A Mobile Health App (DIGICOG-MS®) for Self-Assessment of Cognitive Impairment in People with Multiple Sclerosis: Instrument Validation and Usability Study

Jessica Podda, Andrea Tacchino, Michela Ponzio, Federica Di Antonio, Alessia Susini, Ludovico Pedullà, Mario Alberto Battaglia, Giampaolo Brichetto

Submitted to: JMIR Formative Research
on: January 10, 2024

Table of Contents

A Mobile Health App (DIGICOG-MS®) for Self-Assessment of Cognitive Impairment in People with Multiple Sclerosis: Instrument Validation and Usability Study

Jessica Podda[1]; Andrea Tacchino[1]; Michela Ponzio[1]; Federica Di Antonio[1]; Alessia Susini[1]; Ludovico Pedullà[1]; Mario Alberto Battaglia[2]; Giampaolo Brichetto[1]

[1]Scientific Research Area Italian Multiple Sclerosis Foundation Genoa IT
[2]Department of Physiopathology, Experimental Medicine and Public Health University of Siena Siena IT

Corresponding Author:
Jessica Podda
Scientific Research Area
Italian Multiple Sclerosis Foundation
Via Operai 40
Genoa
IT

Abstract

Background: Mobile health (mHealth) apps have proven useful for people with Multiple Sclerosis (PwMS). Thus, easy-to-use digital solutions are now strongly required to assess and monitor cognitive impairment (CI), one of the most disturbing symptoms in MS, experienced by almost 43-70% of PwMS. In this view, we developed DIGICOG-MS® (DIGItal assessment of COGnitive impairment in Multiple Sclerosis), a smartphone and tablet-based mHealth app for self-assessment of CI in MS.

Objective: This study aimed to test validity and usability of the novel mHealth app in a sample of PwMS.

Methods: DIGICOG-MS® includes four digital tests assumed to evaluate the most affected cognitive domains in MS as visuospatial memory (VSM), verbal memory (VM), semantic fluency (SF) and information processing speed (IPS), taking inspiration from traditional paper-based tests known to assess the same cognitive functions, as 10/36 Spatial Recall Test, Rey Verbal Learning Test, Word List Generation, Symbol Digit Modalities Test. Participants were asked to complete both digital and traditional assessments in two separate sessions. Convergent validity was analysed using the Pearson correlation coefficient (r) to determine the strength of the association between digital and traditional tests. To test reliability of the app, the agreement between two repeated measurements was addressed with use of the intraclass correlation coefficients (ICC). System Usability Scale (SUS) and mHealth App Usability Questionnaire (MAUQ) were thus administered after the DIGICOG-MS® evaluation to test usability of the mHealth app.

Results: The final sample consisted in ninety-two PwMS (female: 60), followed as outpatients at the AISM Rehabilitation Service of Genoa (Italy). They had a mean age of 51.38 (11.36) years, an education of 13.07 (2.74) years, a disease duration of 12.91 (9.51) years and a disability level as measured by the Expanded Disability Status Scale of 3.58 (1.75). Relapsing-remitting MS was most common (73.91%), followed by secondary progressive (16.30%) and primary progressive (9.78%) courses. Pearson correlation analyses indicated significantly strong correlations for VSM, VM, SF and IPS (all ps <.001), with r values ranging from 0.58 to 0.78 for all cognitive domains. Test-retest reliability of the mHealth app was excellent (ICCs > 0.90) for VM and IPS, and good for VSM and SF (ICCs > 0.80). Moreover, SUS score averaged 84.5 (13.34), and total score from MAUQ was 104.02 (17.69), suggesting that DIGICOG-MS® was highly usable and well appreciated by PwMS.

Conclusions: Results indicated that tests from DIGICOG-MS® strongly correlated with traditional paper-based evaluation. Furthermore, PwMS positively evaluated DIGICOG-MS®, finding it highly usable. Since CI poses major limitations to PwMS, the current findings open new paths to deploy digital cognitive tests for MS and further support the use of such novel mHealth app for cognitive self-assessment in PwMS into clinical practice.

(JMIR Preprints 10/01/2024:56074)
DOI: https://doi.org/10.2196/preprints.56074

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

A Mobile Health App (DIGICOG-MS®) for Self-Assessment of Cognitive Impairment in People with Multiple Sclerosis: Instrument Validation and Usability Study

Jessica Podda [1], Andrea Tacchino [1], Michela Ponzio [1], Federica Di Antonio [1], Alessia Susini [1], Ludovico Pedullà [1], Mario Alberto Battaglia [2], Giampaolo Brichetto [1,3]

[1] Scientific Research Area, Italian Multiple Sclerosis Foundation (FISM), Genoa, Italy

[2] Department of Physiopathology, Experimental Medicine and Public Health, University of Siena, Siena, Italy

[3] AISM Rehabilitation Service, Italian Multiple Sclerosis Society, Genoa, Italy

Correspondence to:

Jessica Podda

Scientific Research Area, Italian Multiple Sclerosis Foundation (FISM)

jessica.podda@aism.it

Via Operai 40

16149, Genoa (Italy)

Phone: +39 010 2713306

ABSTRACT

Background. Mobile health (mHealth) apps have proven useful for people with Multiple Sclerosis (PwMS). Thus, easy-to-use digital solutions are now strongly required to assess and monitor cognitive impairment (CI), one of the most disturbing symptoms in MS, experienced by almost 43-70% of PwMS. In this view, we developed DIGICOG-MS® (DIGItal assessment of COGnitive impairment in Multiple Sclerosis), a smartphone and tablet-based mHealth app for self-assessment of CI in MS.

