the WMT was 77.6 seconds (SD 32) and in those failing the WMT it was 100 seconds (SD 54). In contrast, the Loring case had a mean Trail Making B time of 300 seconds. Her Trail Making score was, therefore, an
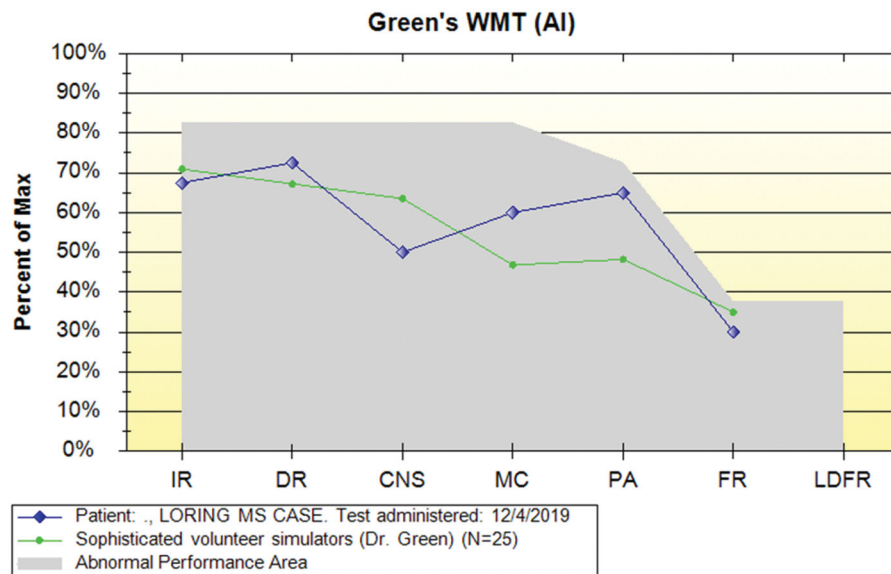
Figure 4. Single case of MS versus the most similar group mean profile from the AI program (sophisticated volunteer simulators).

extreme outlier, even for those MS cases failing WMT with artificially lowered scores on cognitive tests. In addition, it is well outside groups of invalid patients who had a mean time of 158 seconds (Trueblood & Schmidt, 1993) and was beyond that of TBI patients, <5% of whom scored over 200 seconds (Iverson et al., 2002), adding evidence of failed embedded PVTs along with the WMT.

Discussion

Loring et al. (2019) concluded that their single case of MS failed the WMT because of brain disease and/or drug effects. However, in our sample, actual brain disease was not associated with failure on the WMT. Quite the contrary, it was those with the least severe TBI who failed the WMT the most and it was those with the most objective radiological abnormalities of the brain who failed the WMT the least.

Loring et al. (2019) suggested that Lorazepam had been found to have a "robust effect" on the WMT. As argued above, their own data did not support such a conclusion. Those who scored below established cutoffs on the WMT when on Lorazepam had already failed embedded validity tests in the no-drug baseline phase (Rohling, 2013). The logical conclusion would have been that those scoring low on the WMT when on Lorazepam were not completing tests in a manner that would produce valid results. This is certainly the implication of their failure on embedded effort tests in the no-drug baseline phase. The argument that healthy adults volunteering for a study of a commonly used minor tranquilizer often fail the WMT because of the drug is not supported by the evidence. To fail PVTs, a healthy adult would have to perform worse than the average person with early dementia on extremely easy tasks, even though these tasks are unrelated to FSIQ and age and even though their scores are unaffected by most neurologic diseases (Green et al., 2003, Green & Flaro, 2019). If a drug had such an effect, it would probably be withdrawn promptly from the market.

The MS case of Loring et al. (2019) failed the VSVT in the first assessment and she failed the WMT in the second assessment. She produced WMT results that are not typical of MS patients tested in our sample. The majority of MS cases (24 out of 29) passed the WMT, despite there being an external incentive to exaggerate impairment to obtain or keep disability status and their recognition scores on the WMT were no different from healthy adults. Thus, MS does not generally suppress WMT recognition scores or cause WMT failure.

