Received: 29 May 2024 | Accepted: 10 August 2024

DOI: 10.1111/ene.16454

european journal of neurology
the official journal of the european academy of neurology

ORIGINAL ARTICLE

Development and validation of an electronic Symbol-Digit Modalities Test for remote monitoring of people with multiple sclerosis

Michelangelo Dini[1,2] | Giulia Gamberini[3] | Marta Tacchini[1,2] | Angela Boschetti[1,2] | Alessandro Gradassi[3] | Luca Chiveri[3] | Mariaemma Rodegher[3] | Giancarlo Comi[3] | Letizia Leocani[1,2]

[1]Vita-Salute San Raffaele University, Milan, Italy

[2]Experimental Neurophysiology Unit, Institute of Experimental Neurology (INSPE), IRCCS-Scientific Institute San Raffaele, Milan, Italy

[3]Department of Neurorehabilitation Sciences, Casa di Cura Igea, Milan, Italy

Correspondence
Letizia Leocani, Faculty of Medicine, University Vita-Salute San Raffaele and Experimental Neurophysiology Unit, Institute of Experimental Neurology-INSPE, San Raffaele Scientific Institute, Via Olgettina 60, Milan 20132, Italy.
Email: leocani.letizia@unisr.it

Abstract

Background: Computerized cognitive tests may extend the reach of cognitive screening and monitoring to those with mobility issues or living in remote areas. Moreover, it could enable frequent and autonomous remote cognitive assessments in people with multiple sclerosis (pwMS) on account of its reduced economic and organizational costs. This may further improve our understanding of longitudinal trends and significantly improve the standard of care for pwMS living in remote areas or with mobility limitations. We aimed to evaluate the psychometric properties of an electronic Symbol-Digit Modalities Test (eSDMT) designed to allow pwMS to perform a rapid cognitive assessment independently from home using their own PC/laptop.

Methods: Sixty-two participants underwent a neuropsychological evaluation, and then performed the eSDMT in the clinic. Forty-two participants also repeated the eSDMT at home. We assessed concurrent validity (eSDMT vs. oral SDMT), test–retest reliability (in the clinic vs. at home), discriminant validity (pwMS with/without cognitive impairment), and other psychometric characteristics of the eSDMT (effect of age, sex, and education on test scores).

Results: We observed good-to-excellent concurrent validity ($r \geq 0.84$, all $p < 0.0001$) and test–retest reliability (intraclass correlation coefficients [ICCs]>0.87, all $p < 0.0001$). Discriminant validity was excellent (area under the curves [AUCs] >0.84, all $p < 0.0001$). eSDMT scores were only slightly influenced by demographic characteristics (all $R^2 < 0.200$).

Conclusions: We provided evidence which supports the use of our eSDMT as a feasible, valid, and reliable remote assessment of cognitive function in pwMS. Future studies will investigate long-term reliability and predictive power.

KEYWORDS
cognition, digital health, multiple sclerosis, telemedicine

INTRODUCTION

Cognitive impairment is a frequent symptom of multiple sclerosis (MS) which negatively affects daily activities, lowering the quality of life of people with MS (pwMS) [1, 2]. It occurs even in the earlier stages of the disease [3] and it can develop acutely and follow a remitting–relapsing pattern, or develop insidiously and follow a gradual progression [4].

Cognitive deficits of pwMS are typically assessed by administering standardized neuropsychological tests and batteries designed to provide an estimate either of global cognitive functioning or of specific domains [5]. The Symbol-Digit Modalities Test (SDMT) is currently the most recommended test, owing to its sensitivity, test–retest reliability, and predictive validity [6]. The SDMT is considered the most effective test for distinguishing between pwMS with or without cognitive impairment and represents the gold standard outcome measure for cognition in clinical trials [7, 8].

Standardized in-clinic cognitive tests in pwMS are usually repeated at intervals of 6–12 months. Thus, they are blind to cognitive fluctuations between evaluations, and could cause considerable delay in identifying the shift to progressive cognitive worsening. The technological advances of the last 10–15 years have led to the development of software which may enable the long-term monitoring of cognitive functioning in a cost-effective and ecological manner [9]. PCs and high-speed internet have become widely accessible in most countries, and thus can facilitate large-scale repeated remote assessments in a cost-effective and unobtrusive manner [10–12]. This may enable cost-effective long-term monitoring of cognitive functioning in pwMS [9] by allowing patients to perform cognitive tests remotely and autonomously. For researchers, granular longitudinal monitoring may enable the gathering of larger quantities of data and the acquisition of new insights. Clinicians could monitor individual patients and detect meaningful changes more efficiently, improving the standard of care [10].

However, before remote digital cognitive tests can be implemented in research and clinical practice, their psychometric properties should be carefully assessed [13]. Moreover, most published digital cognitive tests are not widely and freely available for researchers or are embedded in software from commercial enterprises which may not be approved by independent ethics committees and is unsuitable for research due to lack of access to raw data. We aimed to evaluate the psychometric properties of a custom-built electronic version of the SDMT which enables pwMS to perform remote cognitive assessments independently, using their own PC/laptop.

METHODS

Participants

Participants were recruited among inpatients attending the Neurorehabilitation Department of Casa di Cura Igea, Milan, Italy between March and December 2023. Inclusion criteria were (1) age>18 years, (2) diagnosis of MS based on the revised McDonald's criteria [14], (3) absence of significant neurological and/or psychiatric comorbidities, as diagnosed by a neurologist and/or psychiatrist, (4) absence of visual and/or motor deficits severe enough to invalidate routine pen-and-paper neuropsychological testing (e.g., patients unable to distinguish the different SDMT symbols due to vision deficits or unable to use a pencil); (5) ability to provide written informed consent; and (6) laptop/PC and internet connection availability at home (only for the remote part of the study).

All study procedures adhered to the principles established by the Declaration of Helsinki. Written informed consent was obtained from all participants prior to the implementation of any study procedure as per the protocols approved by the Institutional Ethics Committee (protocol number: 1162_2022).

Electronic Symbol-Digit Modalities Test (eSDMT)

We developed an electronic version of the SDMT (eSDMT) suitable for online administration using Lab.js [15] and hosted it on a JATOS server [16]. This allows immediate on-demand availability through any internet browser. The test was designed to be easily performed by patients, without the need for any software download or setup, with minimal hardware/software requirements and maximum cross-platform compatibility.