Objective. This study aimed to test validity and usability of the novel mHealth app in a sample of PwMS.

Methods. DIGICOG-MS® includes four digital tests assumed to evaluate the most affected cognitive domains in MS as visuospatial memory (VSM), verbal memory (VM), semantic fluency (SF) and information processing speed (IPS), taking inspiration from traditional paper-based tests known to assess the same cognitive functions, as 10/36 Spatial Recall Test, Rey Verbal Learning Test, Word List Generation, Symbol Digit Modalities Test. Participants were asked to complete both digital and traditional assessments in two separate sessions. Convergent validity was analysed using the Pearson correlation coefficient (r) to determine the strength of the association between digital and traditional tests. To test reliability of the app, the agreement between two repeated measurements was addressed with use of the intraclass correlation coefficients (ICC). Usability of DIGICOG-MS® was evaluated using the System Usability Scale (SUS) and mHealth App Usability Questionnaire (MAUQ), administered at the conclusion of the digital session.

Results. The final sample consisted in ninety-two PwMS (female: 60), followed as outpatients at the AISM Rehabilitation Service of Genoa (Italy). They had a mean age of 51.38 (11.36) years, an education of 13.07 (2.74) years, a disease duration of 12.91 (9.51) years and a disability level as measured by the Expanded Disability Status Scale of 3.58 (1.75). Relapsing-remitting MS was most common (73.91%), followed by secondary progressive (16.30%) and primary progressive (9.78%) courses. Pearson correlation analyses indicated significantly strong correlations for VSM, VM, SF and IPS (all *P* <.001), with r values ranging from 0.58 to 0.78 for all cognitive domains. Test-retest reliability of the mHealth app was excellent (ICCs > 0.90) for VM and IPS, and good for VSM and SF (ICCs > 0.80). Moreover, SUS score averaged 84.5 (13.34), and total score from MAUQ was 104.02 (17.69), suggesting that DIGICOG-MS® was highly usable and well appreciated by PwMS.

Conclusions. Results indicated that tests from DIGICOG-MS® strongly correlated with traditional paper-based evaluation. Furthermore, PwMS positively evaluated DIGICOG-MS®, finding it highly usable. Since CI poses major limitations to PwMS, the current findings open new paths to deploy digital cognitive tests for MS and further support the use of such novel mHealth app for cognitive self-assessment in PwMS into clinical practice.

KEYWORDS

Cognitive assessment; Cognitive impairment; Digital health; mHealth app; Multiple sclerosis; Self-management; Usability.

INTRODUCTION

Multiple sclerosis (MS) is a chronic, inflammatory, demyelinating, neurodegenerative disease of the central nervous system with an unpredictable clinical course and highly variable clinical manifestations [1]. It predominantly affects individuals in their early adult life, and has a huge impact functionally, financially, and on quality of life [2]. Three main clinical courses (phenotypes) of MS have been identified: relapsing–remitting MS (RRMS), secondary-progressive MS (SPMS), and primary-progressive MS (PPMS). These classifications are based on the presence or absence of activity, including relapses and progression, as well as on MRI findings such as new lesions indicating inflammatory activity and atrophy suggesting ongoing neurodegeneration [2]. The most common MS symptoms are fatigue, pain, bladder and bowel issues, movement and coordination problems, visual problems, cognitive deficits and emotional lability [2]. While some symptoms are immediately obvious (e.g., balance or gait disorders), others such as fatigue, mood disorders, and cognitive problems are often hidden, referred to as “invisible” symptoms.

Cognitive impairment (CI), affecting about 43-70% of people with MS (PwMS), is recognized as a prevalent and debilitating symptom. CI is documented in all MS courses, with more severe deficits in progressive forms, both SPMS and PPMS compared to RRMS [6]. Multiple aspects of cognition can be affected in PwMS, including information processing speed, attention, executive functioning, working memory, and long-term memory [7–9]. Cognitive deficits negatively affect many aspects of PwMS' life, as level of activity and participation in daily activities, including work and social life, increased risk of falls and car accidents, reduced interaction with healthcare providers, and decreased compliance with therapy and associated benefits [10]. Although addressing CI is recognised as a key priority in MS care, leading to more timely and targeted treatment interventions, highly reliable and validated traditional paper-and-pencil tests as the Brief International Cognitive Assessment for Multiple Sclerosis [11,12], the Brief Repeatable Battery of Neuropsychological tests [13] and the Minimal Assessment of Cognitive Function in MS [14] are still not widely and routinely used in clinical practice. Even in specialized centers, systematic cognitive assessments are somewhat time-consuming, require specialized examiners, and may not be well accepted by PwMS [15]. Moreover, as MS progresses, traditional paper-based tests may not timely detect subtle cognitive changes. Thus, these limitations may be overcome by alternative easy-to-use digital solutions.

Digital and remote technologies are involved in challenges of great interest for the current and future research in MS, such as the improvement of safety, autonomy, and well-being, during daily activities. Considering the high cost of providing healthcare to people with a neurological chronic disease and the increasing technological advancements in rehabilitation field, the time is right to move from traditional standard care center based-program toward the implementation of innovative and alternative tools in routine practice with high therapeutic benefits [9].

