



Article

Examining the Clinical Utility of Selected Memory-Based Embedded Performance Validity Tests in Neuropsychological Assessment of Patients with Multiple Sclerosis

John W. Lace^{1,*}, Zachary C. Merz² and Rachel Galioto^{1,3}

¹ Neurological Institute, Section of Neuropsychology, Cleveland Clinic Foundation, Cleveland, OH 44195, USA; galiotr@ccf.org

² LeBauer Department of Neurology, The Moses H. Cone Memorial Hospital, Greensboro, NC 27401, USA; zachary.merz@conehealth.com

³ Mellen Center for Multiple Sclerosis, Cleveland Clinic Foundation, Cleveland, OH 44195, USA

* Correspondence: lacej@ccf.org



Citation: Lace, J.W.; Merz, Z.C.; Galioto, R. Examining the Clinical Utility of Selected Memory-Based Embedded Performance Validity Tests in Neuropsychological Assessment of Patients with Multiple Sclerosis. *Neurol. Int.* **2021**, *13*, 477–486. <https://doi.org/10.3390/neurolint13040047>

Academic Editor: Marcello Moccia

Received: 3 August 2021

Accepted: 8 September 2021

Published: 23 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Within the neuropsychological assessment, clinicians are responsible for ensuring the validity of obtained cognitive data. As such, increased attention is being paid to performance validity in patients with multiple sclerosis (pwMS). Experts have proposed batteries of neuropsychological tests for use in this population, though none contain recommendations for standalone performance validity tests (PVTs). The California Verbal Learning Test, Second Edition (CVLT-II) and Brief Visuospatial Memory Test, Revised (BVMT-R)—both of which are included in the aforementioned recommended neuropsychological batteries—include previously validated embedded PVTs (which offer some advantages, including expedience and reduced costs), with no prior work exploring their utility in pwMS. The purpose of the present study was to determine the potential clinical utility of embedded PVTs to detect the signal of non-credibility as operationally defined by below criterion standalone PVT performance. One hundred thirty-three (133) patients (*M* age = 48.28; 76.7% women; 85.0% White) with MS were referred for neuropsychological assessment at a large, Midwestern academic medical center. Patients were placed into “credible” (*n* = 100) or “noncredible” (*n* = 33) groups based on a standalone PVT criterion. Classification statistics for four CVLT-II and BVMT-R PVTs of interest in isolation were poor (AUCs = 0.58–0.62). Several arithmetic and logistic regression-derived multivariate formulas were calculated, all of which similarly demonstrated poor discriminability (AUCs = 0.61–0.64). Although embedded PVTs may arguably maximize efficiency and minimize test burden in pwMS, common ones in the CVLT-II and BVMT-R may not be psychometrically appropriate, sufficiently sensitive, nor substitutable for standalone PVTs in this population. Clinical neuropsychologists who evaluate such patients are encouraged to include standalone PVTs in their assessment batteries to ensure that clinical care conclusions drawn from neuropsychological data are valid.

Keywords: multiple sclerosis; neuropsychological assessment; noncredible performance; embedded performance validity tests

1. Introduction

Multiple sclerosis (MS) is a chronic, inflammatory, demyelinating condition that can present with diverse patterns of physical, cognitive, and psychiatric symptoms [1–4]. In addition to expert care provided by neurologists who may provide clinical monitoring and pharmacological intervention (e.g., dimethyl fumarate, interferons, monoclonal antibodies) [5,6], clinical neuropsychologists are often called upon to assess such patients and provide targeted treatment recommendations for them from a neurocognitive and psychological perspective [7,8]. Several attempts to standardize the clinical neuropsychological assessment of patients with MS (pwMS) have been made. The National MS Society

provided practicable recommendations for cognitive screening in patients with MS [1]. Over the previous two decades, the Minimal Assessment of Cognitive Function in MS (MACFIMS) [9], the Brief International Cognitive Assessment for MS (BICAMS) [10], and the Mercy Evaluation of Multiple Sclerosis (MEMS) [11] have been proposed as compendious neuropsychological batteries for more thorough evaluation (see [11] for a review of similarities and differences among these batteries).

Clinical neuropsychologists are tasked with identifying bona fide neuropsychological dysfunction while ensuring that extraneous factors—which may negatively influence results and clinical conclusions—are considered. Notably, noncredible presentations of cognitive impairment can seriously reduce overall neuropsychological performance above and beyond disease- or injury-related variables [12,13]. Several case studies of noncredible cognitive presentations in pwMS have been reported and debated [14–17]. At a broader level, more than one-fifth of clinically-referred patients with MS have been reported to demonstrate noncredible performance [18,19], consistent with base rates in other clinically referred neuropsychological samples [20]. Despite there being “no generally accepted explanation for suboptimal cognitive performance in MS” [19] (p. 1), many have identified the presence of external incentive (e.g., receipt of disability insurance), significant psychological/psychiatric symptoms, and pain/fatigue as being potential contributing factors in this and other populations [21–23] while suggesting that the degree of the radiological disease may be relatively noncontributory to PVT performance in pwMS [18].

Importantly, clinical neuropsychologists must determine the validity of cognitive data—typically via standalone and/or embedded performance validity tests (PVTs)—to ensure clinical decisions and aspects of continued medical care for pwMS are pursued ethically and responsibly [24,25]. As an aside, experts have encouraged neurologists to become familiar with validity assessment methods to aid in the interpretation of neuropsychological test results [26]. Despite the reported base rates of noncredible performance in MS, none of the aforementioned neuropsychological batteries (nor the National MS Society recommendations) explicitly suggest the inclusion of a standalone PVT in clinical assessment. Only the paper describing the MEMS [11] discussed in clear detail the necessity of PVT use in this context, though it did not specifically suggest a standalone PVT in its battery. Thus, clinicians who utilize these tools are left with the task of evaluating the only available embedded PVTs to determine credibility.