In the eSDMT (Figure 1), a symbol-digit key is displayed at the top center of the screen, where nine symbols are associated with digits 1–9. Symbols appear at the center of the screen, one at a time in a pseudo-random sequence, and participants must press the corresponding number using the keyboard. The eSDMT consists of 54 stimuli (six blocks, pseudo-randomized so that each block contains one instance of each symbol). Once the participant presses a digit key, the given response is stored, and the test sequence moves to the next stimulus.

Before the eSDMT, participants perform a choice reaction task (hereon referred to as 'Numbers') during which random digits (1–9) are displayed at the center of the screen, one at a time in a pseudo-random sequence (27 stimuli; same procedure as for the eSDMT). The Numbers task was included as a baseline reaction task which could be less dependent on higher cognitive processes.

Instructions are provided in writing on the screen, and the test sequence includes a practice run before both the Numbers and eSDMT tasks, during which each possible target stimulus is presented only once.

Reaction times (RTs; in milliseconds) and keypresses are recorded for each target stimulus. Upon completion, summary results are shown, and a text box enables reporting of external disturbances that may have affected test validity (e.g., the doorbell ringing).

Neuropsychological assessment

Upon admission to the inpatient unit, all participants underwent a neuropsychological assessment with the Italian versions

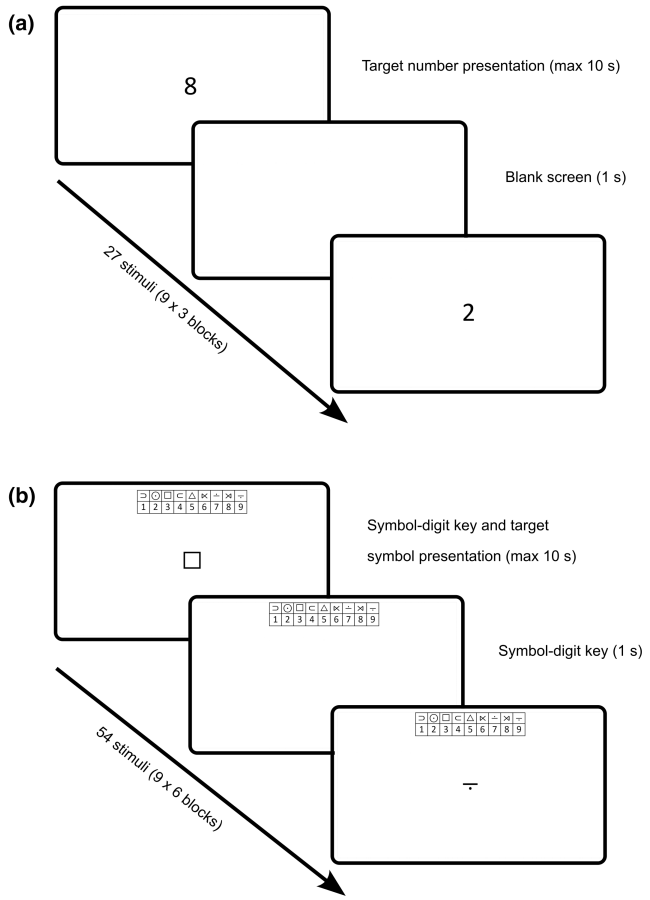

FIGURE 1 Electronic Symbol-Digit Modalities Test (eSDMT) flow diagram. Diagram depicting the flow of the Numbers and eSDMT tasks. (a) In the Numbers task, stimuli are represented by single-digit numbers (1–9), presented serially (number of stimuli = 27, stimulus duration = maximum 10s, response–stimulus interval = 1 s). (b) In the eSDMT task, a key is displayed at the top of the screen, containing nine symbols which are associated with single-digit numbers (1–9). Target stimuli are represented by symbols, presented serially (number of stimuli = 54, maximum stimulus duration = 10s, response–stimulus interval = 1 s). In both tasks, the presentation order is randomized so that each stimulus is presented only once per block.

of the Montreal Cognitive Assessment (MoCA) [17] and the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS) [18]. All neuropsychological assessments were conducted by experienced neuropsychologists as per clinical protocol.

The MoCA is a screening tool widely used to assess global cognitive functioning [19]. The BICAMS battery consists of the oral SDMT, a verbal learning test, and a visuospatial learning test [20].

Study procedures

Participants performed the eSDMT in the clinic, after the standard neuropsychological evaluation, (median [1st–3rd quartile] = 13.25 [8.5–20.5] days after) to minimize interference with the clinical routine. The eSDMT was administered using the same laptop for all participants (screen diagonal = 17.3″, resolution = 1920×1080). Test instructions were provided on the screen, and participants were instructed to perform the test as if they were on their own.

Upon discharge, participants were emailed an individual link, enabling them to perform the eSDMT from home using their own laptop/PC. Participants were instructed to complete the cognitive test autonomously, in an environment as free as possible from possible distractors.

In the clinic, all participants completed the same eSDMT version, whereas the at-home testing version featured randomized symbol–number pairings to minimize practise effects.

Statistical analyses

We calculated mean RTs (based on all stimuli, including incorrect responses) and the number of correct responses in 90s as eSDMT scoring metrics for both timepoints (in-clinic session and first remote session). Mean RTs measure processing speed and penalize slow responses but not inaccuracy. Correct responses in 90s reflect processing speed while also penalizing inaccuracy.

Normality was assessed via statistical tests and by examining skewness and excess kurtosis [21]. Appropriate transformations were performed on non-normally distributed variables, where required (e.g., mean RTs were transformed using Box-Cox transformation for bivariate linear correlations or linear regression analyses, and then rescaled to range 0–100 for better interpretability; higher scores indicate faster RTs).

Multiple linear regressions were fitted to assess the effect of sociodemographic variables (sex, age, education) on oral SDMT and eSDMT scores, separately.