Telemedicine has enabled convenient and effective follow-up visits, reduced unnecessary testing and referrals, maintained good perception of care, and, not least, reduced travel costs and caregiver burden [16]. The increasing use of smartphones and tablets has made mobile health (mHealth) apps a promising tool for empowering and engaging people in the self-management of their health [17]. In this scenario, mHealth apps can screen individuals who are at risk and/or offer self-help intervention or clinical referrals for various health conditions. They represent vehicles for enhanced real-time data capture to screen for, monitor and treat CI in MS that may fundamentally shift traditional paradigms of medicine [18–23]. However, while technology promises to offer PwMS alternative methods for tracking and self-managing symptoms, communicating with their healthcare providers, and improving their health-related quality of life, it is unclear if available mHealth tools are meeting their needs [24]. For instance, while Salimzadeh et al. [17] found 104 MS-related apps in iTunes and Google Play, they noted that there was no corresponding evidence regarding the usability and utility of these solutions in PwMS. To be employed in clinical practice, mHealth tools for cognition should be validated, and their usability and acceptance by end users should be tested.

Given the need to understand more in depth the potentiality of mHealth apps in the assessment of CI in MS, we designed and developed DIGItal assessment of COGnitive impairment in Multiple Sclerosis (DIGICOG-MS®), a smartphone and tablet-based app for self-assessment of CI in PwMS. Aims of our study were to test validity and usability of the novel mHealth app in a sample of PwMS. Furthermore, we evaluated whether demographic and clinical variables would explain any variability in usability and acceptability of DIGICOG-MS®. We hypothesized that: 1) since consistent evidence indicates that scores obtained in traditional and electronic assessments did not differ [20,25], individuals' performances at traditional paper-based and digital assessments would be highly comparable; 2) DIGICOG-MS® would be overall well appreciated by PwMS.

METHODS

Participants

The study sample consisted of PwMS enrolled by the Italian MS Foundation and receiving outpatient care at the AISM Rehabilitation Service in Genoa (Italy). Potential eligible participants were identified by a trained neuropsychologist based on medical reports and were subsequently contacted by phone. Inclusion criteria were an age of 18 years and older, a confirmed MS diagnosis following the McDonald criteria [26], all disease course (RRMS, SPMS, PPMS), the absence of relapsing in the last 3 months, an Expanded Disability Status Scale (EDSS) score [27] ≤ 7, adequate visual, hearing and motor capabilities to work on a tablet. Exclusion criteria were a Montreal Cognitive Assessment score < 18, neurological and major psychiatric illness, past serious head trauma, and alcohol or drug abuse.

Upon enrollment and provision of written consent, each participant underwent two in-person evaluation sessions at the AISM Rehabilitation Service in Genoa, Italy. These sessions, each lasting approximately 30 minutes, were spaced apart by a convenient time interval of up to ten days to minimize potential carryover learning effects. A trained neuropsychologist administered paper-based tests, while digital tests were self-administered by PwMS under supervision. The order of digital and paper-based assessments was counterbalanced.

Demographic information (age, gender, education) and disease-related data (MS course, disease duration, and EDSS score) were collected for each participant. Furthermore, patient-reported outcomes (PROs) were assessed for self-reported mood (Hospital Anxiety and Depression Scale-HADS) [28], upper-limb functionality (Arm Function in Multiple Sclerosis-AMSQ)[29], and impact of MS (Multiple Sclerosis Impact Scale-MSIS-29)[30]. Usability of the app was evaluated using the System Usability Scale (SUS) [31,32] and mHealth App Usability Questionnaire (MAUQ)[33], administered at the conclusion of the digital session. The 10-item SUS evaluates users' personal perceptions about how to use a given system or device, ranging from "strongly disagree" (1 point) to "strongly agree" (5 points). Technology in general has an average score of 60 [34]. The SUS score ranges from zero (lowest usability) to 100 (highest usability), with a value of 68 considered above average. The novel 18-item MAUQ is a validated and reliable questionnaire to assess mHealth app. Each item's score ranges from 1 to 7, which corresponds to the degree of agreement of the digital tool. The highest possible score is 126, and the lowest is 18, with an average total score equal to or greater than 72 indicating that the app is usable [33]. This questionnaire provides three subscales: Ease of use, Interface and Satisfaction, and Usefulness. Furthermore, MAUQ is designed for different users (patients or health care providers) and different interaction modes (interactive or standalone). In standalone mHealth apps, the app users enter/collect/store health information about themselves or other people. The standalone apps may generate reminders or show a summary or details about the collected health information, but these apps do not send the data to the user's health care providers or patients. Thus, for study purpose, the standalone patient version of the scale was administered to PwMS. Data collection occurred between January 2023 and November 2023.

DIGICOG-MS® tool

DIGICOG-MS® (intellectual property of Italian Multiple Sclerosis Foundation; SIAE Registration ID: D000018162, 27-12-2022) is a mHealth app developed to self-assess and monitor the presence of CI (See Figure 1). The app, supported on Android and iOS, facilitates data collection, storage and presentation, incorporating analysis algorithms and a clinician dashboard for user management and data extraction. The frontend of the app was developed using the open-source framework Angular, a JavaScript-based TypeScript development language known for its efficiency and speed. A dedicated backend service for DIGICOG-MS® functionality has been implemented using .Net Standard 6.

DIGICOG-MS® integrated four digital tests, inspired by paper-based tests and assumed to measure most affected cognitive domains in MS as visuospatial memory (VSM), verbal memory (VM), semantic fluency (SF) and information processing speed (IPS) [35,36]:

- *Remember and place* assesses visuospatial episodic memory. A 36-square grid with 10 black checkers is displayed on the screen for ten seconds. After the time elapses, the pattern disappears, and participants must reproduce it on a blank checkerboard. This replicates the 10/36 Spatial Recall Test (SPART) [36], in which a 6 x 6 checkerboard with ten pieces arranged in a particular pattern is shown to the subject for ten seconds. Both tests (digital and traditional) include three consecutive trials, and the score consists in the total number of correct responses for the three trials.

