



Research Article

Validation of the Sydney Language Battery naming subtest and utility of latency analysis in characterizing language impairment in multiple sclerosis

Annabel Hudson¹ , Stefanie Roberts^{2,3}, Charles B. Malpas^{1,2,3} , Genevieve Rayner^{1,4,5} and Fiore D'Aprano¹

¹Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Australia, ²Department of Medicine, Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia, ³Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia, ⁴Florey Institute of Neuroscience and Mental Health, Heidelberg, Australia and ⁵Department of Clinical Neuropsychology, Austin Health, Heidelberg, Australia

Abstract

Background: Language deficits are frequently described by patients with multiple sclerosis (MS); however, objective characterization remains somewhat limited due to its omission from standard MS cognitive evaluation and the inconsistent findings that arise from current language measures. **Objective:** To establish alternative approaches to characterizing single-word level language in MS, this study (i) validates the Sydney Language Battery (SYDBAT) visual confrontation naming subtest and (ii) examines the insights provided by examining naming errors and latencies. **Methods:** 40 MS patients from Royal Melbourne Hospital's Cognitive Neuroimmunology Clinic and 40 matched controls completed a series of neuropsychological tests, including the SYDBAT and 'gold standard' confrontation naming task, the Boston Naming Test (BNT). Error types and latencies on the SYDBAT were extracted from assessment audio recordings. **Results:** SYDBAT and BNT scores were highly correlated ($r = 0.81, p < .001$) and these tasks reported comparable receiver operating characteristic curves ($p = .091$). Latency analysis captured lexical retrieval difficulties, with patients displaying significantly longer mean latencies than controls on the SYDBAT ($p = .012, \beta = 0.54$). **Conclusions:** These findings support the validity of the SYDBAT and value of the latency analysis in characterizing language impairment in MS. Use of the SYDBAT and latency considerations contribute to a broader assessment with a briefer administration time compared to gold-standard evaluation. The study thereby offers clinicians an enhanced toolkit to more effectively and appropriately evaluate language functioning and supplement standard cognitive evaluation in this population.

Keywords: Confrontation naming; lexical retrieval; cognition; neuropsychology; neuroimmunology; psychometrics

(Received 19 March 2025; final revision 15 September 2025; accepted 17 September 2025)

Statement of Research Significance

Research Question(s) or Topic(s): Exploration of alternative approaches to characterize single-word level language functioning in multiple sclerosis. **Main Findings:** The Sydney Language Battery was at least, if not more, effective than the 'gold standard' Boston Naming Test in assessing language function in multiple sclerosis, despite requiring briefer administration times. This measure additionally captured more impairment than standard multiple sclerosis cognitive evaluation alone. Longer mean latencies were also identified on this task for patients as compared to controls, indicative of difficulties during word retrieval. **Study Contributions:** This study presents the Sydney Language Battery as a valid tool for assessing single-word level language function in multiple sclerosis and highlights its clinical value as an adjunct to standard cognitive evaluation. Latency analysis is additionally presented as a valuable approach to extend language characterization, contributing to a broader assessment while maintaining brief administration. This study thereby offers clinicians an enhanced toolkit to more effectively and completely evaluate cognitive functioning in multiple sclerosis.

Introduction

Language impairment is commonly reported, but under-recognized in multiple sclerosis (MS). Seventy-five percent of MS patients self-report language difficulties, particularly problems with single word-

level language production such as word-finding difficulties, which significantly impede quality of life (El-Wahsh et al., 2020). Evidence of objective single-word level language impairment has additionally been captured on tasks of confrontation naming (Beatty & Monson, 1989; Kujala et al., 1996; Tallberg & Bergendal, 2009), though, its

Corresponding author: Fiore D'Aprano; Email: fiore.daprano@unimelb.edu.au

Cite this article: Hudson A., Roberts S., Malpas C.B., Rayner G., & D'Aprano F. Validation of the Sydney Language Battery naming subtest and utility of latency analysis in characterizing language impairment in multiple sclerosis. *Journal of the International Neuropsychological Society*, 1–8, <https://doi.org/10.1017/S1355617725101513>

© The Author(s), 2025. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

characterization remains somewhat limited due to its omission from standard MS cognitive evaluation and the inconsistent findings that arise from current language measures (Brandstadter et al., 2020; Jennekens-Schinkel et al., 1990; Langdon et al., 2012). This study therefore aims to enhance the characterization of language functioning in MS by exploring the clinical utility of alternative instruments and examining additional features of confrontation naming which may better capture the impairment than accuracy alone.

Cognitive evaluation in MS typically relies on brief cognitive batteries that primarily target processing speed, attention, memory, and executive functioning (Benedict et al., 2002; Grzegorski & Losy, 2017; Langdon et al., 2012; Rao et al., 1991). The Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS; Langdon et al., 2012) in particular is currently recommended as the clinical benchmark for cognitive evaluation in this population and assesses processing speed, supraspan verbal memory, and visual memory (Maltby et al., 2020). The omission of language from the BICAMS highlights the under recognition of impairment in this domain and suggests the potential limitations of relying solely upon this measure to sufficiently characterize cognitive impairment in MS.

Current measures of single word-level language production in MS, however, are narrowly defined and tend to exhibit inconsistencies. Confrontation naming tasks are widely endorsed measures of single-word level language production (Strauss et al., 2006). Existing tasks, such as the current 'gold standard' Boston Naming Test (BNT; Kaplan et al., 1983) have, however, produced inconsistent results in assessing MS naming functioning. While some studies report poorer BNT scores for patients compared to controls (Beatty & Monson, 1989; Kujala et al., 1996; Tallberg & Bergendal, 2009), others report no significant group difference (Beatty et al., 1989; Jennekens-Schinkel et al., 1990; Olivares et al., 2005). Scores on the BNT have also been found to poorly correlate with patient-reported word-finding difficulties (Brandstadter et al., 2020). These inconsistencies suggest existing assessment methods used in the MS population may not completely capture single-word level language impairment. Validation of alternative methods capable of providing a more consistent and complete characterization of single-word level language function in this population therefore appears necessary.