Concurrent validity (eSDMT score vs. oral SDMT score) and convergent validity (eSDMT score vs. MoCA score) were evaluated using Pearson's correlation coefficient (r). Test–retest reliability (in-clinic assessment vs. remote assessment) was investigated using intraclass correlation coefficients (ICC). Systematic and proportional bias was assessed with Bland–Altman plots and one-sample t-tests (H_0: mean difference = 0) and Pearson's correlation coefficient, respectively. We used Tukey's fence method to identify extreme outliers in the test–retest data [22]. A sensitivity analysis including all outliers was also conducted.

Discriminant validity (pwMS with vs. without cognitive impairment) was assessed via the area under the curve (AUC) of receiver operating characteristic (ROC) analyses for mean RTs and correct responses in 90s, separately. Impaired MoCA performance was defined as scoring below the 5th percentile according to published normative data [23]. Impaired BICAMS performance was defined as scoring below −1.5 standard deviations (SDs) from the mean in at least one test, based on published normative data [24].

All tests were two-sided ($\alpha = 0.05$). Statistical analyses were performed using Python libraries for data science [25–29]. ROC analyses were performed using IBM SPSS v26.0. Aggregate statistics are reported as mean ± SD, median (1st–3rd quartile), or as

count (%); 95% confidence intervals (95% CIs) are provided where appropriate.

RESULTS

Of $n=82$ consecutive pwMS who attended our clinic, $n=9$ were excluded for neurological (stroke = 2, encephalopathy = 1, parkinsonism = 1) and/or psychiatric (e.g., major depressive disorder = 2, bipolar disorder = 1, schizophrenia = 1) comorbidities and $n=4$ were unable to undergo routine clinical testing with the BICAMS due to visual deficits and/or severe cognitive impairment. Thus, $n=69$ patients were considered eligible, of which $n=63$ agreed to participate in the study; $n=59$ also fitted the inclusion criteria for the remote part of the study, of which $n=51$ agreed to perform the digital cognitive assessment at home.

The study sample consisted primarily of people with progressive MS (54/63, 87.3%), mostly female (38/63, 60.3%), with median age 52.5 (47–60.75) years, median education 13 (11.25–17) years, and median Expanded Disability Status Scale (EDSS) score 6.5 (6–6.5).

Cognitive impairment was detected in five (7.9%) participants based on MoCA score, and in 20 (31.7%) participants based on BICAMS scores. All five participants with impaired MoCA performance were also impaired at the BICAMS (Table 1).

TABLE 1 Study sample descriptive statistics.

Variable	*n* (%)	Mean ± SD or median (1st–3rd quartile)
MS subtype	SPMS = 40 (63.5)	–
	PPMS = 15 (23.8)	
	RRMS = 5 (7.9)	
	MS (n.s.) = 3 (4.8)	
Sex	Female = 38 (60.3)	–
	Male = 25 (39.7)	
Age (years)	–	52.5 (47–60.75)
Education (years)	–	13 (11.25–17)
EDSS	–	6.5 (6–6.5)
Cognitive functioning	Impaired = 20 (31.7)	–
MoCA score	Impaired = 5 (7.9)	26.08 ± 2.85
SDMT score	Impaired = 12 (19.0)	41.61 ± 11.81
CVLT-II score	Impaired = 5 (7.9)	54.35 ± 12.56
BVMT-R score	Impaired = 9 (14.3)	21.39 ± 10.16

Abbreviations: BVMT-R, Brief Visuospatial Memory Test–Revised; CVLT-II, California Verbal Learning Test–2nd edition; EDSS, Expanded Disability Status Scale; MoCA, Montreal Cognitive Assessment; PPMS, Primary Progressive Multiple Sclerosis; RRMS, Relapsing–Remitting Multiple Sclerosis; SD, standard deviation; SDMT, Symbol-Digit Modalities Test; MS (n.s.), multiple sclerosis (not specified); SPMS, Secondary Progressive Multiple Sclerosis.

eSDMT

Sixty-two participants completed the eSDMT session in the clinic successfully. Only one participant could not complete it due to severe upper limb weakness, which rendered them unable to reach all the response keys. Overall, participants completed the entire eSDMT procedure in the clinic in 7:56 ± 1:37 min. This included the time spent reading instructions and performing the practice trials. Fifty-four (85.7%) participants successfully completed the entire eSDMT procedure in less than 10 min, with the slowest time to completion being 12:42 min. The average completion time from home was 7:24 ± 1:48 min, with the slowest time being 12:34 min. Descriptive eSDMT score statistics are reported in Table 2.

We assessed the effect of sex, age, and education on oral SDMT score and the two eSDMT scores separately. In all three cases, the overall model was statistically significant, but the explained variance was very small (Table 3).

Concurrent and convergent validity

Oral SDMT score correlated strongly with eSDMT mean RTs ($r=0.873$, $p<0.0001$) and correct responses in 90s ($r=0.837$, $p<0.0001$) (Figure 2a). MoCA score showed moderate positive correlations with mean RTs ($r=0.460$, $p=0.0002$) and correct responses in 90s ($r=0.429$, $p=0.0003$) (Figure 2b).

Test–retest reliability

Of 51 participants who agreed to perform the eSDMT at home, $n=42$ (82%) performed it as planned, while 9 (18%) never performed the test from home. On average, participants repeated the eSDMT 1 month (25 [17.25–42.75] days) after performing it in the clinic. Two extreme test–retest outliers were excluded from the reliability analysis of mean RTs, based on Tukey's method (mean RTs differences [T0–T1] = −1064 ms and − 1205 ms).

eSDMT test–retest reliability was good-to-excellent for mean RTs (ICC [2, 1]=0.88 [95% CI = 0.79–0.93], $p<0.0001$) and correct responses in 90s (ICC [2, 1]=0.87 [95% CI = 0.77–0.93], $p<0.0001$). Mean RTs score showed no systematic bias (mean difference = −29.96 ms [95% CI = −126.16–66.25], $p=0.532$) or proportional bias ($r=0.168$, $p=0.299$). Correct responses in 90s showed no systematic bias (mean difference = 0.26 [95% CI = −0.51–1.03], $p=0.497$) or proportional bias ($r=-0.177$, $p=0.262$) (Figure 3).

The sensitivity analysis including the two outliers excluded from the previous analysis confirmed the good-to-excellent test–retest reliability of mean RTs (ICC [2, 1] = 0.84 [95% CI = 0.72–0.91], $p<0.0001$), as well as the lack of systematic ($p=0.165$) or proportional bias ($p=0.147$).