- *Listen and repeat* was developed as an electronic version of the Rey Verbal Learning Test (RAVLT) [37], that evaluates verbal memory. Participants listen to a prerecorded list of 15 common nouns and are asked to recall as many words as possible for five times. Responses are recorded and then scored by the neuropsychologist. In the traditional test, words are read aloud to the subject who is asked to repeat as many words as possible in any order. All pronounced nouns in each of the five learning trials are transcribed by the neuropsychologist. For both versions of the test, total score consists in the number of words recalled across the five trials.
- *Generate words* is a digital adaptation of the Word List Generation (WLG) [36] and measures semantic verbal fluency. Participants generate a list of words, typically constrained by a specific semantic category, in 90 seconds. Recordings of pronounced words are processed by the neuropsychologist for scoring. In the traditional test, all words generated within the given semantic category are transcribed by the neuropsychologist. Total score is based on the number of correct words produced.
- *Associate numbers* represents a modified electronic version of the Symbol Digit Modalities Test (SDMT) [11,36], measuring attention and information processing speed. Participants are required to match single digits (1–9) to corresponding symbols as accurately and quickly as possible clicking on a keyboard. A key pairing digits with symbols is available for the entire duration of the test, with a practice phase of 4 items provided to familiarize participants with the task. A 90-second timer is added to the system to avoid using a stopwatch, automatically terminating the procedure. In the traditional test, participants are presented with a page headed by a key of 9 symbols, and their task is to orally report the correct corresponding

number. After completing the first 10 items with guidance, participants are timed to determine how many responses can be made in 90 seconds. In both versions of the test, the total score is based on the number of correct answers provided within 90 seconds. Additionally, in the digital test also the number of correct pairs at 30 sec and 60 sec are provided.

Table 1 outlines digital and traditional tests to assess VSM, VM, SF and IPS. Although DIGICOG-MS® tests were developed to closely mirror the original paper-based versions, minor adjustments were made during the digitalization process. Notably, substantial variations were implemented for the IPS digital test due to copyright restrictions associated with the SDMT (i.e., different symbols and pairing have been settled). In this study, all PwMS used the app with Android OS. Although the current version of DIGICOG-MS® is in Italian language, the app was designed with scalability in mind, allowing for easy extension to other languages.

Table 1. Digital, delivered through DIGICOG-MS®, and traditional tests administered to PwMS.

Cognitive domains	Digital	Traditional
Visuospatial memory	Remember and place	10/36 SPART[a]
Verbal memory	Listen and repeat	RAVLT[b]
Semantic fluency	Generate words	WLG[c]
Information processing speed	Associate numbers	SDMT[d]

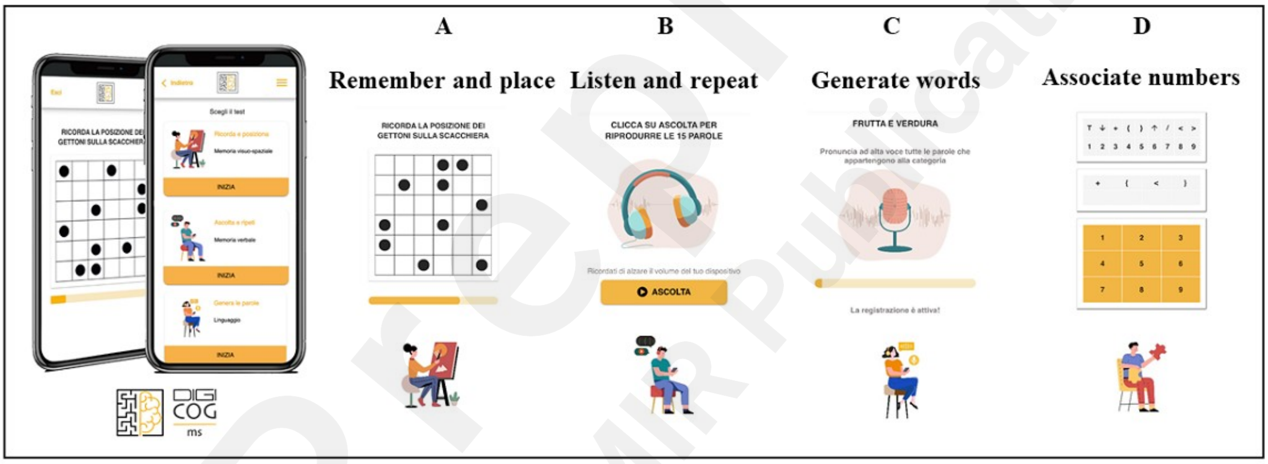
[a] 10/36 Spatial Recall Test

[b] Rey Verbal Learning Test

[c] Word List Generation

[d] Symbol Digit Modalities Test

Figure 1. Visualization of DIGICOG-MS®. The four digital tests implemented in DIGICOG-MS® that assess visuospatial memory (A), verbal memory (B), semantic fluency (C), and information processing speed (D).



Statistical Analysis

Categorical data were summarized by numbers and percentages, while numerical data were indicated by the mean and standard deviation (SD).