An alternative tool that may characterize language functioning in MS is the naming subtest of the SYDBAT; an Australian visual confrontation naming task (Savage et al., 2013). The SYDBAT naming subtest, simply referred to as the SYDBAT henceforth, has previously demonstrated its suitability in assessing naming impairment, reporting high convergent validity with the short form BNT in its original primary progressive aphasia population (Savage et al., 2013). Unlike the increasingly outdated BNT (Beatty et al., 2017), this task consists of a set of contemporary items that may more appropriately and effectively assess naming function in a modern-day population (Savage et al., 2013). The SYDBAT additionally has a briefer administration time, making it a potentially superior alternative to the BNT as an adjunct to existing brief MS cognitive batteries (Savage et al., 2013). Specific validation in this population is necessary to accurately ascertain suitability.

Examining alternative features of confrontation naming beyond accuracy scores may also be valuable in enhancing the characterization of MS language. Two viable approaches include error and latency analyses, with specific error types indicating breakdown in distinct cognitive–linguistic processes (Lethlean & Murdoch, 1994), and response times capturing more subtle

difficulties during the retrieval processes (Goodglass et al., 1984). Previous studies employing these techniques in MS reveal high rates of semantic errors indicative of impaired semantic selection (De Dios Pérez et al., 2020; Kujala et al., 1996; Lethlean & Murdoch, 1994), and longer mean latencies reflecting lexical retrieval difficulties (Beatty & Monson, 1989; De Dios Pérez et al., 2020; Kujala et al., 1996). This retrieval difficulty was further observed irrespective of overall confrontation naming score (Beatty & Monson, 1989). These techniques may therefore extend the characterization of single-word level language beyond that of total scores alone, thereby providing valuable insight in developing a more complete assessment of the cognitive–linguistic impairment.

This study will therefore explore alternative approaches for assessing single-word level language to improve the characterization of language functioning in MS. It aims to achieve this by (1) validating the SYDBAT against the 'gold standard' confrontation naming task, the BNT, and considering its value as an adjunct to standard MS cognitive evaluation, the BICAMS, and (2) extending the analysis of confrontation naming performance beyond accuracy to investigate potential insights provided by error and latency analysis.

Method

Participants

This cross-sectional study employed a sample drawn from a broader dataset collected as part of an ongoing project at the Royal Melbourne Hospital (RMH), Melbourne, Australia. All participants were referred to the Cognitive Neuroimmunology Clinic for specialist cognitive opinion. Inclusion criteria for the MS group were: (1) a diagnosis of MS, (2) aged 18 years or over, (3) English proficiency, (4) able to provide their own informed consent, (5) sufficient vision and audition to perceive test materials, and 6) no history of other neurological conditions likely to affect cognition (i.e. stroke, head injury, dementia). In addition to these broader project criteria, participants included in this study were also required to have complete data for the relevant neuropsychological measures. Age-, sex-, and education-matched controls were recruited via community snowball sampling by liaising with friends and family of patients and clinicians. Control eligibility criteria were as above, excluding criteria 1). While not a formal exclusion criterion, it is also noted that specific developmental language or learning disorders were not reported or identified in this sample. Ethics approval was obtained from the Human Research Ethics Committee of the Royal Melbourne Hospital, Melbourne (2020.240 RMH67046), and all research was conducted in accordance with ethical standards of the 1964 Declaration of Helsinki. All participants provided written informed consent.

Setting and procedures

The appointment was conducted either face-to-face at the RMH Cognitive Neuroimmunology Clinic or via telehealth using Zoom (Zoom Video Communications Incorporated, 2016). The complete assessment comprised a clinical interview, a comprehensive neuropsychological examination conducted by a neuropsychologist, and a set of self-report questionnaires relating to health, mood, and cognitive complaint. The battery was conducted over a single session and took approximately 1.5 to 2 hours to complete. Tasks such as the SYDBAT, which were not originally designed for remote delivery, were adapted into a pre-prepared presentation file

Table 1. Overview of relevant measures

Measure	Description
Sydney Language Battery (Naming Subtest)	30-item visual confrontation naming test assessing single-word level language production.
Boston Naming Test	60-item visual confrontation naming test assessing single-word level language production.
BICAMS	
Symbol Digits Modality Test	Processing speed test.
California Verbal Learning Test – Second Edition	List A total trials 1 – 5. Supraspan verbal memory test.
Brief Visuospatial Memory Test-Revised	Total recall. Visual learning and memory test.

Note. BICAMS = Brief International Cognitive Assessment for Multiple Sclerosis.

that would display each stimulus item sequentially to ensure standardization across participants. For appointments conducted via Zoom, the examiner would share their screen to present these materials consistently. Demographic data were obtained from medical records for MS patients and via a pre-study online questionnaire for controls.

Measures

From the broader dataset, a subset of measures relevant to the present study were extracted, as outlined Table 1.

Audio analysis

Audio recordings were obtained for all participant assessments. Audio output for face-to-face sessions was recorded using a Yeti microphone or the Voice Memo application (Apple Inc., 2023). For telehealth assessments, audio output was recorded directly from the Zoom session (Zoom Video Communications Incorporated, 2016) for control participants and via the Voice Memo application for patients (Apple Inc., 2023). Recordings of the SYDBAT were extracted to undergo latency and error analysis conducted by a single-blinded researcher.