TABLE 2 Electronic Symbol-Digit Modalities Test descriptive statistics.

Parameter	Mean ± SD or median (1st–3rd quartile)	Minimum	Maximum
Numbers accuracy (%)			
In the clinic	0.992 ± 0.021	0.889	1.000
At home	0.993 ± 0.025	0.852	1.000
eSDMT accuracy (%)			
In the clinic	0.987 ± 0.019	0.907	1.000
At home	0.988 ± 0.024	0.870	1.000
Numbers mean RTs (ms)			
In the clinic	1199.23 ± 254.09	837.79	2141.86
At home	1224.33 ± 307.80	825.36	2164.19
eSDMT mean RTs (ms)			
In the clinic	2470.78 ± 660.21	1420.74	4679.84
At home	2524.55 ± 709.67	1446.52	4754.04
eSDMT correct responses in 90s			
In the clinic	37 (31–43)	20	54
At home	36.5 (30.25–42)	15	54

Note: eSDMT statistics from the first assessment in the clinic ($n = 62$) and at home ($n = 42$).

Abbreviations: eSDMT, electronic Symbol-Digit Modalities Test; RT, reaction time; SD, standard deviation.

TABLE 3 Impact of sociodemographic variables on oral Symbol-Digit Modalities Test (SDMT) and electronic SDMT (eSDMT) score.

Test score	Overall model significance	Predictors	β	p
Oral SDMT score	$F_{(3,58)} = 4.64$ $p = 0.006$ Adjusted $R^2 = 0.152$	Age	−0.353	0.005
		Education	0.240	0.054
		Sex	0.071	0.568
eSDMT – mean RTs	$F_{(3,58)} = 5.57$ $p = 0.002$ Adjusted $R^2 = 0.183$	Age	−0.386	0.002
		Education	0.241	0.043
		Sex	0.075	0.525
eSDMT – correct responses in 90s	$F_{(3,58)} = 5.28$ $p = 0.003$ Adjusted $R^2 = 0.174$	Age	−0.368	0.003
		Education	0.221	0.064
		Sex	0.135	0.257

Note: Results of linear regression analyses to assess the impact of age, education, and sex on oral SDMT and both eSDMT scores. Overall model significance is indicated by the model's ANOVA $F_{(\text{degrees of freedom})}$ statistic and its related p-value, as well as by the explained variance of the model (adjusted R^2). Standardized beta coefficients (β) and their p-values are reported for individual variables.

Abbreviations: eSDMT, electronic SDMT; RT, reaction time; SDMT, Symbol-Digit Modalities Test.

Discriminant validity

Given the low proportion of participants with cognitive impairment based on MoCA scores ($n = 5$), we investigated the discriminant validity of eSDMT based only on impaired BICAMS performance. There were no significant age, sex, or education differences between the two groups (all $p > 0.433$). We observed good discriminant validity for both mean RTs (ROC-AUC = 0.846 [95% CI = 0.745–0.948], $p < 0.0001$) and correct responses in 90s (ROC-AUC = 0.815 [95% CI = 0.699–0.931], $p < 0.0001$) (Figure 4). We identified optimal sensitivity/specificity cut-off values for mean RTs (score > 2500 ms; sensitivity = 0.750, specificity = 0.810) and correct responses in 90s (score ≤ 25; sensitivity = 0.800, specificity = 0.714).

DISCUSSION

We aimed to evaluate an eSDMT designed to allow pwMS to perform a rapid cognitive assessment independently. This eSDMT is quick and easy to use, and it shows good-to-excellent concurrent validity with the oral SDMT. We also provided evidence that it can be performed by pwMS independently from home, using their own PC/laptop, with excellent reliability. Finally, we showed that it can accurately discriminate between pwMS with/without cognitive impairment, and that it correlates with a gold standard measure of global cognitive functioning (MoCA).

Feasibility was demonstrated by all participants performing the eSDMT successfully in the clinic, with only one exception, owing to severe upper limb weakness. This indicates that a minority

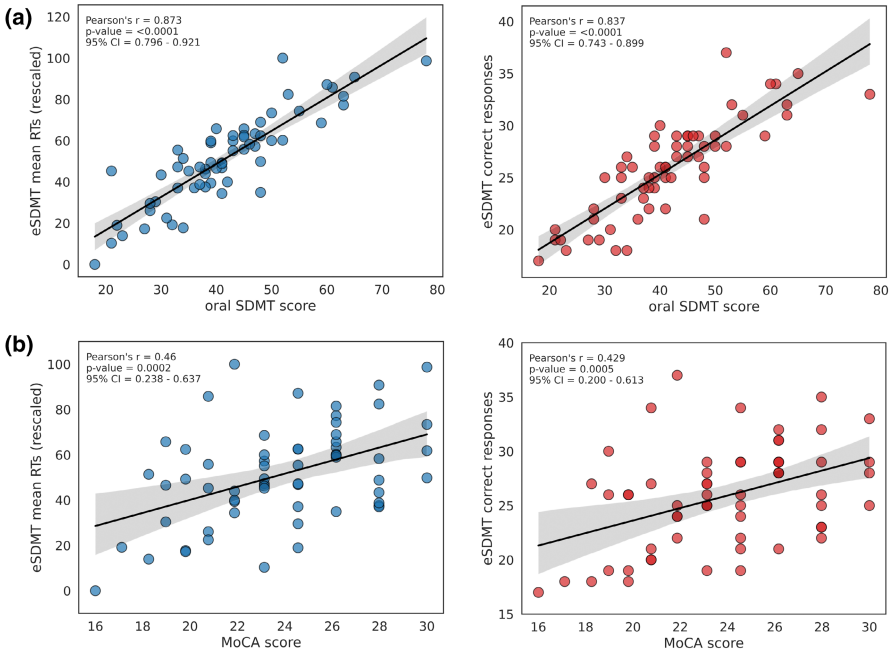

FIGURE 2 Concurrent and convergent validity. Scatterplots representing correlations between electronic Symbol-Digit Modalities Test (eSDMT) mean reaction times (RTs) (left side) and correct responses in 90 s (right side) and validated neuropsychological tests. (a) Strong positive correlations between both eSDMT scores and oral SDMT score. (b) Moderate positive correlations between both eSDMT scores and Montreal Cognitive Assessment (MoCA) Score. eSDMT RTs have been inverted and rescaled to range 0–100 (higher values indicate faster RTs). Pearson's correlation coefficient (*r*) and associated *p*-values are reported in the top left corner. Linear fits (black line) with 95% confidence interval (CI) are displayed for every correlation.