Convergent validity was computed using the Pearson correlation coefficient r to determine the strength of the association between digital (i.e., Remember and place, Listen and repeat, Generate words, Associate numbers) and traditional (i.e., 10/36 SPART, RAVLT, WLG and SDMT) tests. Pearson correlation was considered as strong ($r \geq 0.5$), moderate ($0.5 > r > 0.3$), and weak ($0.3 > r >$

0.1) [38]. Known as common tool to compare two methods of measurement to determine the overall degree of agreement, Bland-Altman scatterplots were created to estimate the disagreement between traditional and digital tests as a function of the mean of the two measurements for each cognitive domain. On the y-axis, the difference between every pair of measurements for each cognitive domain is plotted, and the x-axis shows the mean of these measures. Horizontal reference lines on the plot represent the average difference between the measurements (dotted line), which is expected to be zero for a good level of agreement; and limits of agreement (sharp lines), which are fitted on the plus and minus 1.96 SD. When there was only a single score without a pair (i.e., digital or traditional assessment was missing), the subject was excluded from the final analysis.

The test-retest reliability of DIGICOG-MS® was addressed with use of the intraclass correlation coefficients (ICC) in a cohort of PwMS two weeks apart the first digital evaluation. Using power analysis calculations to test reliability between two different observations described by Bujang and Baharum (2017) [39], a minimum of 22 individuals is needed to have an acceptable ICC value ≥ 0.5 (α = 0.05, power = 80%, n = 2). A two-week window was chosen to reduce variability that is independent of general disease status and might be attributed to good or bad days or to differences between weekdays and weekends [20]. A two-way mixed effects model on absolute agreement for single measurements was used. Reliability was considered as excellent (ICC ≥ 0.9), good (ICC ≥ 0.75), acceptable (0.75 > ICC ≥ 0.5) and poor (ICC < 0.5) [40].

To investigate potential factors influencing the usability as measured by SUS and MAUQ, a univariate linear regression was conducted using a selection of predictors selected from demographic, i.e., age and educational level, and clinical variables, i.e., EDSS, MS course, MS duration, HADS-a (anxiety subscale), HADS-d (depression subscale), MSIS-29, and AMSQ. Min-max rescaling was previously applied to PROs items. The median values were employed to fill in the missing values. All confidence intervals were two-sided and will use 95% confidence levels. The *P*-values <.05 were considered statistically significant. Statistical analysis was performed using STATA (version 17) and MATLAB (version 9.14.0 (R2023a) Update 4).

Ethical considerations

The study was approved by Regional Ethics Committee of Azienda Ospedaliera "San Martino" of Genoa (Italy) (N. 240/2022 - DB id 12354) and conducted according to the Declaration of Helsinki [41]. Before entering the study, participants had to read, complete, and sign an informed consent. Data collected were stored in an anonymized format to properly protect the privacy and confidentiality of participants, ensuring that no individual can be identified from the data provided. Participants have been informed that data collection could be used only for research purposes.

RESULTS

At the beginning, 105 PwMS met eligibility criteria and then were contacted by the neuropsychologist. Seven PwMS postponed the first evaluation due to multiple reason (e.g., health problems, travel constraints) several times until the neuropsychologists decided to withdraw them from the study. Ninety-eight PwMS were enrolled. Six participants were excluded from the analyses due to missing score in one assessment (i.e., digital or traditional modality). The final sample consisted in 92 PwMS (female: 60). A summary of demographic and clinical characteristics is reported in Table 2.

Table 2. Demographic and clinical sample characteristics (N = 92).

Characteristic		Value
Age (in years), mean (SD)		51.38 (11.36)
Min-max		(20-69)
Gender, n (%)		
	Female	60 (65.20%)
	Male	32 (34.80%)
Years of education, mean (SD)		13.07 (2.74)
Min-max		(8-22)
Disease duration, mean (SD)		12.91 (9.51)
Min-max		(1-32)
Disease course, n (%)		
	RRMS [a]	68 (73.91%)
	SPMS [b]	15 (16.30%)
	PPMS [c]	9 (9.78%)
EDSS, mean (SD)		3.58 (1.75)
Min-max		(1-7.5)
HADS [d], mean (SD)		
Min-max		
	HADS-d [d]	5.87 (4.17)
		(0-19)
	HADS-a [d]	7.6 (4.9)
		(0-19)
AMSQ [e], mean (SD)		54.14 (28.10)
Min-max		(31-136)
MSIS-29 [f], mean (SD)		65.9 (27.6)
Min-max		(29-136)

[a] Relapsing-Remitting Multiple Sclerosis

[b] Secondary Progressive Multiple Sclerosis

[c] Primary Progressive Multiple Sclerosis

[d] Hospital Anxiety and Depression Scale

[d] Hospital Anxiety and Depression Scale – depression subscale

[d] Hospital Anxiety and Depression Scale – anxiety subscale

[e] Arm Function in Multiple Sclerosis

[f] Multiple Sclerosis Impact Scale

Correlation between digital and traditional assessment tests

Pearson correlation analysis indicated significantly strong correlations for VSM, VM, SF and IPS (all *P* values <.001), with associations ranging from 0.58 to 0.78 (See Table 3 and Figure 2). Bland–Altman plot for each cognitive domain showed a larger systematic bias for IPS, with a mean difference of -26.01 (limit of agreement: 3.84; -48.18). For VM and SF, the mean difference was close to zero (1.54 and 0.85, respectively), indicating that digital scores were slightly larger than traditional ones, while for the VSM, a negative mean difference was observed (-1.16) (Figure 2).

Table 3. Correlation coefficients (Pearson r and two-tailed *P* value) between digital and traditional tests.