Latency extraction. The Audacity software (Audacity Team, 2024) was used to extract spontaneous latencies for each completed SYDBAT item for all participants. Latencies were obtained by measuring the time from a beep that corresponded with picture onset to the generation of a spontaneous response (*i.e.* an uncued response), excluding filler words (*e.g.* ‘um’ or ‘that’s a-’).

Error coding. Errors for all spontaneous responses were coded according to an error classification system (Table 2), adapted from previous error analysis guidelines (Kohn & Goodglass, 1985; Kujala et al., 1996; Lethlean & Murdoch, 1994). The coding approach was discussed and agreed upon by the researcher and three clinical neuropsychologists before coding commenced. To achieve coding agreement, any ambiguous error codes were resolved via a consensus decision discussion between three clinical neuropsychologists. This applied to 96 of the 441 errors.

Statistical analyses

The R statistical software (v4.4.0; R Core Team, 2024) was used for all analyses. More detailed descriptions of the analyses are provided in Appendix A.

Group differences: group characteristics and cognitive measure scores

Group differences for continuous sample characteristics and the BICAMS subtests were computed using a Welch’s independent

samples *t*-test and Cohen’s *d* for effect size. Group differences for discrete variables were analyzed using χ^2 tests of independence and phi-coefficient for effect size. For SYDBAT and BNT scores, group differences were computed using general linear models (GLM), with consult modality (face-to-face or telehealth), age, and education included as covariates. Parameters were extracted with 95% confidence intervals, using standardized beta coefficients for effect size.

Aim one: validating the SYDBAT for confrontation naming evaluation

Validation of the SYDBAT against the BNT. Spearman’s correlation coefficients (*r*) were computed to investigate the association between the SYDBAT and BNT scores. Receiver operating characteristic (ROC) curves were produced for the raw and standardized scores to examine diagnostic performance. The area under the curve (AUC) was extracted with 95% confidence and compared using bootstrapping. Hierarchical logistic regression models were estimated to assess additional insight into group membership (patient or control) provided by each confrontation naming measure, over and above the other. A logistic regression model was first fit with the BNT raw score as the sole predictor of diagnostic group. A second model was then fit, with the SYDBAT raw score included as an additional predictor. The difference in residual deviance of the two models was assessed to evaluate model improvement. This analysis was then repeated in reverse, with the initial model including only the SYDBAT score, followed by a second model incorporating both SYDBAT and BNT scores.

Clinical utility as an adjunct to BICAMS. To investigate the value of the SYDBAT as an adjunct to standard cognitive evaluation in MS, the concordance between the rates of impairment captured by the SYDBAT and that of the BICAMS was assessed. Patients’ cognitive status was classified as ‘cognitively impaired’ or ‘cognitively intact’ following typical BICAMS protocol (impaired = scored 1.5 SD below the standardized control mean on one or more of the three subtests; Dusankova et al., 2012). The number of patients in each cognitive status group classified as ‘language impaired’ according to the SYDBAT (*i.e.* 1.5 SD below the standardized control mean) was established and compared using an χ^2 test of independence and phi-coefficient for effect size.

Aim two: exploring insight provided by error and latency analysis

Error analysis. Robust general linear mixed models (GLMM) were estimated using restricted maximum likelihood to investigate group differences in *overall error types* (semantic, phonological, visual, mis-focus, non-response, unrelated response) on the SYDBAT. Error frequency was specified as the dependent variable, while diagnostic group and error type were specified as

Table 2. Classification system for error coding of the confrontation naming tasks

Error Type	Description	Example
Semantic		
a. Semantic-visual	Incorrect response, semantically and visually related to the target word.	"hippopotamus" for rhinoceros
b. Semantic-associative	Incorrect response, semantically, but not visually, related to the target word.	"giraffe" for hippopotamus
c. Superordinate	Class or category to which target belongs.	"animal" for rhinoceros
d. Circumlocutory response	Multiword response showing accurate identification of the target by describing its function, action or attributes, including semantic negation.	"to listen to a heart" for stethoscope, or "not a hippopotamus" for rhinoceros
Phonological		
a. Phonemic	Phonemically based error.	"pi-gola" for pagoda
b. Non-word	Non-identifiable response.	"slyso" for balaclava
Visual error	Incorrect response, semantically unrelated but visually similar to the target.	"stone" for potato
Mis-focus	Misinterpretation of intended focus.	"waist" for leotard
Non-response	No verbal response, general comment or question, incomplete noun phrase with a modifier, or less than a syllable.	"that's a hard one" or "a princess's -" for tiara
Unrelated response	Incorrect response unrelated semantically, visually, or phonemically to target word.	"flute" for shuttlecock

independent variables, along with an interaction term between the two. A random intercept was specified for each participant to account for within-subjects dependence. Consult modality, age, and education were entered as covariates. Parameters were extracted with 95% confidence intervals, using partial omega squared for effect size. Simple effects analyses were conducted using GLMs to assess group differences in each error type. Parameters were extracted with 95% confidence intervals, using standardized beta coefficients for effect size. If a significant group difference was identified for semantic errors, this analysis would be repeated for the *semantic error subtypes*.

Latency analysis. Mean latency, standard deviation of the latency, and frequency of 'extreme' latencies (i.e. latencies 1.96 SDs longer than the mean latency) for all spontaneous responses, for correct spontaneous responses, and for incorrect spontaneous responses on the SYDBAT were extracted. These variables were selected to capture the typical, positively skewed response time distribution under various retrieval conditions to form a rounded latency profile. Ex-Gaussian parameters were considered, however, were not possible due to the relatively small number of items per participant, producing unstable estimates. Group differences for each variable under each response condition (e.g. mean latency of correct spontaneous responses) were assessed via GLMs. Parameters were extracted with 95% confidence intervals, using standardized beta coefficients for effect size.