In contrast to standalone PVTs, which require their own administration time and materials and tend to have improved psychometric properties, embedded PVTs offer their own balance of benefits and detriments [27]. On one hand, they tend to offer lower classification accuracies and sensitivities compared to their standalone analogs, with many embedded PVTs approaching the so-called “Larrabee limit” [28] of approximately 0.50 sensitivity when maintaining ideal specificity (i.e., ≥ 0.90). On the other hand, they may increase efficiency in neuropsychological testing procedures by reducing the duration of examination time and decreasing the overall test burden on examinees. This consideration is of clinical relevance given the frequency of reported fatigue in MS [1,10]. Each of the aforementioned batteries recommends the use of the California Verbal Learning Test, Second Edition (CVLT-II) [29] and Brief Visuospatial Memory Test, Revised (BVMT-R) [30], both of which are well-validated list-learning and visual learning/memory tasks (respectively) that include embedded PVTs [31,32]. Specifically, the CVLT-II includes embedded PVTs of Recognition Hits (C-RH) and Forced Choice Recognition (C-FCR), and the BVMT-R includes Recognition Hits (B-RH) and Recognition Discrimination (B-RD). Each of these variables has been supported as embedded PVTs in recent literature, mostly in individuals with traumatic brain injuries or mixed clinical samples [33–37].

Unfortunately, these embedded PVTs remain infrequently examined in the context of MS despite their potential for clinical value. To the authors’ knowledge, only Domen and colleagues [38] have explored this question, and they reported specificity values ≥ 0.92 for previously validated cutoffs for C-FCR and B-RD, suggesting that traditional cutoffs may not yield high rates of false-positive errors. Importantly, their findings are limited by

their decision to exclude patients performing below expected limits on a standalone PVT and subsequently not reporting classification statistics besides specificity (i.e., sensitivity, classification accuracy).

As such, the present study examined the clinical utility of embedded PVT variables in CVLT-II and BVMT-R—recommended by many experts for routine use in pwMS—in identifying noncredible performance, operationally defined as below criterion performance on a well-validated, standalone, memory-based PVT, in a sample of pwMS. We evaluated select variables both in isolation and in combination (using both arithmetically- and logistic regression-derived formulas) to determine their possible utility.

2. Materials and Methods

2.1. Patient Characteristics

Patients with MS (pwMS) were identified by retrospective analysis of an archival clinical dataset of individuals referred for clinical neuropsychological assessment within a large, Midwestern, academic medical center. All patients underwent neuropsychological assessment between 2017 and 2021. Inclusion criteria were: (1) complete data for all variables of interest; (2) estimated premorbid intellectual ability, as measured by a word reading test, ≥ 70 ; and (3) most recent or only neuropsychological evaluation.

The sample included 133 patients (M age = 48.28, SD = 12.55; M Education = 13.96 years, SD = 2.49). Most patients were women (76.7%) and identified as White/Caucasian (85.0%). Most patients were diagnosed with a relapsing-remitting subtype (75.2%), with relatively fewer primary (15.0%) and secondary (9.8%) progressive subtypes. Patients (n = 118 with data for this variable) had an average disease duration (since reported symptom onset) of 13.34 years (SD = 10.53).

2.2. Measures

2.2.1. Wide Range Achievement Test, Fourth Edition

The Word Reading subtest of the Wide Range Achievement Test, Fourth Edition (WRAT-4) [39] was used as an estimate of premorbid intellectual ability.

2.2.2. Victoria Symptom Validity Test

The Victoria Symptom Validity Test (VSVT) [40] is a widely used, standalone, memory-based PVT. A recent systematic review and meta-analysis identified the best performing variable as the total score of ≤ 40 to indicate noncredible performance [41]. The VSVT was chosen for use in this sample of patients with MS because it is minimally affected by bona fide deficits in working memory, processing speed, and memory [42] and is not likely related to MS disease burden [18].

2.2.3. Embedded PVTs

Forced Choice Recognition (C-FCR) and Recognition Hits (C-RH) from the CVLT-II [29] and Recognition Hits (B-RH) and Recognition Discrimination (B-RD) from the BVMT-R [30] were chosen due to their availability within the MACFIMS, BICAMS, and MEMS [9–11].

2.3. Procedures

Approval from the first and third authors' institution's IRB was obtained prior to data analysis. All patients were referred for neuropsychological assessment as part of medical care and were evaluated by a board-certified clinical neuropsychologist or clinical psychologist trained under Houston Conference Guidelines. All measures were administered by either a board-certified or trained psychometrist, clinical psychology doctoral student, or postdoctoral neuropsychology fellow proficient in test administration and scoring. Patients with VSVT Total Accuracy scores ≥ 41 were categorized as "credible" and those with VSVT Total Accuracy scores ≤ 40 were categorized as "noncredible" [41].