Variable	VSM traditional	VM traditional	SF traditional	IPS traditional
VSM digital				

r	0.580	0.454	0.248	0.526
P value	<.001	<.001	0.017	<.001
VM digital				
r	0.439	0.785	0.405	0.649
P value	<.001	<.001	<.001	<.001
SF digital				
r	0.465	0.392	0.636	0.625
P value	<.001	<.001	<.001	<.001
IPS digital				
r	0.518	0.519	0.539	0.720
P value	<.001	<.001	<.001	<.001

[a] Visuospatial memory

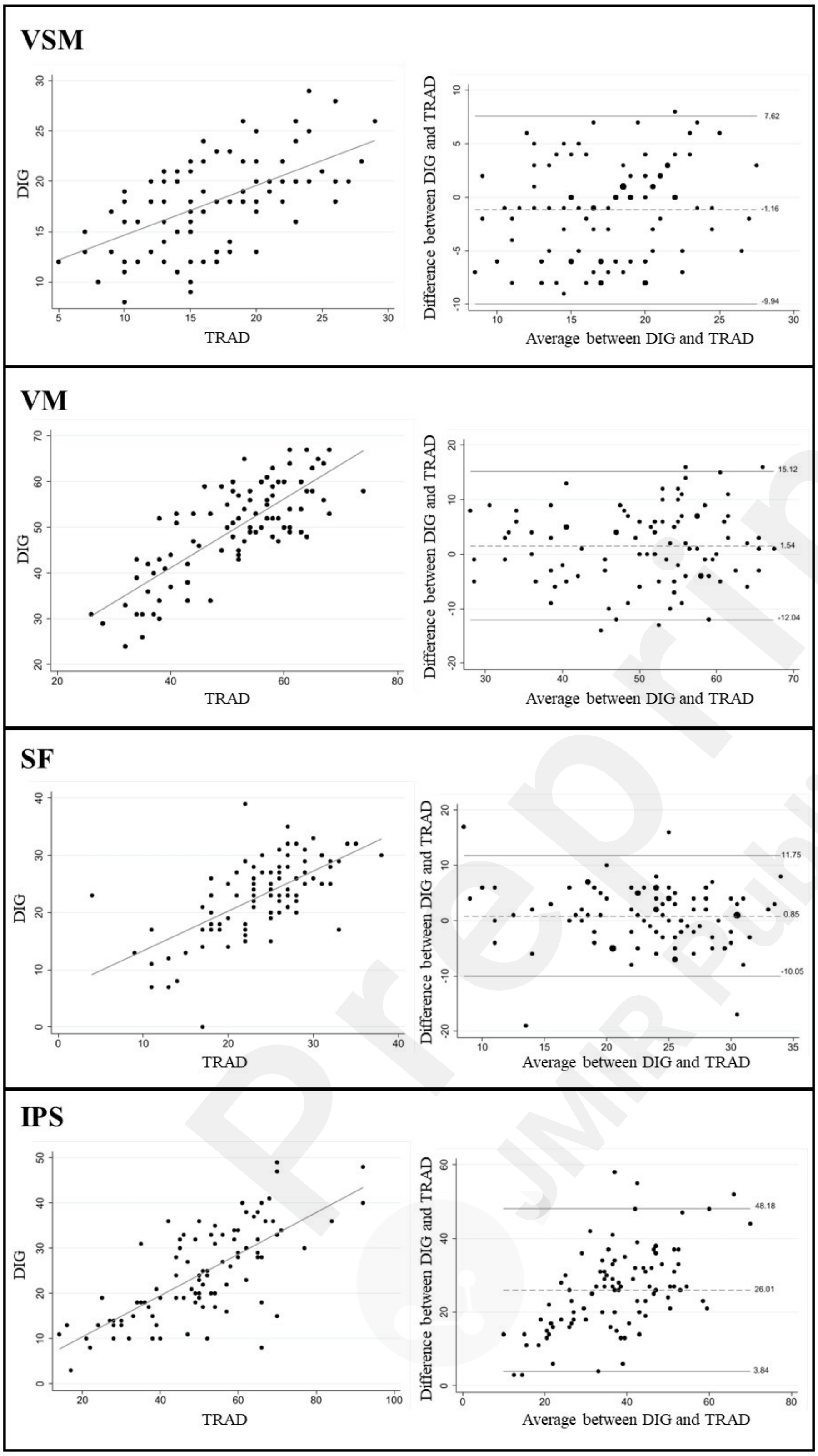
[b] Verbal memory

[c] Semantic fluency

[d] Information processing speed

Figure 2. Scatter and Bland–Altman plots of digital and traditional assessment tests for visuospatial memory (VSM), verbal memory (VM), semantic fluency (SF) and information processing speed (IPS). Correlations between digital (y-axis) and traditional (x-axis) are shown (first column). In the Bland-Altman plots (second column), agreement between DIG and TRAD is shown. The dashed line represents the mean difference, that is expected to align with y = 0 for an optimal level of agreement; the solid lines represent the 95% limits of agreement, within which all

points are expected to fall. Following digital and traditional tests were used to measure cognitive domains frequently altered in MS: Remember and place and 10/36 Spatial Recall Test (VSM); Listen and repeat and Rey Verbal Learning Test (VM); Generate words and Word List Generation (SF); Associate numbers and Symbol Digit Modalities Test (IPS).



Test-retest reliability

Test-retest reliability values of the mHealth app were obtained from a randomly selected subset of

PwMS (n = 27). Results showed excellent reliability for VM and IPS (ICCs ≥ 0.95), and good for VSM and SF (ICCs ≥ 0.83). Scores from digital tests by DIGICOG-MS® increased during the second test after two weeks, likely due to practice effects. Results of the test-retest reliability are summarized in Table 4.