Relative to the MS cases shown in the tables, the MS case presented by Loring et al. (2019) produced extremely low scores on the main validity subtests of the WMT, which are based on recognition memory and are generally insensitive to impairment, including impairment from MS, as shown in current data. It was previously reported that neurologic patients as a whole easily pass the WMT and that their scores on the recognition subtests are the same as seen in healthy adults (Green & Allen, 1999). This was true also for those who were selected as having impaired verbal memory on the CVLT (Green et al., 2003). They scored just as highly on the recognition tasks of the WMT as the neurological cases with normal range memory. The MS case under review produced extremely low WMT scores relative to other MS cases and relative to other neurologic cases on subtests which are hardly affected by brain disease in most cases, with the exception of certain dementia like conditions.

WMT failures are always viewed in the clinical context. In any case of WMT failure, Slick et al. (1999) Criterion D (e.g. behaviors are not fully accounted for by neurologic, psychiatric, or developmental factors) must be applied and we have to decide between the two explanations using other available information. For example, in the dementia group in Table 2, there were cases of normal pressure hydrocephalus, fronto-temporal dementia, Alzheimer's disease and olivo-pontine degeneration with marked expressive aphasia. Such conditions can cause WMT failure. Note that the mean IR and DR scores in the dementia group are actually

above the standard cutoffs for invalid data and well above those of the MS case, although some did fail the WMT.

The MS case scored substantially lower than the mean from the dementia group on the WMT recognition measures. If we were to argue that she genuinely had much more impairment than the average dementia patient, we would have to explain the overall profile of results which she produced, as well as providing evidence that she was actually functioning worse than most dementia patients in daily life. There was no evidence of such low functioning in the report by Loring et al. (2019), such as requirements for 24-hour care, as needed in severe dementia patients. In fact, she performed normally on the RAVLT (74th percentile), demonstrating memory abilities far above those of neurologic patients with impaired memory who were still able to pass the WMT effort subtests, and negating any credible neurologic reason for failing the measure. Her impaired score on Trail Making B, far from being typical of MS patients, was an extreme outlier when compared with either valid or invalid MS cases presented above (i.e. pass versus fail the WMT, as in Table 2).

Most importantly, her WMT profile contains internally inconsistent results, which are at odds with what may be seen in genuine cases of cognitive impairment. In all three figures, the MS case scored lower than impaired groups on the easiest subtests but she scored relatively much higher on the harder subtests. She did not show the expected easy-hard difference of at least 30 points, which was observed in all dementia cases in the study by Green et al. (2001). Instead she produced a non-credible profile, with lower scores than dementia patients on very easy subtests and higher scores on harder subtests (Figure 1). Her profile was most similar to that of volunteers who have been asked to simulate impairment (Figure 4). Going forward, further research is encouraged with regard to PVTs in MS patients and those with other neurologic diseases. More detailed analysis of the relationship between PVTs and disease severity, objective measures of brain involvement, and functional skills could yield fruitful insights.

The conclusion of Loring et al. (2019) that their single MS case was performing validly is not supported by our analysis of the data. Their conclusion encourages faulty clinical judgment and it stands in direct contradiction to copious data on performance validity tests and the true impact on the WMT of neurologic diseases, including MS.