Results

The final sample ($n = 80$) comprised 40 MS patients and 40 healthy controls. Demographic, clinical, and cognitive characteristics are outlined in Table 3. No significant group differences were found for age, gender, and years of education. A moderate significant association was found for consult modality, $X^2(1) = 5.08$, $p = .024$, $\phi = 0.25$, with patients more likely to be seen face-to-face than controls (Rea & Parker, 1992). There was however no evidence of consult modality impacting performance (Appendix B). Patient scores were significantly lower than corresponding control group scores for all cognitive measures. Of note, effect sizes for SYDBAT and BNT scores between groups were large and medium, respectively (Cohen, 1988).

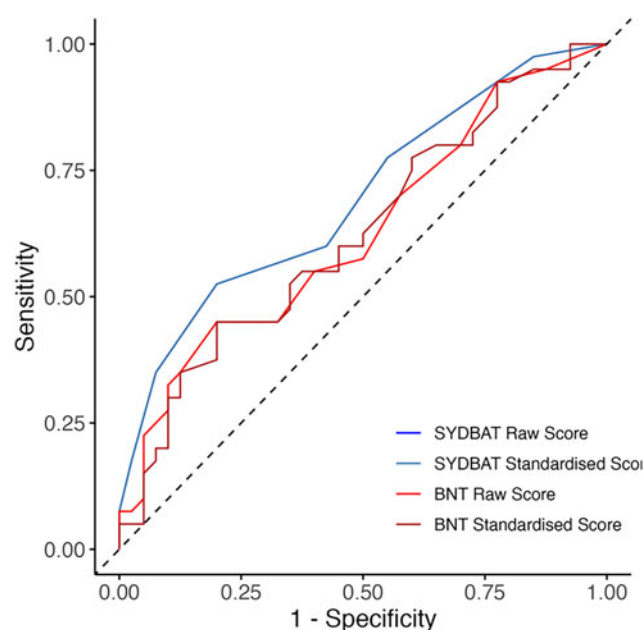


Figure 1. Receiver operating characteristic curves for confrontation naming scores. Note. SYDBAT = Sydney language battery (Naming subtest), BNT = Boston naming test. Dashed line represents an area under the curve of 0.5.

Validating the SYDBAT for confrontation naming evaluation

Validation of the SYDBAT against the BNT

Correlation coefficients were interpreted following the conventions of Cohen (1988). A large positive correlation was identified between SYDBAT and BNT, $r = 0.81$, $p < .001$. Diagnostic performance was assessed by computing ROC curves for the raw and standardized scores, as shown in Figure 1. AUC for both raw and standardized SYDBAT scores were significant, at 0.69 (95% CI [0.58, 0.80]) and 0.69 (95% CI [0.58, 0.81]), respectively. AUC for both raw and standardized BNT scores were also significant, at 0.63 (95% CI [0.51, 0.75]) and 0.62 (95% CI [0.51, 0.75]), respectively. The raw and standardized score ROC curves for two naming tests were comparable, with non-significant differences in their AUC ($p = .091$ and $p = .082$, respectively).

Table 3. Sample demographic, clinical, and cognitive characteristics

	Control (n = 40)	Patient (n = 40)	p	Effect Size	95% CI
Demographic					
Age, <i>M(SD)</i> , range	40.01 (14.78), 19 – 71	42.55 (11.53), 20 – 65	.406	–0.11	[–0.35, 0.14]
Education (years), <i>M(SD)</i> , range	14.20 (2.02), 10 – 17	13.25 (2.25), 9 – 17	.050	0.24	[–0.01, 0.46]
Sex, Female, <i>n</i> (%)	25 (62.5)	25 (62.5)	> .99	–	–
Consult Modality, face-to-face, <i>n</i> (%)	17 (42.5)	28 (70)	.024*	0.25	–
Clinical					
Clinical Course					
RRMS, <i>n</i> (%)	–	36 (90)	–	–	–
PPMS, <i>n</i> (%)	–	3 (0.08)	–	–	–
SPMS, <i>n</i> (%)	–	1 (0.025)	–	–	–
Disease Duration (years), <i>M(SD)</i> , range	–	10.58 (8.32), 0.76 – 34.25	–	–	–
EDSS score, <i>M(SD)</i>	–	2.51 (2.08) ^a	–	–	–
Cognitive					
SYDBAT					
Raw Score, <i>M(SD)</i>	27.88 (1.60)	26.43 (2.24)	.008***†	–0.69	[–1.05, –0.25]
Z-Score, <i>M(SD)</i>	0.61 (0.76)	–0.08 (1.07)	.002***†	–0.70	[–1.09, –0.24]
BNT					
Raw Score, <i>M(SD)</i>	55.20 (3.65)	53.03 (4.81)	.024*†	–0.49	[–0.89, –0.08]
Z-Score, <i>M(SD)</i>	–0.25 (1.22)	–0.90 (1.55)	.046*†	–0.45	[–0.83, –0.01]
SDMT					
Raw Score, <i>M(SD)</i>	63.88 (9.22)	50.58 (10.26)	< .001***†	1.36	[0.87, 1.85]
Z-Score, <i>M(SD)</i>	0.37 (1.03)	–0.81 (1.01)	< .001***†	1.16	[0.68, 1.63]
CVLT					
Raw Score, <i>M(SD)</i>	58.93 (8.20)	52.68 (12.27)	< .001***†	0.60	[0.15, 1.05]
T-Score, <i>M(SD)</i>	59.73 (10.55)	54.20 (13.24)	.043*†	0.46	[0.02, 0.90]
BVMT-R					
Raw Score, <i>M(SD)</i>	58.93 (8.20) ^a	52.68 (12.17) ^b	.004***†	0.71	[0.23, 1.18]
T-Score, <i>M(SD)</i>	53.47 (8.86) ^a	46.37 (14.12) ^b	.014*†	0.61	[0.14, 1.08]