2.4. Statistical Analyses

All analyses were conducted with SPSS 26.0. Demographic variables were compared between credible and noncredible groups using χ^2 or Mann-Whitney U tests. Embedded PVTs in isolation and arithmetically and logistic regression-derived formulas were evaluated. Five arithmetic formulas were computed to determine if combining variables resulted in improved classification. These combinations of variables were:

1. C-FCR + C-HR + B-HR + B-RD;
2. C-FCR + B-RH;
3. C-FCR + B-RD;
4. C-RH + B-RH; and
5. C-RH + B-RD.

Five logistic regressions (LRs) were performed with the combinations of variables described in the formulas above entered as predictors as one set (i.e., all four were entered as predictors for the first, C-FCR and B-RH were entered as predictors for the second, etc.) with credible/noncredible group membership as the binary outcome variable

Mann-Whitney U tests, areas under the receiver operating characteristic curve (AUCs), sensitivity, and specificity were calculated for PVTs in isolation using credible vs. non-credible group membership as the criterion variable. AUCs were additionally computed for each arithmetically- and logistic regression-derived formula. A conservative adjusted criterion α of 0.013 (i.e., 0.05/4) was used for χ^2 , Mann-Whitney U, and AUC analyses due to the abundance of repeated analyses to minimize the likelihood of Type 1 error. AUC values ≥ 0.70 suggest at least acceptable discriminability, and values below 0.70 indicate poor classification ability and are generally considered unacceptable [43].

3. Results

On average, the premorbid intellectual ability was in the average range (M WRAT-4 Reading Standard Score = 96.79, SD = 10.79). Regarding the VSVT, 100 (75.2%) patients were classified as having “credible” neuropsychological test performance and 33 (24.8%) were classified as having “noncredible” neuropsychological test performance [41].

Demographic characteristics for both groups are displayed in Table 1. The credible and noncredible groups did not significantly differ in terms of age, racial/ethnic identity (coded and White vs. Non-White), MS subtype, nor symptom duration (all $ps \geq 0.08$). The credible and noncredible groups differed in terms of gender and years of education ($ps \leq 0.01$). However, the magnitude (i.e., effect size) of the gender difference was negligible-to-weak (Cramer’s $V = 0.23$) [44], consistent with similar previous research [45], and was not considered to be impactful. Additionally, the noted difference in years of education between groups was not considered to be clinically meaningful, as has been the case with prior literature with similar aims [46,47].

Mann-Whitney U tests revealed that each of the four embedded PVT variables (C-FCR, C-HR, B-HR, and B-RD) were not significantly different between groups (all $ps > 0.017$), with notably small effect sizes (ds ranged from 0.22–0.37). AUCs for each of the four embedded PVT variables ranged from 0.58 (C-FCR) to 0.62 (B-RD), all of which were not statistically significant (ps ranged from 0.034 to 0.195) and unacceptable (i.e., all ≤ 0.70). Sensitivities for these variables at previously validated cutoff scores ranged from 0.12 (B-RD ≤ 3) to 0.33 (C-FCR ≤ 15 , C-RH ≤ 10 , and C-RH ≤ 11). Specificity values generally hovered around 0.90, with the exception of C-FCR ≤ 15 , which yielded lower specificity of 0.80. Classification accuracy ranged from 0.68 (C-FCR ≤ 15) to 0.77 (C-RH ≤ 10 and B-RH ≤ 4). Various cutoff scores and their sensitivity, specificity, and total accuracy values are displayed in Table 2.

Table 1. Demographic characteristics of credible and noncredible groups.

	Credible (<i>n</i> = 100)	Noncredible (<i>n</i> = 33)	χ^2 (<i>p</i>) or <i>U</i> (<i>p</i>)
	<i>M</i> (<i>SD</i>) or <i>n</i> (%)		
Age, Years	48.83 (13.13)	47.18 (10.67)	0.25 (0.76)
Gender	-	-	7.30 (0.01)
Male	29 (29)	2 (6)	-
Female	71 (71)	31 (94)	-
Education, Years	14.36 (2.49)	12.76 (2.11)	1020.00 (≤0.01)
Racial/Ethnic Identity	-	-	0.29 (0.59)
White	84 (84)	29 (88)	-
Non-White	16 (16)	4 (12)	-
MS Subtype	-	-	0.91 (0.64)
Relapsing-Remitting	75 (75)	25 (76)	-
Secondary Progressive	11 (11)	2 (6)	-
Primary Progressive	14 (14)	6 (18)	-
Symptom Duration, Years (<i>n</i> = 118)	14.32 (10.93)	10.18 (8.54)	979.50 (0.08)

N = 133 unless otherwise noted. *p* values displayed in boldface are considered significant at the adjusted criterion α of 0.013.

Table 2. Embedded CVLT-II and BVMT-R variables and classification statistics.

Variable	Raw Cutoff	AUC (<i>p</i> ¹)	Credible <i>M</i> (<i>SD</i>)	Noncredible <i>M</i> (<i>SD</i>)	<i>U</i> (<i>p</i> ¹), <i>d</i> ²	Sen.	Spec.	Acc.
C-FCR	≤15 ^a	0.58 (0.195)	15.58 (1.03)	14.97 (1.98)	1401.50 (0.080), 0.22	0.33	0.80	0.68
	≤14 ^b	-	-	-	-	0.24	0.90	0.74
C-RH	≤11 ^c	0.61 (0.066)	13.71 (2.36)	12.45 (3.16)	1296.50 (0.062), 0.32	0.33	0.89	0.75
	≤10 ^c	-	-	-	-	0.33	0.91	0.77
B-RH	≤4 ^d	0.60 (0.096)	5.66 (1.28)	5.15 (1.28)	1330.50 (0.042), 0.29	0.24	0.95	0.77
B-RD	≤4 ^e	0.62 (0.034)	5.44 (0.88)	4.85 (1.46)	1244.00 (0.017), 0.37	0.30	0.90	0.75
	≤3 ^d	-	-	-	-	0.12	0.92	0.72

Credible *n* = 100. Noncredible *n* = 33. ¹ Uncorrected *p* value reported, with conservative critical α of 0.013 chosen to interpret statistical significance to minimize Type 1 error. ² Cohen's *d* effect size converted from Mann-Whitney *U* values. [48]. ^a [49]. ^b [37]. ^c [34]. ^d [35]. ^e [33]. C-FCR = California Verbal Learning Test, Second Edition (CVLT-II) Forced Choice Recognition. C-RH = CVLT-II Recognition Hits. B-RH = Brief Visuospatial Memory Test, Revised (BVMT-R) Recognition Hits. B-RD = BVMT-R Recognition Discrimination. Sen. = Sensitivity. Spec. = Specificity. Acc. = Overall classification accuracy.