References

Alverson, W. A., O'Rourke, J. F., & Soble, J. R. (2019). The Word Memory Test genuine memory impairment profile discriminates genuine memory impairment from invalid performance in a mixed clinical sample with cognitive impairment. *The Clinical Neuropsychologist*, *33*(8), 1420–1435. https://doi.org/10.1080/13854046.2019.1599071

An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2017). Performance validity in undergraduate research participants: A comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist*, *31*(1), 193–206. https://doi.org/10.1080/13854046.2016.1217046

Carone, D. A., Green, P., & Drane, D. L. (2014). Word Memory Test profiles in two cases with surgical removal of the left anterior hippocampus and parahippocampal gyrus. *Applied Neuropsychology: Adult*, *21*(2), 155–160. https://doi.org/10.1080/09084282.2012.755533

Clark, A. L., Amick, M. M., Fortier, C., Milberg, W. P., & McGlinchey, R. E. (2014). Poor performance validity predicts clinical characteristics and cognitive test performance of OEF/OIF/OND veterans in a research setting. *The Clinical Neuropsychologist*, *28*(5), 802–825.

DeRight, J., & Jorgensen, R. S. (2015). I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *The Clinical Neuropsychologist*, *29*(1), 101–117. https://doi.org/10.1080/13854046.2014.989267

Eichstaedt, K. E., Clifton, W. E., Vale, F. L., Benbadis, S. R., Bozorg, A. M., Rodgers-Neame, N. T., & Schoenberg, M. R. (2014). Sensitivity of Green's Word Memory Test genuine memory impairment profile to temporal pathology: A study in patients with temporal lobe epilepsy. *The Clinical Neuropsychologist*, *28*(6), 941–953. https://doi.org/10.1080/13854046.2014.942374

Eijlers, A. J. C., van Geest, Q., Dekker, I., Steenwijk, M. D., Meijer, K. A., Hulst, H. E., Barkhof, F., Uitdehaag, B. M. J., Schoonheim, M. M., & Geurts, J. J. G. (2018). Predicting cognitive decline in multiple sclerosis: A 5-year follow-up study. *Brain*, *141*, 2605–2618. https://doi.org/10.1093/brain/awy202

Erdodi, L. A., Green, P., Sirianni, C. D., & Abeare, C. A. (2019). The myth of high false-positive rates on the word memory test in mild TBI. *Psychological Injury and Law*, *12*(2), 155–169. https://doi.org/10.1007/s12207-019-09356-8

Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2009). Word Memory Test performance in amnesic patients with hippocampal damage. *Neuropsychology*, *23*(4), 529–534. https://doi.org/10.1037/a0015444

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. https://doi.org/10.1080/02699050110088254

Green, P., & Allen, L. (1999). *Performance of neurologic patients on the Word Memory Test (WMT) and Computerized Assessment of Response Bias (CARB). May 1999 Supplement to the Word Memory Test and CARB '97 Manuals.* Cognisyst, N.C. & Green's Publishing.

Green, P., & Astner, K. (1995). *Manual for the Oral Word Memory Test.* Green's Publishing.

Green, P. (2003, revised 2005). *Green's Word Memory Test for Windows; User's manual.* Green's Publishing.

Green, P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America*, *18*(1), 43–68. https://doi.org/10.1016/j.pmr.2006.11.002

Green, P. (2008). *The Advanced Interpretation (AI) program.* Green's Publishing.

Green, P., & Flaro, F. (2015). Results from three performance validity tests (PVTs) in adults with intellectual deficits. *Applied Neuropsychology: Adult*, *22*(4), 293–303. https://doi.org/10.1080/23279095.2014.925903

Green, P., & Flaro, L. (2016). Results from three performance validity tests in children with intellectual disability. *Applied Neuropsychology: Child*, *5*(1), 25–34. https://doi.org/10.1080/21622965.2014.935378

Green, P., & Flaro, L. (2019). Performance validity test failure predicts suppression of neuropsychological test results in developmentally disabled children. *Applied Neuropsychology: Child.* https://doi.org/10.1080/21622965.2019.1604342

Green, P., Allen, L., & Astner, K. (1996). *Manual for the computerized Word Memory Test.* Green's Publishing.