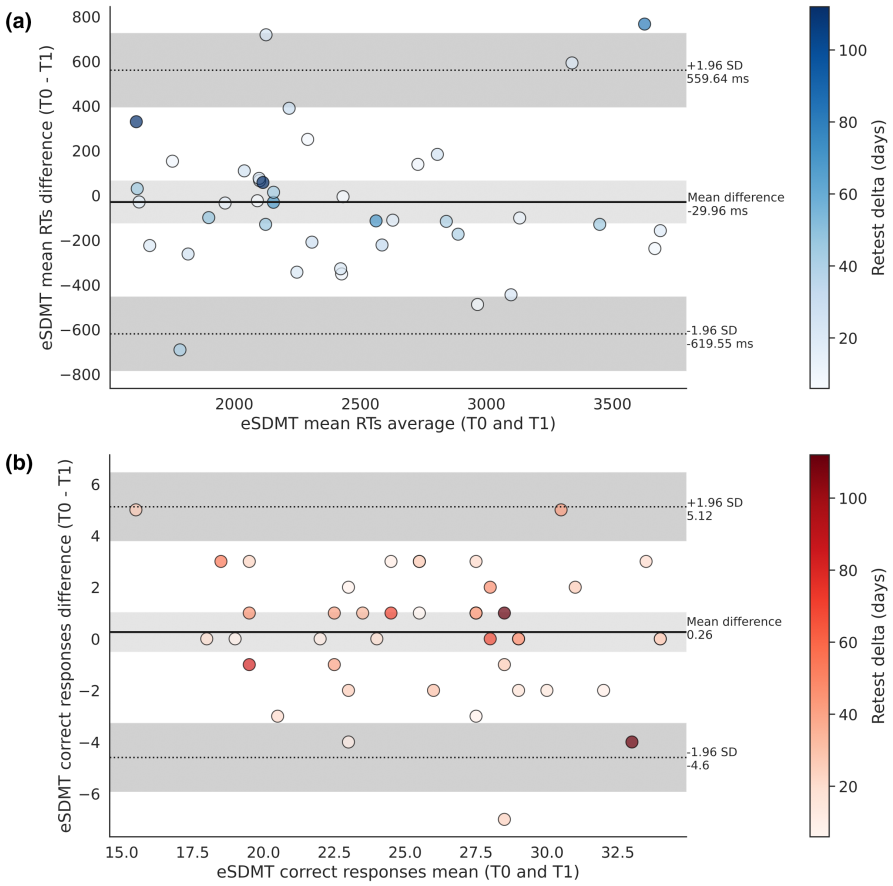

FIGURE 3 Test–retest reliability. Bland–Altman plots displaying test–retest reliability of electronic Symbol-Digit Modalities Test (eSDMT) mean reaction times (RTs) (a) and correct responses in 90 s (b). Circles represent individual bivariate cases; darker fill colours indicate longer test–retest delta (in days). Mean difference with 95% confidence interval (CI) (solid horizontal line with shaded area) and limits of agreement with 95% CI (dotted lines with shaded area) are displayed, as well as numerical values (right axis). SD, standard deviation.

of patients may be able to perform the oral SDMT, but not the eSDMT, due to motor deficits. As opposed to other eSDMT studies [30–34], we opted not to exclude participants based on vision and/or upper limbs impairment. By including all pwMS who were able to undergo pen-and-paper neuropsychological evaluations, we improved the generalizability of our results to a wider range of pwMS. Most participants were able to complete the test in under 10 min, that is, the average administration time of the oral SDMT [7]. Our sample had a high prevalence of participants with progressive MS and high disability. Therefore, pwMS with lower disability may complete the eSDMT even faster. Nevertheless, this is just a hypothesis, which should be evaluated in future by recruiting

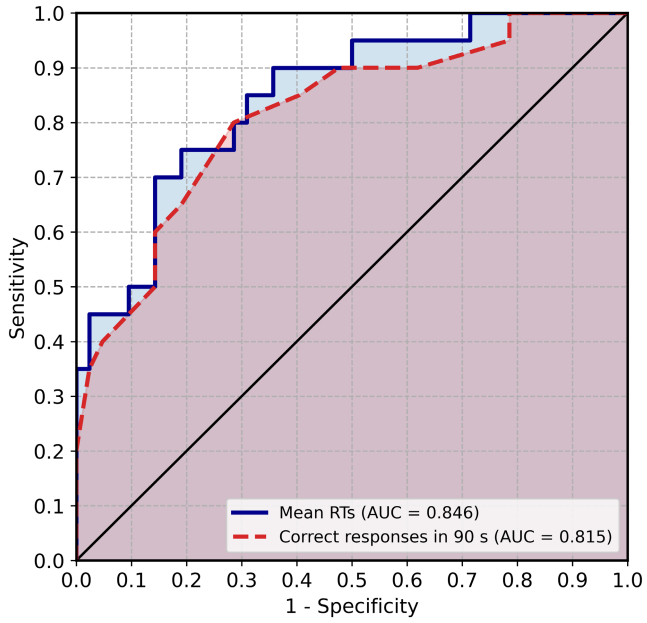

FIGURE 4 Discriminant validity. Receiver operating characteristic curves of electronic Symbol-Digit Modalities Test (eSDMT) mean reaction times (RTs) (solid line) and correct responses in 90 s (dashed line) for the detection of cognitive impairment in people with multiple sclerosis. The area under the curve (AUC) is represented by the lightly shaded area under the two curves.

pwMS with lower disability. Although four participants could not be included in the remote part of the study due to lack of laptop/PC and/or internet connection, those included did not report any technical issues which impeded independent eSDMT execution at home. However, nine participants were lost at follow-up. While there is no evidence that they encountered technical difficulties which led them to abandon the study, this possibility should not be ruled out. We did not administer self-reported usability questionnaires, and therefore cannot draw any conclusion regarding the perceived usability of the eSDMT, which may not be perfectly correlated with objective measures. Future research will need to evaluate patients' perspectives, adherence to long-term monitoring, and associated sociodemographic, clinical, and practical factors.