As stated above, five arithmetic formulas were computed to determine if combining embedded PVT variables resulted in improved classification. AUCs for these variables ranged from 0.61 to 0.63 and were all nonsignificant compared to the conservative critical α (*ps* > 0.03) and unacceptable. Additionally, five logistic regressions (LRs) were performed with the combinations of variables described in the five formulas above entered as predictors as one set (i.e., all four were entered as predictors for the first, C-FCR and B-RH were entered as predictors for the second, etc.) with credible/noncredible group membership as the binary outcome variable. Exponentiated equations were derived from LR results, with similarly unacceptable AUC variables for each (0.61–0.64) such that were all nonsignificant compared to the conservative critical α (*ps* > 0.02). Due to these poor results and lack of potential clinical utility, cutoff scores and sensitivity/specificity values were not identified for these novel exponentiated equation variables as they appeared psychometrically inadequate.

4. Discussion

The present study sought to evaluate the clinical utility of four embedded performance validity test (PVT) variables in the CVLT-II and BVMT-R, commonly used in various clinical neuropsychological samples, in patients with MS. These variables were considered in isolation and in combination with both arithmetical and logistic regression-derived methods. Given the base rates of noncredible performance reported in patients with MS [18] and the important implications of non-credibility in neuropsychological assessment [25], the psychometric evaluation of PVTs in the neuropsychological assessment of patients in this population is necessary and clinically warranted [11]. Several findings deserve further discussion.

First and foremost, current findings extended prior work and indicated that general discriminability for several embedded CVLT-II and BVMT-R PVTs were unacceptably low, with none of the AUC values exceeding 0.70 (a generally accepted lower bound criterion denoting at least acceptable discriminability [43,50,51]) nor emerging as statistically significant compared to conservative, adjusted α . Additionally, a series of arithmetically- and logistic regression-derived formulas did not appear to bolster discriminability, despite recent work supporting these methods to derive useful multivariate composite formulas for this purpose [52]. In all, current findings suggested that these select embedded CVLT-II and BVMT-R PVTs are likely not appropriate in isolation nor in combination at detecting noncredible performance, let alone substitutable for standalone PVTs in this population.

Relatedly, current findings, at least in part, further replicate recent work on embedded PVTs in MS. Domen and colleagues [38] reported specificities for previously validated cutoff scores for C-FCR and B-RD ranged from 0.98–0.99 and 0.88–0.93, respectively. Current findings suggested specificities for these variables ranged from 0.80–0.90 and 0.90–0.92, respectively. Although the specificity for C-FCR ≤ 15 was 0.98 as reported by Domen and colleagues [38], these results suggested that 20% (i.e., specificity = 0.80) of the credible group were misidentified as noncredible, suggesting that a more conservative cutoff of ≤ 14 [37] may yield fewer false positives. Additionally, these results further extend Domen and colleagues' [38] research as they revealed that specificities for C-RH and B-RH—neither of which were discussed by Domen and colleagues [38]—hovered around 0.90 (0.89–0.95). Current findings support their conclusion that various previously validated cutoff scores may avoid excessive false-positive errors in patients with MS. Nonetheless, despite broadly adequate levels of specificity for the embedded PVTs in isolation, their *sensitivities* were notably lacking, such that they likely do not have sufficient ability to detect the “signal” of noncredible performance during the neuropsychological assessment of pwMS. The embedded PVTs herein appeared to fall short of the “Larrabee limit” [28] (p. 1088), which states that sensitivities tend to hover around 0.50 while maintaining specificities of ≥ 0.90 . Such psychometric considerations may result in false negatives in clinical decision-making (i.e., concluding that a patient is providing credible performance when he or she is not).

Importantly, current findings highlight a glaring insufficiency in recommended neuropsychological procedures and care regarding the evaluation of pwMS. Of note, neither the National MS Society's recommendations for neuropsychology [1] nor the articles proposing the MACFIMS [9] and BICAMS [10] discuss the need for nor the role of performance validity assessment in MS. Only the MEMS [11] clearly identified and thoroughly discussed the continued need for this venture, though it did not explicitly recommend the inclusion of a standalone PVT in its battery. These findings indicate that embedded PVTs within the aforementioned batteries, which are certainly effective at reducing the time needed to complete neuropsychological testing, may be insensitive to noncredible performance.

In line with emerging literature on this topic [18,38], the authors strongly recommend that clinicians who perform neuropsychological evaluations with patients with MS include at least one standalone PVT in their battery and consider clinical guidelines in interpreting one or more data suggesting noncredible performance [53]. Performance on PVTs may account for extraordinary amounts of variance in neuropsychological test scores [54–56].

The reliance on invalid neuropsychological data may result in incorrect clinical conclusions with implications regarding patients' continued medical care (e.g., receiving unwarranted medical treatment/diagnosis) [24,57,58] or extra-medical considerations (e.g., in the context of disability application or medicolegal assessments) [59,60].

From a clinical perspective, ideally, the choice of which PVT(s) to use will be made on the basis of the best available evidence and clinical judgment according to each patient's presenting problems and neuropsychologist's expertise. As an aside, the authors believe that the VSVT may be appropriate given its routine use in previous MS literature [18,21] and reported robustness against bona fide deficits in processing speed, working memory, and memory [42], each of which may be jeopardized in patients with MS.