Table 4. The test-retest reliability of DIGICOG-MS® tests on four cognitive domains (n=27)

Variable	Time 1	Time 2		
	Mean (SD)	Mean (SD)	Mean Difference (SD)	ICC [e] (95%CI)
VSM	17.6 (4.8)	18.7 (4.7)	1.11 (2.50)	0.84 (0.67-0.93)
VM	50.4 (11.2)	51.9 (10.9)	1.52 (2.76)	0.96 (0.89-0.98)
SF	22.8 (7.6)	23.8 (6.9)	1.04 (4.12)	0.83 (0.67-0.92)
IPS	23.9 (11.4)	25.3 (10.6)	1.44 (3.08)	0.95 (0.89-0.99)

[a] Visuospatial memory

[b] Verbal memory

[c] Semantic fluency

[d] Information processing speed

[e] Intraclass Correlation Coefficient

DIGICOG-MS® usability testing

SUS score averaged 84.5 (SD: 13.34), indicating “best imaginable” usability (range 84.1-100) [42]. Total score from MAUQ was 104.02 (SD: 17.69), well above the cut-off of 72, suggesting that DIGICOG-MS® was significant usable and well appreciated by PwMS [33]. Subscales were scored as follows: Ease of use: 30.57 (SD: 5.18); Interface and satisfaction: 42.96 (SD: 7.93); Usefulness: 30.49 (SD: 7.14). Depressive symptoms, as indicated by HADS-d, negatively correlate with both MAUQ (P < .001, coefficient = –0.31) and SUS (P = .02, coefficient = –0.17), and show the major effects on usability outcome measures compared to all other covariates; anxiety from HADS-a was only associated with MAUQ (*P* = .04, coefficient = –0.16), even if with a minor but confirmative effect. See Table 5 for a summary of usability total scores from SUS and MAUQ.

A higher level of education shows a positive correlation with Ease of use subscale from MAUQ (*P* = .03, coefficient = 0.012), while EDSS correlates with Usefulness (*P* = .03, coefficient = 0.025). Both anxiety and depression appeared to have a significant impact on mHealth app usability. HADS-d was inversely proportional to all MAUQ subscales (from *P* < .001 to *P* = .03, coefficients from -0.2 to -0.37); HADS-a was inversely proportional to the Interface and satisfaction subscale only (*P* = .016, coefficient = -0.21). Interestingly, upper limb functionality by AMSQ has an important negative effect (coefficient -0.25) on individual evaluation of the Interface and satisfaction subscale (*P* = .02). Except for HADS-d, neither clinical nor demographic variables correlated with the SUS score (see Table 6 for a detailed overview of MAUQ subscale results).

Table 5. Associations between usability as indicated by SUS and MAUQ total scores and demographic and clinical variables.

Variable	MAUQ [a] total		SUS [b] total	
	Coefficients (SE)	*P* values	Coefficients (SE)	*P* values
Age	-0.0015 (0.0014)	.28	-0.0021 (0.0012)	.09
Years of education	0.005 (0.006)	.41	0.010 (0.005)	.05
MS course (RR) [c]	0.013 (0.04)	.76	0.033 (0.038)	.39
Gender (M)	-0.006 (0.032)	.85	0.039 (0.029)	.19
Disease duration	-0.0014 (0.0016)	.37	-0.0022 (0.0015)	.14
EDSS [d]	0.007 (0.009)	.45	-0.006 (0.008)	.50
MSIS-29 [e]	-0.10 (0.07)	.19	-0.11 (0.06)	.08
AMSQ [f]	-0.145 (0.096)	.13	-0.15 (0.08)	.05
HADS-a [g]	-0.16 (0.07)	.04	-0.11 (0.06)	.08
HADS-d [g]	-0.31 (0.08)	<.001	-0.17 (0.07)	.02

[a] mHealth App Usability Questionnaire

[b] System Usability Scale

[c] Relapsing-Remitting Multiple Sclerosis

[d] Expanded Disability Status Scale

[e] Multiple Sclerosis Impact Scale

[f] Arm Function in Multiple Sclerosis

[g] Hospital Anxiety and Depression Scale – anxiety subscale

[g] Hospital Anxiety and Depression Scale – depression subscale

Table 6. Correlations between MAUQ subscales and demographic and clinical variables.

Variable	MAUQ [a] subscale Ease of use Coefficients (SE)	*p* values	MAUQ [a] subscale Interface and satisfaction Coefficients (SE)	*p* values	MAUQ [a] subscale Usefulness Coefficients (SE)	*p* values
Age	-0.0025 (0.0013)	.07	-0.0020 (0.0015)	.21	-0.0001 (0.0018)	.97
Years of education	0.0123 (0.005)	.03	0.003 (0.006)	.62	0.000 (0.008)	.99
MS course (RR)[b]	0.014 (0.04)	.73	0.01 (0.05)	.86	0.02 (0.06)	.83
Gender (M)	0.001 (0.032)	.97	-0.013 (0.037)	.73	-0.01 (0.04)	.91
Disease duration	-0.0025 (0.0016)	.13	-0.0016 (0.0018)	.40	-0.0005 (0.0022)	.83
EDSS [c]	-0.006 (0.009)	.51	0.000 (0.010)	.98	0.025 (0.012)	.03
MSIS-29 [d]	-0.08 (0.08)	.32	-0.14 (0.09)	.11	-0.07 (0.09)	.45
AMSQ [e]	-0.12 (0.10)	.24	-0.25 (0.11)	.02	-0.04 (0.12)	.72
HADS-a [f]	-0.11 (0.08)	.16	-0.21 (0.09)	.016	-0.13 (0.09)	.18
HADS-d [f]	-0.20 (0.09)	.03	-0.33 (0.10)	.001	-0.37 (0.10)	< .001