Note. RRMS = Relapsing Remitting Multiple Sclerosis, PPMS = Primary Progressive Multiple Sclerosis, SPMS = Secondary Progressive Multiple Sclerosis, EDSS = Expanded Disability Status Scale (ranging 1 – 10, with higher values indicating more severe disability). SYDBAT = Sydney Language Battery (Naming Subtest). BNT = Boston Naming Test. SDMT = Symbol Digits Modality Test. CVLT-II = California Verbal Learning Test – Second Edition. BVMT-R = Brief Visuospatial Memory Test-Revised. **p* < .05, ***p* < .01, ****p* < .001. † = significance holds on false discovery rate correction. ^a*n* = 38. ^b*n* = 35.

Table 4. SYDBAT improvement in group membership prediction over the BNT

	Residual Deviance	df	Δ Deviance	Δ df	p
Model 1 BNT	105.76	78	–	–	–
Model 2 BNT + SYDBAT	99.74	77	6.02	1	.014*

Note. SYDBAT = Sydney Language Battery (Naming Subtest). BNT = Boston Naming Test. *df* = degrees of freedom. Δ = change in variable. **p* < .05, ***p* < .01, ****p* < .001.

Hierarchical logistic regression models were estimated and model fit was compared to assess insight provided by the SYDBAT relative to the BNT. The addition of SYDBAT score significantly increased the proportion of variance explained (Table 4), indicating improved prediction of group membership (i.e. patient or control). The addition of BNT score however, did not significantly increase the proportion of variance explained (Table 5), providing no evidence for an improved prediction of group membership.

The value of the SYDBAT as an adjunct to BICAMS

Following standard BICAMS protocol (Dusankova et al., 2012), 20 patients were classified as ‘cognitively impaired,’ and 20 patients were classified as ‘cognitively intact.’ Of the 20 ‘cognitively impaired’ patients, nine (45%) were additionally classified as ‘language impaired’ according to the SYDBAT. Of the 20 ‘cognitively intact’ patients, five (25%) were incongruously identified as ‘language impaired.’ The number of ‘language impaired’ patients classified as ‘cognitively intact’ or as ‘cognitively impaired’ did not significantly differ, *p* = .285, ϕ = 0.14.

Table 5. BNT improvement in group membership prediction over the SYDBAT

	Residual Deviance	df	Δ Deviance	Δ df	p
Model 1 SYDBAT	100.12	78	–	–	–
Model 2 SYDBAT + BNT	99.74	77	0.38	1	.537

Note. SYDBAT = Sydney Language Battery (Naming Subtest). BNT = Boston Naming Test. *df* = degrees of freedom. Δ = change in variable. **p* < .05, ***p* < .01, ****p* < .001.

Latency analysis insight into language impairment

The mean latency, standard deviation of the latency, and frequency of ‘extreme’ latencies (i.e. latencies 1.96 SDs longer than the mean latency) for all spontaneous responses, for correct spontaneous responses, and for incorrect spontaneous responses for patients and controls on the SYDBAT are summarized in Table 6. A large significant group difference was identified for mean latency across all spontaneous responses (Cohen, 1988). No other significant differences were identified.

Inconclusive error analysis

Mean frequencies of each overall error type on the SYDBAT for each group are outlined in Table 7. The GLMM indicated a significant medium main effect for group, $F(1, 12.35) = 7.84$, $p = .007$, $\omega_p^2 = 0.08$, and a significant large main effect for error type, $F(5, 1073.79) = 136.29$, $p < .001$, $\omega_p^2 = 0.63$ (Kirk, 1996). The interaction effect for group and error type was additionally significant, $F(5, 34.19) = 4.34$, $p < .001$, $\omega_p^2 = 0.04$, indicating a

Table 6. Latencies for all, correct, and incorrect spontaneous responses on the SYDBAT

	Control (<i>n</i> = 40)		Patient (<i>n</i> = 40)		<i>p</i>	Effect Size	95% CI
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
All Responses							
Mean (s)	1.75	0.55	1.99	0.65	.012*†	0.54	[0.16, 0.89]
Variation (s)	1.50	1.29	1.83	1.21	.348	0.22	[−0.21, 0.68]
Extreme	1.80	0.72	1.83	0.71	.968	0.02	[−0.44, 0.52]
Correct Responses							
Mean (s)	1.45	0.38	1.52	0.47	.090	0.35	[−0.03, 0.70]
Variation (s)	0.86	0.71	1.10	0.75	.508	0.13	[−0.32, 0.56]
Extreme	1.25	0.63	1.25	0.44	.872	−0.03	[−0.45, 0.41]
Incorrect Responses							
Mean (s)	3.32	1.81	3.89	2.23	.152	0.33	[−0.11, 0.70]
Variation (s)	2.23	2.40	2.49	2.22	.632	0.12	[−0.32, 0.55]
Extreme	0.03	0.17	0.10	0.31	.156	0.16	[−0.03, 0.21]

Note. SYDBAT = Sydney Language Battery (Naming Subtest). s = seconds. **p* < .05, ***p* < .01, ****p* < .001. † = significance holds on false discovery rate correction.