Limitations

The present study is not without limitations. First, noncredible performance was defined by the below criterion scores on a single, standalone PVT. Although the PVT chosen is widely used clinically and is psychometrically robust at detecting the "signal" of non-credibility without being negatively affected by bona fide weaknesses in aspects of processing speed, working memory, and attention [42], this methodological decision nonetheless may be critiqued in light of work highlighting multivariate models that rely on more than one datum to detect non-credibility [52,61]. Future work should seek to include enough psychometrically appropriate standalone and embedded PVTs to be consistent with these recommendations. Furthermore, these results focus exclusively on the CVLT-II and not the newly published CVLT-3 [62], which is broadly similar in nature albeit with differences most meaningfully in its forced-choice recognition items (which are not described in any detail herein for test security). Furthermore, the PVTs of interest (both standalone and embedded) utilized memory-based paradigms. It may be that PVTs tapping non-memory-based abilities (e.g., attention, visuospatial ability, language) provide improved classification statistics and deserve investigation. Additionally, given the multiplicity of variables that may play a role in noncredible presentations (e.g., psychiatric symptomatology, presence of secondary gain), future research is strongly encouraged to parse out the relevant dimensions that may meaningfully contribute to and/or explain noncredible performance on cognitive tests. Finally, while differences in gender and years of education between credible and noncredible groups were not considered to be clinically meaningful in this study (as is consistent with prior research with similar aims) [45–47], future work may seek to consider the impact of baseline patient characteristics on neuropsychological test scores and, by extension, performance validity in pwMS, as some have suggested that level of education and/or disease subtype may differentially influence aspects of test performance in MS [63].

5. Conclusions

In all, the present paper is the first to explore the possible clinical utility of four embedded PVTs in the CVLT-II and BVMT-R specifically within a sample of pwMS. Current findings revealed fairly poor classification statistics for these variables and highlight the need for a psychometrically sound assessment of performance validity in this population. The authors encourage clinicians who work with pwMS in a neuropsychological context to not rely solely on the available embedded metrics explored herein, but rather to routinely utilize standalone PVT(s) to maximize the interpretive quality of their cognitive data. Furthermore, the authors recommend that researchers explore the utility of various types of PVTs in pwMS.

Author Contributions: The contributions of authors are as follows: J.W.L.: contributed to the concept, methodology, data curation, formal analysis, interpretation, drafting, and critical revision of the manuscript; Z.C.M.: contributed to the interpretation, drafting, and critical revision of the manuscript; R.G.: contributed to the concept, data curation, interpretation, critical revision, and supervision of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Data collection and analysis was conducted according to the guidelines of the Declaration of Helsinki and was approved by the Institutional Review Board of Cleveland Clinic Foundation (#17-1719, approved 20 December 2020) as part of a larger study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study may be available on reasonable, approved request from the corresponding author.