We proposed two different eSDMT scores, hypothesizing that both would reflect processing speed, although one would focus more on pure speed, and the other on speed and accuracy. These two scores demonstrated comparable psychometric properties, suggesting that their similarities (i.e., their dependence on processing speed) far outweigh their differences. Nevertheless, having multiple valid and reliable scoring metrics should not be overlooked, as they possess different strengths and weaknesses. RTs allow the evaluation of test performance much more granularly, by selecting specific time intervals (e.g., block by block) or specific symbol properties (e.g., shape, position). This may be especially useful for research focused on deeper investigations of cognitive processes (i.e., learning, cognitive fatigability). Correct responses, conversely, are more readily interpreted, since they follow the same scoring procedure of the oral SDMT, and higher scores intuitively reflect better performance. Future studies are needed to evaluate the usefulness of these and other metrics for the detection of cognitive worsening in pwMS. Their combination could emerge as a novel, data-driven approach to characterizing individual patient profiles (e.g., 'low-speed/high-accuracy' vs. 'low-speed/low-accuracy').

Other published eSDMTs [30–37] have reported comparable reliability, although we must highlight the great methodological heterogeneity across existing studies (study sample, test–retest interval, statistical approach). Notably, our study was designed to assess reliability across test settings (in-clinic vs. at-home) rather than in time. Although our results demonstrate excellent reliability in this regard, two participants showed test–retest score differences highly suggestive of invalid performances at home. Such differences may be random (e.g., tiredness, distractions) but may also stem from systematic issues (e.g., outdated hardware, poor internet connection). Further research is needed to demonstrate eSDMT reliability for remote long-term monitoring, focusing on these potential confounders, and devising new ways to detect unreliable performances.

Concerning concurrent validity, our results are comparable, if not superior, to those obtained by other eSDMTs [30–32, 35, 37–40]. One explanation could be that most of these studies validated smartphone eSDMTs, which may be more susceptible to vision and/or motor impairment. Computer screens allow stimuli to be displayed in much larger fonts, and physical keyboards may afford higher response accuracy than touchscreen, especially for patients with upper limb spasticity and/or dexterity deficits. Moreover, we tried to closely replicate the SDMT features, by selecting symbols with the same properties (three unique symbols, two pairs of identical symbols mirrored horizontally, and one pair mirrored vertically), similar to other studies [37, 41]. Finally, we displayed only black stimuli on a white background, without unnecessary graphical elements which could reduce contrast or affect visual salience of the target stimuli, similar to other studies [37, 40]. Nonetheless, this is speculative, as conducting accurate cross-study comparisons is hindered by significant methodological differences. Notably, our strong results may represent a lower bound estimate, due to the possible increased variability in performance due to clinical fluctuations across different days. Future studies should implement designs which allow comparing computerized and smartphone eSDMT versions, to determine if test-related and device-related characteristics significantly affect validity.

Most available eSDMTs use the number of correct responses in 90s as a scoring metric, mirroring the oral SDMT [30, 31, 35–37]. Although any eSDMT based on keyboard/touch responses may potentially provide stimulus-by-stimulus RTs, our test design allows the recording of both correct responses in 90s and mean RTs reliably. Measuring timed performance with unbalanced symbols frequency distributions relies on RTs being equal across different symbols and symbol positions, with the risk of producing biased estimates. We pseudo-randomized each symbol to appear an equal number of times, with an even distribution across stimuli blocks, similar to one

previous study [38], allowing future exploratory analyses (e.g., quantifying cognitive fatigability). Our eSDMT also allows the calculation of correct responses across various time intervals. One study [37] showed that concurrent validity was maximized for eSDMT correct responses in 75 s, without noticeable improvement at 90 s. Our design allows the investigation of scores well beyond the common 90-s threshold with a data-driven approach, which may yield higher concurrent validity.

We showed that the eSDMT can reliably detect cognitive impairment in pwMS. Our results were slightly inferior to those of Lam et al. [30], although they defined cognitive impairment based only on SDMT scores, while it is well documented that verbal and/or visuospatial learning deficits are common in pwMS [42], lending greater external validity to our results, as we defined cognitive impairment based on all three BICAMS tests. One limitation of our study was the lack of a healthy control group, which would have allowed us to also assess discriminant validity between pwMS and healthy controls. Moreover, we demonstrated moderate convergent validity between eSDMT scores and MoCA score, suggesting that eSDMT scores can capture global cognitive functioning, and not just processing speed. However, there clearly are cognitive domains/functions included in the MoCA that the eSDMT is unable to capture (e.g., language, visuo-constructive abilities, executive functions). This warrants further studies to determine the validity of eSDMT as a monitoring tool for pwMS with atypical cognitive impairment profiles.

We designed the eSDMT to enable pwMS to perform frequent remote cognitive assessments independently, promoting a shift from the conventional cross-sectional 'healthy/impaired' framework to an intra-individual longitudinal 'stable/worsening' framework. The latter greatly reduces the impact of inter-individual sensory motor differences which may bias between-groups comparisons. The same also applies to potential confounders like sex, age, and education. Nevertheless, our study found that age and education only marginally affected eSDMT scores. An almost identical pattern was observed for the oral SDMT in our sample and in an Italian normative sample [24], supporting the validity of this eSDMT as an outcome measure for cross-sectional analyses.

The availability of granular intra-individual longitudinal data may also be more sensitive to cognitive worsening, since cross-sectional comparisons rely on normative data which need constant updating [43] and often fail to account for regional differences and ethnicity [44]. Moreover, intra-individual longitudinal data may be able to shed light on the inconsistent findings regarding subjective cognitive impairment [45], by focusing on intra-individual changes rather than group comparisons, which seem unable to capture subtle decline which many pwMS nevertheless perceive as meaningful [4]. However, these new tests must be able to detect whether individual longitudinal trends change due to confounding factors or due to 'pure' cognitive worsening. This was the rationale behind the inclusion of the 'Numbers' test embedded in our eSDMT procedure. With the exception of the FloodLight [35] and Konectom [41] studies, no other study reported similar dedicated screening measures, with the risk of failing to differentiate between cognitive worsening and sensory motor worsening. Performing a single screening at baseline assumes that patients will not experience sensory and/or motor fluctuations/worsening, which is certainly not the case for most pwMS [46]. Future studies must evaluate the ability of eSDMT scores to detect true cognitive worsening by leveraging a screening task preceding the cognitive task.