Table 7. Overall error type frequencies on the SYDBAT

	Control (<i>n</i> = 40)		Patient (<i>n</i> = 40)		<i>p</i>	Effect Size	95% CI
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Semantic	3.53	2.18	5.08	3.17	.084	0.36	[−0.09, 0.79]
Phonological	0.63	0.87	1.00	1.18	.064	0.43	[−0.07, 0.90]
Visual	0.03	0.16	0	0	.684	−0.03	[−0.09, 0.00]
Mis-focus	0.25	0.44	0.33	0.47	.116	0.17	[−0.10, 0.75]
Non-Response	0.45	0.75	0.75	1.05	.092	−0.28	[−0.61, 0.04]
Unrelated Response	0.03	0.16	0.13	0.33	.288	0.06	[−0.05, 0.16]

Note. SYDBAT = Sydney Language Battery (Naming Subtest).

small effect (Kirk, 1996). To assess group differences in each *overall error type*, a simple effects analysis was conducted. As outlined in Table 7, no significant group differences were identified for any *overall error type*.

Discussion

To date, objective characterization of single-word level language in MS has been somewhat limited due to omission of language screening in standard cognitive evaluation (Langdon et al., 2012), and inconsistencies in performance on existing language measures (Brandstadter et al., 2020; Jennekens-Schinkel et al., 1990). To address this limitation, the present study explored alternative approaches to characterizing single-word level language impairment in MS patients. Consistent with the language difficulties frequently described in the MS population (El-Wahsh et al., 2020), we observed significantly poorer confrontation naming in the patient group compared to controls. The SYDBAT naming subtest total score and latency analyses demonstrated potential value in characterizing this language impairment.

Validity and clinical value of the SYDBAT

We posit that the SYDBAT is a valid and valuable tool for characterizing single-word level language functioning among MS patients. This is supported by congruent naming scores identified between the SYDBAT and ‘gold standard’ BNT, and the capacity

for each of these tasks to comparably differentiate between patients and controls. The SYDBAT scores also improved group membership prediction beyond that of the BNT alone, suggesting this tool may contribute additional insights into the impairment beyond that of the current ‘gold standard.’ Such findings align with previous accounts of the shortcomings of the BNT in characterizing language in MS and its diminishing suitability in a modern population (Beatty et al., 2017; Beatty et al., 1989; Brandstadter et al., 2020; Jennekens-Schinkel et al., 1990; Olivares et al., 2005). Regardless, comparable performance alone suggests the SYDBAT may nonetheless be a valuable alternative. The measures were, at a minimum, equally effective in identifying the impairment, however, the SYDBAT achieved this characterization with a briefer 30 items compared to the 60 items of the BNT (Kaplan et al., 1983; Savage et al., 2013). It is acknowledged that shorter versions of the BNT do exist (Mack et al., 1992), however, these remain only combinations of the increasingly outdated original items (Beatty et al., 2017). The SYDBAT therefore presents a brief, yet contemporary valid alternative to these existing single-word level tests.

The proposed clinical value of the SYDBAT relates to its potential use as an adjunct to standard cognitive assessment in the MS population. The present study found 25% of MS patients who were classified as ‘cognitively intact’ by the BICAMS were in fact ‘language impaired’ according to the SYDBAT. While the BICAMS does not claim to assess language, it is widely recommended as a cognitive screening tool in the MS population (Maltby et al., 2020). This discrepancy in the identified impairment rates highlights a critical limitation of using the BICAMS in isolation in clinical practice, as it may lead to the misclassification of language-impaired patients as cognitively intact. Under-detection of impairment not only invalidates patient concerns and difficulties but can potentially lead to missed signs of relapse and hamper informed disease management decision-making. This study therefore presents the clinical value of the SYDBAT as a brief and valid adjunct to the BICAMS battery. In the context of this clinical application, it is worth noting that an app-based version of the SYDBAT is available (Piguet, 2022), facilitating easier administration and scoring for clinicians. Such clinical recommendations do however remain somewhat constrained by the limited normative data available for the SYDBAT. Future research to develop appropriate normative data may therefore be considered to maximize the clinical applicability of this tool.

Insight provided by the latency analysis

Latency analysis additionally emerged as a valuable method for providing a more nuanced characterization of the language impairment in MS. Consistent with previous latency analyses (Beatty & Monson, 1989; De Dios Pérez et al., 2020; Kujala et al., 1996), this study observed an overall delay in spontaneous responses for MS patients compared to controls. These findings extend the characterization beyond the inaccuracy identified by naming scores alone, capturing difficulties that emerge during the lexical retrieval process. These difficulties are plausibly driven by a range of underlying cognitive–linguistic mechanisms. One proposed account is a semantic access impairment at the level of lexical selection, with the longer response times reflecting a greater reliance on manualised retrieval processes which require increased effort to suppress lexical competitors (Goodglass et al., 1984; Levelt et al., 1999). Alternatively, the delay may reflect a general processing speed deficit, as is commonly observed in MS

(Grzegorski & Losy, 2017). Regardless of the precise mechanism, the latency analysis nonetheless captures language production inefficiencies that are otherwise uncharacterized by overall naming scores alone. This study provides support for latency analysis as a useful approach to improving the characterization of single-word level language function in MS. From a practical perspective, these findings highlight potential clinical utility of incorporating latencies into naming assessments. Informally, this may entail clinicians being cognizant of longer response latencies in their patients as possible evidence of dysfunctional naming, even in the absence of an impaired accuracy score. More formal metrics may alternatively be incorporated into assessment, such as mean response times or frequency of longer latencies, as applied in naming tasks in other clinical populations (Hamberger & Seidel, 2003).

Limitations and future directions

These interpretations must be considered within the constraints of the study. The patient group was recruited from those referred to the RMH Cognitive Neuroimmunology Clinic for specialist cognitive investigation following some level of patient or clinician concern. While the prevalence of objective impairment in the sample may therefore be higher than the general MS community, the sample composition is appropriate for validating language measures by ensuring the tools are tested in the population they ultimately aim to assess.