Acknowledgments: The authors acknowledge MDPI, *Neurology International*, and the anonymous reviewers for their time and expertise.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kalb, R.; Beier, M.; Benedict, R.H.; Charvet, L.; Costello, K.; Feinstein, A.; Gingold, J.; Goverover, Y.; Halper, J.; Harris, C.; et al. Recommendations for Cognitive Screening and Management in Multiple Sclerosis Care. *Mult. Scler. J.* **2018**, *24*, 1665–1680. [[CrossRef](#)]
- Macaron, G.; Ontaneda, D. Diagnosis and Management of Progressive Multiple Sclerosis. *Biomedicines* **2019**, *7*, 56. [[CrossRef](#)]
- Patten, S.B.; Marrie, R.A.; Carta, M.G. Depression in Multiple Sclerosis. *Int. Rev. Psychiatry* **2017**, *29*, 463–472. [[CrossRef](#)] [[PubMed](#)]
- Petracca, M.; Pontillo, G.; Moccia, M.; Carotenuto, A.; Cocozza, S.; Lanzillo, R.; Brunetti, A.; Brescia Morra, V. Neuroimaging Correlates of Cognitive Dysfunction in Adults with Multiple Sclerosis. *Brain Sci.* **2021**, *11*, 346. [[CrossRef](#)] [[PubMed](#)]
- Torkildsen, Ø.; Myhr, K.-M.; Bø, L. Disease-modifying Treatments for Multiple Sclerosis—A Review of Approved Medications. *Eur. J. Neurol.* **2016**, *23*, 18–27. [[CrossRef](#)]
- Giovannoni, G. Disease-Modifying Treatments for Early and Advanced Multiple Sclerosis: A New Treatment Paradigm. *Curr. Opin. Neurol.* **2018**, *31*, 233–243. [[CrossRef](#)] [[PubMed](#)]
- Benedict, R.H.B.; DeLuca, J.; Enzinger, C.; Geurts, J.J.G.; Krupp, L.B.; Rao, S.M. Neuropsychology of Multiple Sclerosis: Looking Back and Moving Forward. *J. Int. Neuropsychol. Soc.* **2017**, *23*, 832–842. [[CrossRef](#)] [[PubMed](#)]
- Stimmel, M.; Shagalow, S.; Seng, E.K.; Portnoy, J.G.; Archetti, R.; Mendelowitz, E.; Sloan, J.; Botvinick, J.; Glukhovskiy, L.; Foley, F.W. Short Report: Adherence to Neuropsychological Recommendations in Patients with Multiple Sclerosis. *Int. J. MS Care* **2019**, *21*, 70–75. [[CrossRef](#)]
- Benedict, R.H.B.; Cookfair, D.; Gavett, R.; Gunther, M.; Munschauer, F.; Garg, N.; Weinstock-Guttman, B. Validity of the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS). *J. Int. Neuropsychol. Soc.* **2006**, *12*, 549–558. [[CrossRef](#)]
- Langdon, D.; Amato, M.; Boringa, J.; Brochet, B.; Foley, F.; Fredrikson, S.; Härmäläinen, P.; Hartung, H.-P.; Krupp, L.; Penner, I.; et al. Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). *Mult. Scler. J.* **2012**, *18*, 891–898. [[CrossRef](#)]
- Merz, Z.C.; Wright, J.D.; Vander Wal, J.S.; Gfeller, J.D. A Factor Analytic Investigation of the Mercy Evaluation of Multiple Sclerosis. *Clin. Neuropsychol.* **2018**, *32*, 1431–1453. [[CrossRef](#)] [[PubMed](#)]
- Larrabee, G.J. Performance Validity and Symptom Validity in Neuropsychological Assessment. *J. Int. Neuropsychol. Soc.* **2012**, *18*, 625–630. [[CrossRef](#)]
- Rohling, M.L.; Demakis, G.J. 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? *Clin. Neuropsychol.* **2010**, *24*, 119–136. [[CrossRef](#)] [[PubMed](#)]
- Bayard, S.; Adnet Bonte, C.; Nibbio, A.; Moroni, C. Exagération de symptômes mnésiques hors contexte médical légal chez un patient atteint de sclérose en plaques. *Rev. Neurol.* **2007**, *163*, 730–733. [[CrossRef](#)]
- Graver, C.; Green, P. Misleading Conclusions about Word Memory Test Results in Multiple Sclerosis (MS) by Loring and Goldstein (2019). *Appl. Neuropsychol. Adult* **2020**, 1–9. [[CrossRef](#)]
- Loring, D.W.; Goldstein, F.C. If Invalid PVT Scores Are Obtained, Can Valid Neuropsychological Profiles Be Believed? *Arch. Clin. Neuropsychol.* **2019**, *34*, 1192–1202. [[CrossRef](#)]
- Loring, D.W.; Meador, K.J.; Goldstein, F.C. Valid or Not: A Critique of Graver and Green. *Appl. Neuropsychol. Adult* **2020**, 1–4. [[CrossRef](#)]
- Galioto, R.; Dhima, K.; Berenholz, O.; Busch, R. Performance Validity Testing in Multiple Sclerosis. *J. Int. Neuropsychol. Soc.* **2020**, *26*, 1028–1035. [[CrossRef](#)]
- Nauta, I.; Bertens, D.; van Dam, M.; Huiskamp, M.; Driessen, S.; Geurts, J.; Uitdehaag, B.; Fasotti, L.; Hulst, H.; de Jong, B.; et al. Performance Validity in Outpatients with Multiple Sclerosis and Cognitive Complaints. *Mult. Scler. J.* **2021**, 135245852110257. [[CrossRef](#)] [[PubMed](#)]
- Martin, P.K.; Schroeder, R.W. Base Rates of Invalid Test Performance Across Clinical Non-Forensic Contexts and Settings. *Arch. Clin. Neuropsychol.* **2020**, *35*, 717–725. [[CrossRef](#)] [[PubMed](#)]