Green, P., Flaro, L., Brockhaus, R., & Montijo, J. (2012). Performance on the WMT, MSVT, and NV-MSVT in children with developmental disabilities and in adults with mild traumatic brain injury. In A. Horton, & C. Reynolds (Eds.), *Detection of malingering during head injury litigation,* 2nd edition. Springer.

Green, P., Lees-Haley, P. R., & Allen, L. M. (2003). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, *2*(3–4), 97–124. https://doi.org/10.1300/J151v02n03_05

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology*, *18*(2), 86–94. https://doi.org/10.1080/09084282.2010.523389

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*(3), 218–224. https://doi.org/10.1037/1040-3590.6.3.218

Hill, A. B. (1965). The environment and disease: Association or causation. *Proceedings of the Royal Society of Medicine*, *58*(5), 295–300. https://doi.org/10.1177/003591576505800503

Iverson, G. L., Lange, R. T., Green, P., & Franzen, M. D. (2002). Detecting exaggeration and malingering with the trail making test. *The Clinical Neuropsychologist*, *16*(3), 398–406. https://doi.org/10.1076/clin.16.3.398.13861

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The Value of the WISC-IV Digit Span Subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology*, *26*(5), 377–384. https://doi.org/10.1093/arclin/acr040

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*(3), 410–425. https://doi.org/10.1076/clin.17.3.410.18089

Loring, D. W., & Goldstein, F. C. (2019). If invalid PVT scores are obtained, can valid neuropsychological profiles be believed? *Archives of Clinical Neuropsychology*, *34*(7), 1192–1111. https://doi.org/10.1093/arclin/acz028

Loring, D. W., Marino, S. E., Drane, D. L., Parfitt, D., Finney, G. R., & Meador, K. J. (2011). Lorazepam's effects on WMT Performance: A randomized, double-blind, placebo controlled, cross-over trial. *The Clinical Neuropsychologist*, *25*(5), 799–811. https://doi.org/10.1080/13854046.2011.583279

Macciocchi, M. N., Seel, R. T., Alderson, A., & Godsall, R. (2006). Victoria Symptom Validity Test performance in acute severe traumatic brain injury: Implications for test interpretation. *Archives of Clinical Neuropsychology*, *21*(5), 395–404. https://doi.org/10.1016/j.acn.2006.06.003

Rohling, M. L. (2013). *When are your trial data real?* [Poster presentation]. 53rd Annual Conference of the American Society for Clinical Psychopharmacology (NCDEU) in Hollywood, FL, USA.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561. https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545

Slick, D. J., Tan, J. E., Strauss, E., Mateer, C. A., Harnadek, M., & Sherman, E. M. (2003). Victoria Symptom Validity Test scores of patients with profound memory impairment: Nonlitigants case studies. *The Clinical Neuropsychologist*, *17*(3), 390–394. https://doi.org/10.1076/clin.17.3.390.18090

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *VSVT: Victoria Symptom Validity Test, version 1.0. Professional manual.* Psychological Assessment Resources, Inc.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *15*(4), 578–590. https://doi.org/10.1080/01688639308402580

Uher, T., Blahova-Dusankova, J., Horakova, D., Bergsland, N., Tyblova, M., Benedict, R. H. B., Kalincik, T., Ramasamy, D. P., Seidl, Z., Hagermeier, J., Vaneckova, M., Krasensky, J., Havrdova, E., & Zivadinov, R. (2014). Longitudinal MRI and neuropsychological assessment of patients with clinically isolated syndrome. *Journal of Neurology*, *261*(9), 1735–1744. https://doi.org/10.1007/s00415-014-7413-9

Viterbo, R. G., Iaffaldano, P., & Trojano, M. (2013). Verbal fluency deficits in clinically isolated syndrome suggestive of multiple sclerosis. *Journal of the Neurological Sciences*, *330*(1–2), 56–60. https://doi.org/10.1016/j.jns.2013.04.004

Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. *Clinical Neuropsychologist*, *32*(4), 657–670. https://doi.org/10.1080/13854046.2017.1415374