The somewhat limited sample size must also be considered. The sample size was suitable for the validation of the SYDBAT, comparable to previous validations of this task (Janssen et al., 2022; Savage et al., 2013). It may have however increased the likelihood of type two errors in other analyses, particularly the error analysis. While the frequency of *overall error types* did vary with group, no one error type significantly differed between patients and controls, including semantic errors, a previously consistent distinguishing factor in this population (De Dios Pérez et al., 2020; Lethlean & Murdoch, 1994). As the frequency of each error was minimal, it is reasonable to speculate that the lack of significant simple effects may be attributed to insufficient power. Future studies may therefore consider employing larger cohorts to capture a greater number of errors and yield more robust findings.

The coding process used in this error analysis may also be developed for application in future studies. This component was treated as a pilot analysis, secondary to the primary aim of validating the SYDBAT and applying a novel coding system adapted from previous literature (Kohn & Goodglass, 1985; Kujala et al., 1996; Lethlean & Murdoch, 1994). The methodology supported strong internal consistency, with a single blinded researcher ensuring consistent application of the coding framework, and resolution of all ambiguous responses through consensus discussion between three clinical neuropsychologists promoting reliable coding agreement. A key limitation, however, is that inter-rater reliability could not be formally calculated. Future studies may consider involving multiple independent coders to enable such estimation and strengthen the robustness of the findings.

The limited evaluation of language must also be acknowledged. This study exclusively examined visual confrontation naming and thereby omits other linguistic functions that may be additionally affected. Discourse analyses, which capture higher-level linguistic features such as fluency, cohesion, and coherence (D'Aprano et al., 2024), may be a particularly pertinent approach to consider in

future research. As a diffuse neurological condition (Thompson et al., 2018), it is reasonable to speculate that higher-level language, which relies upon widely distributed cognitive processes (Mesulam, 1990), may be fundamentally compromised and warrant assessment in MS. Such approaches are less clinically applicable due to their time and expertise demands, though may be valuable in enhancing theoretical understanding and contributing to a more complete characterization of language function in MS.

Conclusions

This study presents the SYDBAT naming subtest and latency analysis as valuable adjunct methods for assessing single-word level language in the MS population. The study offers clinicians a broader toolkit to characterize language functioning and supplement existing cognitive batteries, namely the BICAMS. This represents a positive step toward increased accuracy and complete assessment, informing patient care in this clinical population. The characterization provided by these approaches is, however, by no means exhaustive, and continued efforts to improve the understanding of language function in MS presents an important avenue for future research.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S1355617725101513>.

Acknowledgements. We would like to thank the patients who were referred to the Royal Melbourne Hospital Cognitive Neuroimmunology Clinic and took the time to participate in this study. We would also like to extend this thanks to all healthy control participants who additionally took part in this study. Finally, we acknowledge Aboriginal and Torres Strait Islander people of the unceded land on which we work, learn, and live: we pay respect to Elders past, present, and future, and acknowledge the importance of Indigenous knowledge at The University of Melbourne.

Author contributions

A.H: Writing – original draft preparation (lead); Writing – review, and editing (equal); Data curation (equal); Formal analysis (supporting)

S.R: Conceptualization (equal); Methodology (equal); Investigation (lead); Data curation (equal); Writing – review, and editing (equal); Supervision (equal)

C.B.M: Conceptualization (equal); Methodology (equal); Formal analysis (lead); Writing – review, and editing (equal); Supervision (equal)

G.R: Conceptualization (equal); Methodology (equal); Writing – review, and editing (equal); Supervision (equal)

F.D: Conceptualization (equal); Methodology (equal); Writing – review, and editing (equal); Supervision (lead)

Funding statement. Funding relevant to G.R. includes a Medical Research Future Fund (MRFF) Australian Epilepsy Research Fund grant for project funding as well as salary support in part from an Australian National Health & Medical Research Council (NHMRC) Investigator grant (APP2008737).

Competing interests. The authors declare none.

Ethical standard. The study was approved by the Human Research Ethics Committee of the Royal Melbourne Hospital, Melbourne (2020.240 RMH67046), and all research was conducted in accordance with ethical standards of the 1964 Declaration of Helsinki. All participants provided written informed consent prior to participating. Data supporting the findings of this study are available from the corresponding author, upon reasonable request

References

- Apple Inc. (2023). Voice Memos (Version 2.4) [Mobile app]. App Store. <https://apps.apple.com/us/app/voice-memos/id1069512134>.
- Audacity Team. (2024). Free Audio Editor and Recorder (Version 3.5.1) [Computer application]. <https://www.audacityteam.org/>.