21. Suchy, Y.; Chelune, G.; Franchow, E.I.; Thorgusen, S.R. Confronting Patients about Insufficient Effort: The Impact on Subsequent Symptom Validity and Memory Performance. *Clin. Neuropsychol.* **2012**, *26*, 1296–1311. [[CrossRef](#)]
22. Klimczak, N.J.; Donovic, P.J.; Burrig, R. The Malingering of Multiple Sclerosis and Mild Traumatic Brain Injury. *Brain Inj.* **1997**, *11*, 343–352. [[CrossRef](#)] [[PubMed](#)]
23. Bigler, E.D. Effort, Symptom Validity Testing, Performance Validity Testing and Traumatic Brain Injury. *Brain Inj.* **2014**, *28*, 1623–1638. [[CrossRef](#)]
24. Sweet, J.J.; Heilbronner, R.L.; Morgan, J.E.; Larrabee, G.J.; Rohling, M.L.; Boone, K.B.; Kirkwood, M.W.; Schroeder, R.W.; Suhr, J.A. Conference Participants American Academy of Clinical Neuropsychology (AACN) 2021 Consensus Statement on Validity Assessment: Update of the 2009 AACN Consensus Conference Statement on Neuropsychological Assessment of Effort, Response Bias, and Malingering. *Clin. Neuropsychol.* **2021**, *35*, 1053–1106. [[CrossRef](#)]
25. Young, G. Resource Material for Ethical Psychological Assessment of Symptom and Performance Validity, Including Malingering. *Psychol. Inj. Law* **2014**, *7*, 206–235. [[CrossRef](#)]
26. Lockhart, J.; Satya-Murti, S. Symptom Exaggeration and Symptom Validity Testing in Persons with Medically Unexplained Neurologic Presentations. *Neurol. Clin. Pract.* **2015**, *5*, 17–24. [[CrossRef](#)]
27. Greher, M.R.; Wodushek, T.R. Performance Validity Testing in Neuropsychology: Scientific Basis and Clinical Application—A Brief Review. *J. Psychiatr. Pract.* **2017**, *23*, 134–140. [[CrossRef](#)] [[PubMed](#)]
28. Erdodi, L.A.; Hurtubise, J.L.; Charron, C.; Dunn, A.; Enache, A.; McDermott, A.; Hirst, R.B. The D-KEFS Trails as Performance Validity Tests. *Psychol. Assess.* **2018**, *30*, 1082–1095. [[CrossRef](#)] [[PubMed](#)]
29. Delis, D.C.; Kaplan, E.; Kramer, J.H.; Ober, R.A. *The California Verbal Learning Test*, 2nd ed.; Adult Version: A Comprehensive Assessment of Verbal Learning and Memory; The Psychological Corporation: San Antonio, TX, USA, 2001.
30. Benedict, R.H. *Brief Visuospatial Memory Test-Revised: Professional Manual*; PAR: Odessa, FL, USA, 1997.
31. Donders, J. A Confirmatory Factor Analysis of the California Verbal Learning Test—Second Edition (CVLT-II) in the Standardization Sample. *Assessment* **2008**, *15*, 123–131. [[CrossRef](#)]
32. Benedict, R.H.B.; Schretlen, D.; Groninger, L.; Dobraski, M.; Shpritz, B. Revision of the Brief Visuospatial Memory Test: Studies of Normal Performance, Reliability, and Validity. *Psychol. Assess.* **1996**, *8*, 145–153. [[CrossRef](#)]
33. Olsen, D.H.; Schroeder, R.W.; Heinrichs, R.J.; Martin, P.K. Examination of Optimal Embedded PVTs within the BVMT-R in an Outpatient Clinical Sample. *Clin. Neuropsychol.* **2019**, *33*, 732–742. [[CrossRef](#)]
34. Persinger, V.C.; Whiteside, D.M.; Bobova, L.; Saigal, S.D.; Vannucci, M.J.; Basso, M.R. Using the California Verbal Learning Test, Second Edition as an Embedded Performance Validity Measure among Individuals with TBI and Individuals with Psychiatric Disorders. *Clin. Neuropsychol.* **2018**, *32*, 1039–1053. [[CrossRef](#)] [[PubMed](#)]
35. Pliskin, J.I.; DeDios Stern, S.; Resch, Z.J.; Saladino, K.F.; Ovsiew, G.P.; Carter, D.A.; Soble, J.R. Comparing the Psychometric Properties of Eight Embedded Performance Validity Tests in the Rey Auditory Verbal Learning Test, Wechsler Memory Scale Logical Memory, and Brief Visuospatial Memory Test-Revised Recognition Trials for Detecting Invalid Neuropsychological Test Performance. *Assessment* **2020**, 107319112092909. [[CrossRef](#)]
36. Resch, Z.J.; Pham, A.T.; Abramson, D.A.; White, D.J.; DeDios-Stern, S.; Ovsiew, G.P.; Castillo, L.R.; Soble, J.R. Examining Independent and Combined Accuracy of Embedded Performance Validity Tests in the California Verbal Learning Test-II and Brief Visuospatial Memory Test-Revised for Detecting Invalid Performance. *Appl. Neuropsychol. Adult* **2020**, 1–10. [[CrossRef](#)] [[PubMed](#)]
37. Schwartz, E.S.; Erdodi, L.; Rodriguez, N.; Ghosh, J.J.; Curtain, J.R.; Flashman, L.A.; Roth, R.M. CVLT-II Forced Choice Recognition Trial as an Embedded Validity Indicator: A Systematic Review of the Evidence. *J. Int. Neuropsychol. Soc.* **2016**, *22*, 851–858. [[CrossRef](#)] [[PubMed](#)]
38. Domen, C.H.; Greher, M.R.; Hosokawa, P.W.; Barnes, S.L.; Hoyt, B.D.; Wodushek, T.R. Are Established Embedded Performance Validity Test Cut-Offs Generalizable to Patients with Multiple Sclerosis? *Arch. Clin. Neuropsychol.* **2020**, *35*, 511–516. [[CrossRef](#)] [[PubMed](#)]
39. Wilkinson, G.S.; Robertson, G.J. *Wide Range Achievement Test*, 4th ed.; (WRAT-4); PAR: Lutz, FL, USA, 2006.
40. Slick, D.; Hopp, G.; Strauss, E.; Thompson, G. *The Victoria Symptom Validity Test Professional Manual*; PAR: Odessa, FL, USA, 1997.
41. Resch, Z.J.; Webber, T.A.; Bernstein, M.T.; Rhoads, T.; Ovsiew, G.P.; Soble, J.R. Victoria Symptom Validity Test: A Systematic Review and Cross-Validation Study. *Neuropsychol. Rev.* **2021**, *31*, 331–348. [[CrossRef](#)]
42. Resch, Z.J.; Soble, J.R.; Ovsiew, G.P.; Castillo, L.R.; Saladino, K.F.; DeDios-Stern, S.; Schulze, E.T.; Song, W.; Pliskin, N.H. Working Memory, Processing Speed, and Memory Functioning Are Minimally Predictive of Victoria Symptom Validity Test Performance. *Assessment* **2021**, *28*, 1614–1623. [[CrossRef](#)] [[PubMed](#)]
43. Larrabee, G.J.; Berry, D.T.R. Diagnostic classification statistics and diagnostic validity of malingering assessment. In *Assessment of Malingered Neuropsychological Deficits*; Oxford University Press: New York, NY, USA, 2007; pp. 14–26.
44. Lwoga, E.T.; Questier, F. Open Access Behaviours and Perceptions of Health Sciences Faculty and Roles of Information Professionals. *Health Inf. Libr. J.* **2015**, *32*, 37–49. [[CrossRef](#)]
45. Ord, J.S.; Greve, K.W.; Bianchini, K.J.; Aguerrevere, L.E. Executive Dysfunction in Traumatic Brain Injury: The Effects of Injury Severity and Effort on the Wisconsin Card Sorting Test. *J. Clin. Exp. Neuropsychol.* **2010**, *32*, 132–140. [[CrossRef](#)]
46. Shura, R.D.; Miskey, H.M.; Rowland, J.A.; Yoash-Gantz, R.E.; Denning, J.H. Embedded Performance Validity Measures with Postdeployment Veterans: Cross-Validation and Efficiency with Multiple Measures. *Appl. Neuropsychol. Adult* **2016**, *23*, 94–104. [[CrossRef](#)] [[PubMed](#)]