In conclusion, evidence supports this eSDMT as a feasible, valid, and reliable remote tool for assessing cognitive function in pwMS. Its availability may prove invaluable for research, as it allows the collection of more data, for longer periods of time, in a cost-effective way. These features allow large-scale decentralized data collection for longitudinal validation. From a clinical standpoint, such validation may enable clinicians to stop relying solely on infrequent clinical visits, thus providing a clearer clinical picture and allowing tailored interventions based on intra-individual longitudinal trends.

AUTHOR CONTRIBUTIONS

Michelangelo Dini: Conceptualization; visualization; methodology; software; investigation; data curation; formal analysis; writing – original draft. **Giulia Gamberini**: Investigation; writing – review and editing. **Marta Tacchini**: Investigation; writing – review and editing. **Angela Boschetti**: Investigation; writing – review and editing. **Alessandro Gradassi**: Resources; writing – review and editing. **Luca Chiveri**: Resources; writing – review and editing. **Mariaemma Rodegher**: Resources; writing – review and editing. **Giancarlo Comi**: Supervision; resources; writing – review and editing. **Letizia Leocani**: Conceptualization; methodology; project administration; supervision; resources; writing – review and editing.

ACKNOWLEDGMENTS

This original research paper was submitted in partial fulfilment of Michelangelo Dini's requirements for the degree of PhD in Cognitive Neuroscience at Vita-Salute San Raffaele University, Milan, Italy.

FUNDING INFORMATION

The authors received no financial support for the research, authorship, and/or publication of this article.

CONFLICT OF INTEREST STATEMENT

G.C. received consulting and speaking fees from Biogen, Merck, Novartis, Roche, Sanofi Genzyme, Almirall, Teva, Actelion, Cellgene, BMS, and Janssen-Cilag (none related to the present study). L.L. received research support from Novartis, Almirall, Biogen, and Merck and consultancy or speaker fees from Novartis, Almirall, Biogen, Merck, Janssen-Cilag, Bristol-Myers Squibb, and Roche (none related to the present study). All other authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request, pending approval from the institutional data protection officer.

ORCID

Michelangelo Dini https://orcid.org/0000-0002-2430-8902

REFERENCES

1. Mitchell AJ, Benito-León J, González J-MM, Rivera-Navarro J. Quality of life and its assessment in multiple sclerosis: integrating physical and psychological components of wellbeing. *Lancet Neurol*. 2005;4(9):556-566. doi:10.1016/S1474-4422(05)70166-6
2. Ozakbas S, Cagiran I, Ormeci B, Idiman E. Correlations between multiple sclerosis functional composite, expanded disability status scale and health-related quality of life during and after treatment of relapses in patients with multiple sclerosis. *J Neurol Sci*. 2004;218(1–2):3-7. doi:10.1016/j.jns.2003.09.015
3. Labiano-Fontcuberta A, Martínez-Ginés ML, Aladro Y, et al. A comparison study of cognitive deficits in radiologically and clinically isolated syndromes. *Mult Scler*. 2016;22(2):250-253. doi:10.1177/1352458515591072
4. Sumowski JF, Benedict R, Enzinger C, et al. Cognition in multiple sclerosis. *Neurology*. 2018;90(6):278-288. doi:10.1212/WNL.0000000000004977
5. Kalb R, Beier M, Benedict RHB, et al. Recommendations for cognitive screening and management in multiple sclerosis care. *Mult Scler J*. 2018;24(13):1665-1680. doi:10.1177/1352458518803785
6. Benedict RHB, Amato MP, DeLuca J, Geurts JJG. Cognitive impairment in multiple sclerosis: clinical management, MRI, and therapeutic avenues. *Lancet Neurol*. 2020;19(10):860-871. doi:10.1016/S1474-4422(20)30277-5
7. Benedict RHB, Deluca J, Phillips G, LaRocca N, Hudson LD, Rudick R. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler*. 2017;23(5):721-733. doi:10.1177/1352458517690821
8. Strober L, DeLuca J, Benedict RHB, et al. Symbol Digit Modalities Test: a valid clinical trial endpoint for measuring cognition in multiple sclerosis. *Mult Scler*. 2019;25(13):1781–1790. doi:10.1177/1352458518808204
9. Dillenseger A, Weidemann ML, Trentzsch K, et al. Digital biomarkers in multiple sclerosis. *Brain Sci*. 2021;11(11):1-26. doi:10.3390/brainsci11111519
10. Marziniak M, Brichetto G, Feys P, Meyding-Lamadé U, Vernon K, Meuth SG. The use of digital and remote communication technologies as a tool for multiple sclerosis management: narrative review. *JMIR Rehabil Assist Technol*. 2018;20(4):1-21. doi:10.2196/rehab.7805
11. Organisation for Economic Co-operation and Development (OECD). Internet access (indicator). 2024. doi:10.1787/69c2b997-en
12. Organisation for Economic Co-operation and Development (OECD). Access to computers from home (indicator). 2024. doi:10.1787/a70b8a9f-en
13. Miller JB, Barr WB. The technology crisis in neuropsychology. *Arch Clin Neuropsychol*. 2017;32(5):541-554. doi:10.1093/arclin/acx050
14. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17(2):162-173. doi:10.1016/S1474-4422(17)30470-2
15. Henninger F, Shevchenko Y, Mertens UK, Kieslich PJ, Hilbig BE. Lab.Js: a free, open, online study builder. *Behav Res Methods*. 2022;54(2):556-573. doi:10.3758/s13428-019-01283-5
16. Lange K, Kühn S, Filevich E. "Just another tool for online studies" (JATOS): an easy solution for setup and management of web servers supporting online studies. *PLoS One*. 2015;10(6):1-14. doi:10.1371/journal.pone.0130834
17. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699. doi:10.1111/j.1532-5415.2005.53221.x
18. Langdon DW, Amato MP, Boringa J, et al. Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). *Mult Scler J*. 2012;18(6):891-898. doi:10.1177/1352458511431076
19. Charest K, Tremblay A, Langlois R, Roger É, Duquette P, Rouleau I. Detecting subtle cognitive impairment in multiple sclerosis with the Montreal Cognitive Assessment. *Can J Neurol Sci*. 2020;47(5):620-626. doi:10.1017/CJN.2020.97
20. Benedict RHB, 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(2):145-153. doi:10.1037/1040-3590.8.2.145
21. Kim H-Y. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013;38(1):52-54. doi:10.5395/rde.2013.38.1.52
22. Tukey JW. *Exploratory Data Analysis: Past, Present and Future*. Defense Technical Information Center; 1993.
23. Conti S, Bonazzi S, Laiacona M, Masina M, Coralli MV. Montreal Cognitive Assessment (MoCA)-Italian version: regression based norms and equivalent scores. *Neurol Sci*. 2015;36(2):209-214. doi:10.1007/s10072-014-1921-3
24. Goretti B, Niccolai C, Hakiki B, et al. The Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS): normative values with gender, age and education corrections in the Italian population. *BMC Neurol*. 2014;14(1):171. doi:10.1186/s12883-014-0171-6
25. McKinney W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*. 2010; 1 (scipy):56-61. doi:10.25080/majora-92bf1922-00a
26. Vallat R. Pingouin: statistics in Python. *J Open Source Softw*. 2018;3(31):1026. doi:10.21105/joss.01026
27. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357-362. doi:10.1038/s41586-020-2649-2
28. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python Science Conference*. 2010;(Scipy):92-96. doi:10.25080/majora-92bf1922-011
29. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2
30. Lam K, van Oirschot P, den Teuling B, et al. Reliability, construct and concurrent validity of a smartphone-based cognition test in multiple sclerosis. *Mult Scler J*. 2022;28(2):300-308. doi:10.1177/13524585211018103
31. van Dongen L, Westerik B, van der Hiele K, et al. Introducing Multiple Screener: an unsupervised digital screening tool for cognitive deficits in MS. *Mult Scler Relat Disord*. 2019;2020(38):101479. doi:10.1016/j.msard.2019.101479
32. Woelfle T, Pless S, Reyes O, et al. Reliability and acceptance of dreaMS, a software application for people with multiple sclerosis: a feasibility study. *J Neurol*. 2023;270(1):262-271. doi:10.1007/s00415-022-11306-5
33. Khaligh-Razavi SM, Sadeghi M, Khanbagi M, Kalafatis C, Nabavi SM. A self-administered, artificial intelligence (AI) platform for cognitive assessment in multiple sclerosis (MS). *BMC Neurol*. 2020;20(1):1-13. doi:10.1186/s12883-020-01736-x
34. Hsu W-Y, Rowles W, Anguera JA, et al. Application of an adaptive, digital, game-based approach for cognitive assessment in multiple sclerosis: observational study. *J Med Internet Res*. 2021;23(1):e24356. doi:10.2196/24356
35. Montalban X, Graves J, Midaglia L, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. *Mult Scler J*. 2022;28(4):654-664. doi:10.1177/13524585211028561
36. Maillart E, Labauge P, Cohen M, et al. MSCopilot, a new multiple sclerosis self-assessment digital solution: results of a comparative