- Beatty, R. A., Murphy, H., Cornwell, M., Braun, T., Stein, V., Goldstein, M., & Bender, H. A. (2017). Caution warranted in extrapolating from Boston naming test item gradation construct. *Applied Neuropsychology: Adult*, 24(1), 65–72.
- Beatty, W. W., Goodkin, D. E., Monson, N., & Beatty, P. A. (1989). Cognitive disturbances in patients with relapsing remitting multiple sclerosis. *Archives of Neurology*, 46(10), 1113–1119.
- Beatty, W. W., & Monson, N. (1989). Lexical processing in Parkinson's disease and multiple sclerosis. *Journal of Geriatric Psychiatry and Neurology*, 2(3), 145–152.
- Benedict, R. H. B., Fischer, J. S., Archibald, C. J., Arnett, P. A., Beatty, W. W., Bobholz, J., Chelune, G. J., Fisk, J. D., Langdon, D. W., Caruso, L., Foley, F., LaRocca, N. G., Vowels, L., Weinstein, A., DeLuca, J., Rao, S. M., & Munschauer, F. (2002). Minimal neuropsychological assessment of MS patients: A consensus approach. *The Clinical Neuropsychologist*, 16(3), 381–397.
- Brandstadter, R., Fabian, M., Leavitt, V. M., Krieger, S., Yeshokumar, A., Katz Sand, I., Klineova, S., Riley, C. S., Lewis, C., Pelle, G., Lublin, F. D., Miller, A. E., & Sumowski, J. F. (2020). Word-finding difficulty is a prevalent disease-related deficit in early multiple sclerosis. *Multiple Sclerosis Journal*, 26(13), 1752–1764.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- D'Aprano, F. (2023). Deconstructing discourse: the window to the temporal lobe epilepsy patient [Doctoral dissertation, University of Melbourne]. <https://minerva-access.unimelb.edu.au/items/1778fc84-3bba-4535-9c30-905940e7469c>
- D'Aprano, F., Malpas, C. B., Roberts, S., & Saling, M. M. (2024). Macrolinguistic function in temporal lobe epilepsy: a reinterpretation of circumstantiality. *Aphasiology*, 38(2), 181–204.
- De Dios Pérez, B., Cordova Luna, E., Cloutman, L., Rog, D., Preston, E., & Conroy, P. (2020). Anomia in people with rapidly evolving severe relapsing-remitting multiple sclerosis: Both word retrieval inaccuracy and delay are common symptoms. *Aphasiology*, 34(2), 195–213.
- Dusankova, J. B., Kalincik, T., Havrdova, E., & Benedict, R. H. B. (2012). Cross cultural validation of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS) and the brief international cognitive assessment for multiple sclerosis (BICAMS). *The Clinical Neuropsychologist*, 26(7), 1186–1200.
- El-Wahsh, S., Ballard, K., Kumfor, F., & Bogaardt, H. (2020). Prevalence of self-reported language impairment in multiple sclerosis and the association with health-related quality of life: An international survey study. *Multiple Sclerosis and Related Disorders*, 39, 101896.
- Goodglass, H., Theurkauf, J. C., & Wingfield, A. (1984). Naming latencies as evidence for two modes of lexical retrieval. *Applied Psycholinguistics*, 5(2), 135–146.
- Grzegorski, T., & Losy, J. (2017). Cognitive impairment in multiple sclerosis – a review of current knowledge and recent research. *Reviews in the Neurosciences*, 28(8), 845–860.
- Hamberger, M. J., & Seidel, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society: JINS*, 9(3), 479.
- Janssen, N., Roelofs, A., van den Berg, E., Eikelboom, W. S., Holleman, M. A., in de Braek, D. M. J. M., Piguet, O., Piai, V., & Kessels, R. P. C. (2022). The diagnostic value of language screening in primary progressive aphasia: Validation and application of the Sydney language battery. *Journal of Speech, Language, and Hearing Research*, 65(1), 200–214.
- Jennekens-Schinkel, A., Lanser, J. B., van der Velde, E. A., & Sanders, E. A. (1990). Performances of multiple sclerosis patients in tasks requiring language and visuocognition. Assessment of outpatients in quiescent disease stages. *Journal of the Neurological Sciences*, 95(1), 89–103.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test*. Lea and Febiger.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746–759.
- Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and Language*, 24(2), 266–283.
- Kujala, P., Portin, R., & Ruutiainen, J. (1996). Language functions in incipient cognitive decline in multiple sclerosis. *Journal of the Neurological Sciences*, 141(1-2), 79–86.
- Langdon, D. W., Amato, M. P., Boringa, J., Brochet, B., Foley, F., Fredrikson, S., Hämäläinen, P., Hartung, H.-P., Krupp, L., Penner, I. K., Reder, A. T., & Benedict, R. H. B. (2012). Recommendations for a brief international cognitive assessment for multiple sclerosis (BICAMS). *Multiple Sclerosis Journal*, 18(6), 891–898.
- Lethlean, J. B., & Murdoch, B. E. (1994). Naming errors in multiple sclerosis: Support for a combined semantic/perceptual deficit. *Journal of Neurolinguistics*, 8(3), 207–223.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38.
- Mack, W. J., Freed, D. M., Williams, B. W., & Henderson, V. W. (1992). Boston Naming Test: shortened versions for use in Alzheimer's disease. *Journal of Gerontology*, 47(3), P154–P158.
- Maltby, V. E., Lea, R. A., Ribbons, K., Lea, M. G., Schofield, P. W., & Lechner-Scott, J. (2020). Comparison of BICAMS and ARCS for assessment of cognition in multiple sclerosis and predictive value of employment status. *Multiple Sclerosis and Related Disorders*, 41, 102037.
- Mesulam, M.-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 28(5), 597–613.
- Olivares, T., Nieto, A., Sánchez, M. P., Wollmann, T., Hernández, M. A., & Barroso, J. (2005). Pattern of neuropsychological impairment in the early phase of relapsing-remitting multiple sclerosis. *Multiple Sclerosis Journal*, 11(2), 191–197.
- Piguet, O. (2022). The Sydney Language Battery [Mobile app]. App Store <https://apps.apple.com/us/app/the-sydney-language-battery/id1509136039>.
- R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rao, S. M., Leo, G. J., Bernardin, L., & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis I. Frequency, patterns, and prediction. *Neurology*, 41(5), 685–691.
- Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research* (4th ed.). Jossey-Bass
- Savage, S., Hsieh, S., Leslie, F., Foxe, D., Piguet, O., & Hodges, J. R. (2013). Distinguishing subtypes in primary progressive aphasia: Application of the Sydney language battery. *Dementia and Geriatric Cognitive Disorders*, 35(3–4), 208–218.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.
- Tallberg, I. M., & Bergendal, G. (2009). Strategies of lexical substitution and retrieval in multiple sclerosis. *Aphasiology*, 23(9), 1184–1195.
- Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., ... Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2), 162–173.
- Zoom Video Communications Incorporated. (2016). Security Guide.