47. Sawyer, R.J.; Testa, S.M.; Dux, M. Embedded Performance Validity Tests within the Hopkins Verbal Learning Test—Revised and the Brief Visuospatial Memory Test—Revised. *Clin. Neuropsychol.* **2017**, *31*, 207–218. [[CrossRef](#)]
48. Lenhard, W.; Lenhard, A. Computation of Effect Sizes. 2017. Available online: https://www.psychometrica.de/effect_size.html (accessed on 3 August 2021).
49. Erdodi, L.A.; Abeare, C.A.; Medoff, B.; Seke, K.R.; Sagar, S.; Kirsch, N.L. A Single Error Is One Too Many: The Forced Choice Recognition Trial of the CVLT-II as a Measure of Performance Validity in Adults with TBI. *Arch. Clin. Neuropsychol.* **2018**, *33*, 845–860. [[CrossRef](#)] [[PubMed](#)]
50. Haanes, G.G.; Kirkevold, M.; Hofoss, D.; Eilertsen, G. Discrepancy between Self-Assessments and Standardised Tests of Vision and Hearing Abilities in Older People Living at Home: An ROC Curve Analysis. *J. Clin. Nurs.* **2015**, *24*, 3380–3388. [[CrossRef](#)] [[PubMed](#)]
51. Snyder, C.F.; Blackford, A.L.; Brahmer, J.R.; Carducci, M.A.; Pili, R.; Stearns, V.; Wolff, A.C.; Dy, S.M.; Wu, A.W. Needs Assessments Can Identify Scores on HRQOL Questionnaires That Represent Problems for Patients: An Illustration with the Supportive Care Needs Survey and the QLQ-C30. *Qual. Life Res.* **2010**, *19*, 837–845. [[CrossRef](#)]
52. Lace, J.W.; Grant, A.F.; Ruppert, P.; Kaufman, D.A.S.; Teague, C.L.; Lowell, K.; Gfeller, J.D. Detecting Noncredible Performance with the Neuropsychological Assessment Battery, Screening Module: A Simulation Study. *Clin. Neuropsychol.* **2021**, *35*, 572–596. [[CrossRef](#)] [[PubMed](#)]
53. Odland, A.P.; Lammy, A.B.; Martin, P.K.; Grote, C.L.; Mittenberg, W. Advanced Administration and Interpretation of Multiple Validity Tests. *Psychol. Inj. Law* **2015**, *8*, 46–63. [[CrossRef](#)]
54. Gorissen, M.; Sanz, J.C.; Schmand, B. Effort and Cognition in Schizophrenia Patients. *Schizophr. Res.* **2005**, *78*, 199–208. [[CrossRef](#)]
55. Green, P.; Rohling, M.L.; Lees-Haley, P.R.; Iii, L.M.A. Effort Has a Greater Effect on Test Scores than Severe Brain Injury in Compensation Claimants. *Brain Inj.* **2001**, *15*, 1045–1060. [[CrossRef](#)] [[PubMed](#)]
56. Meyers, J.E.; Volbrecht, M.; Axelrod, B.N.; Reinsch-Boothby, L. Embedded Symptom Validity Tests and Overall Neuropsychological Test Performance. *Arch. Clin. Neuropsychol.* **2011**, *26*, 8–15. [[CrossRef](#)]
57. Zasler, N.D.; Bender, S.D. Validity Assessment in Traumatic Brain Injury Impairment and Disability Evaluations. *Phys. Med. Rehabil. Clin. N. Am.* **2019**, *30*, 621–636. [[CrossRef](#)] [[PubMed](#)]
58. Fuermaier, A.B.M.; Tucha, O.; Koerts, J.; Tucha, L.; Thome, J.; Faltraco, F. Feigning ADHD and Stimulant Misuse among Dutch University Students. *J. Neural Transm.* **2021**, *128*, 1079–1084. [[CrossRef](#)] [[PubMed](#)]
59. Sherman, E.M.S.; Slick, D.J.; Iverson, G.L. Multidimensional Malingering Criteria for Neuropsychological Assessment: A 20-Year Update of the Malingered Neuropsychological Dysfunction Criteria. *Arch. Clin. Neuropsychol.* **2020**, *35*, 735–764. [[CrossRef](#)]
60. Merten, T. Logical Paradoxes and Paradoxical Constellations in Medicolegal Assessment. *Psychol. Inj. Law* **2017**, *10*, 264–273. [[CrossRef](#)]
61. Erdodi, L.A. Aggregating Validity Indicators: The Salience of Domain Specificity and the Indeterminate Range in Multivariate Models of Performance Validity Assessment. *Appl. Neuropsychol. Adult* **2019**, *26*, 155–172. [[CrossRef](#)]
62. Delis, D.C.; Kramer, J.H.; Kaplan, E.; Ober, R.A. *The California Verbal Learning Test*, 3rd ed.; The Psychological Corporation: San Antonio, TX, USA, 2017.
63. Estrada-López, M.; García-Martín, S.; Cantón-Mayo, I. Cognitive Dysfunction in Multiple Sclerosis: Educational Level as a Protective Factor. *Neurol. Int.* **2021**, *13*, 335–342. [[CrossRef](#)] [[PubMed](#)]