study versus standard tests. *Eur J Neurol*. 2020;27(3):429-436. doi:10.1111/ene.14091
37. Pham L, Harris T, Varosanec M, Morgan V, Kosa P, Bielekova B. Smartphone-based Symbol-Digit Modalities Test reliably captures brain damage in multiple sclerosis. *Npj Digit Med*. 2021;4(1):1-13. doi:10.1038/s41746-021-00401-y
38. Patel VP, Shen L, Rose J, Feinstein A. Taking the tester out of the SDMT: a proof of concept fully automated approach to assessing processing speed in people with MS. *Mult Scler J*. 2019;25(11):1506-1513. doi:10.1177/1352458518792772
39. Rao SM, Losinski G, Mourany L, et al. Processing speed test: validation of a self-administered, iPad®-based tool for screening cognitive dysfunction in a clinic setting. *Mult Scler*. 2017;23(14):1929-1937. doi:10.1177/1352458516688955
40. Middleton RM, Pearson OR, Ingram G, et al. A rapid electronic cognitive assessment measure for multiple sclerosis: validation of cognitive reaction, an electronic version of the Symbol Digit Modalities Test. *J Med Internet Res*. 2020;22(9):1-14. doi:10.2196/18234
41. Scaramozza M, Chiesa PA, Zajac L, et al. Konectom™ cognitive processing speed test enables reliable remote, unsupervised cognitive assessment in people with multiple sclerosis: exploring the use of substitution time as a novel digital outcome measure. *Mult Scler J*. 2024;1-12. doi:10.1177/13524585241259650
42. Benedict RHB, Cookfair D, Gavett R, et al. Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *J Int Neuropsychol Soc*. 2006;12(4):549-558. doi:10.1017/s1355617706060723
43. Waskowiak PT, Ruitenberg MF, Hulst HE. Neuropsychological assessment in MS is outdated and is in need for innovation: yes. *Mult Scler J*. 2024;30:150-151. doi:10.1177/13524585241230184
44. Elbulok-Charcape MM, Rabin LA, Spadaccini AT, Barr WB. Trends in the neuropsychological assessment of ethnic/racial minorities: a survey of clinical neuropsychologists in the United States and Canada. *Cultur Divers Ethnic Minor Psychol*. 2014;20(3):353-361. doi:10.1037/a0035023
45. Merlo D, Kalincik T, Zhu C, et al. Subjective versus objective performance in people with multiple sclerosis using the MSReactor computerised cognitive tests. *Mult Scler Relat Disord*. 2022;58:103393. doi:10.1016/j.msard.2021.103393
46. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. 2014;83(3):278-286. doi:10.1212/WNL.0000000000000560

How to cite this article: Dini M, Gamberini G, Tacchini M, et al. Development and validation of an electronic Symbol-Digit Modalities Test for remote monitoring of people with multiple sclerosis. *Eur J Neurol*. 2024;00:e16454. doi:10.1111/ene.16454