[a] mHealth App Usability Questionnaire

[b] Relapsing-Remitting Multiple Sclerosis

[c] Expanded Disability Status Scale

[d] Multiple Sclerosis Impact Scale

[e] Arm Function in Multiple Sclerosis

[f] Hospital Anxiety and Depression Scale – anxiety subscale

[f] Hospital Anxiety and Depression Scale – depression subscale

DISCUSSION

Principal Findings

This study aimed to determine whether DIGICOG-MS®, a mHealth app for self-cognitive assessment including digital tests for visuospatial memory (VSM), verbal memory (VM), semantic fluency (SF), and information processing speed (IPS), could effectively evaluate cognitive impairment (CI) in adults with MS. Overall, the findings revealed strong correlations between digital and traditional paper-based tests across all cognitive domains, with correlation coefficients (r) ranging from 0.58 to 0.78. Test-retest reliability was excellent for VM and IPS (ICCs ≥ 0.95) and good for VSM and SF (ICCs ≥ 0.83). Additionally, PwMS positively evaluated the DIGICOG-MS® app, finding it highly usable.

Comparison With Prior Work

For VSM, no previous studies on digital SPART assessment in MS are currently available. However, van Dongen et al. [19] tested on a group of 235 healthy subjects an electronic version of the SPART, reporting good consistency for digital and traditional tests. Although significant (r = 0.58), the correlation between SPART and "Remember and place" task was lower compared to others cognitive domains investigated. One speculation could be the intrinsic heterogeneity in the task procedure: in paper-based version, participants had to pick and move black coins in the correct position on the checkerboard, while in the digital test individuals had to tap on the empty grid. Although a swipe command was more similar to the gesture of pick and place [19], we decided to ask participants to click on the screen to avoid that participants could meet any difficulty in pressure on the screen device.

Regarding VM domain, results are in line with a previous study that developed an electronic version of RAVLT in PwMS [18]. However, in the study by Beier et al. [18], the word list was available on screen, and as the participants verbally recalled words, a single tap on each word recorded a correct response. The highly significant correlation (r = 0.78) observed in our study reflects that the digital version resembles the traditional one in terms of stimuli administration and scoring procedures.

A similar pattern was observed for SF task. To our knowledge "Generate words" is the first digital test to evaluate fluency for semantic stimuli in MS. A strong correlation with the paper-based versions of the WLG (r = 0.64) was observed, reflecting an excellent similarity in the administration procedure between the two modalities.

Notably, although the correlation between traditional and digital test for IPS was strong (r = 0.72) and comparable to the previously reported associations between smartphone-based and paper-based versions of the SDMT [20,43], individual performances at "Associate numbers" tended to be lower than the oral SDMT. In a recent study by Costabile et al. [22], that tested potentiality of iBICAMS, authors found a very small mean difference between two administration modalities (M= -2.72). However, they administered an electronic SDMT as if it was a paper-based test and use the iPad as a backend for neuropsychologists to record the answers given orally by participants. Our decision to implement a written version for the digital IPS test was to allow PwMS to complete the task in a fully unsupervised and autonomous way. Montalban et al. [20] reported several possible sources of performance discrepancy. Firstly, in the digital test, a longer time is required to tap the correct response on a keyboard compared to saying the correct response aloud. Secondly, the digital IPS test displayed four symbols at a time, while the SDMT, on the contrary, provides participants the entire symbol key printed on a sheet of paper, leading to a massive involvement of eye tracking and visual working memory. In addition, since the SDMT is recognized as the recommended screening tool for CI in MS [44], most of PwMS may have familiarity with this test and thus have memorized the symbol-digit code of the paper-based test.

Although the potential benefits of digital tools in terms of assessment and prediction of cognitive changes over time, and thus plan tailored rehabilitation intervention, are confirmed, the usability of the proliferating mHealth apps is sometimes unknown and under investigated [45]. The majority of studies investigating the potential of novel digital tools for assessment purpose missed or did not report a detailed usability examination [18,19,22]. A meaningful mHealth app for PwMS needs to be usable, useful, and satisfying, thus understanding its usability provides insight into the quality and overall satisfaction of the user's experience. In our study, PwMS positively evaluated DIGICOG-MS®, finding it highly usable and motivating.

Overall, an inversed correlation between depressive symptoms and usability was found. Compared to other covariates, these symptoms exhibit the greatest effect strength, closely followed by anxiety. This is in line with Ly et al. [46] that noted the key impact of major depressive illness on participant's ability to interact with mHealth apps in general. Since the user experience of a digital tool is crucial, especially as it might be relevant in avoiding drop-out, the biased impression of a person with depressive symptoms and resulting needs of those users should be considered when designing and evaluating digital (mental) health platforms [47]. Literature supports this concern with Torous, Nicholas et al. [48] advising that many apps are not designed with the end user primarily in mind and that apps are unhelpful in emergencies. They reported a study by Sarkar et al. [49] in which more than 50% of participants with major depressive disorder had some degree of difficulty in both entering and accessing their mood data across four common and popular mood tracking apps.

As expected, upper limb functionality correlates with Interface and satisfaction subscale, confirming that upper limb limitations can also prevent people, especially elderly, from using mHealth app [50]. Indeed, the degree of influence exerted by the upper limb functionality on the appreciation of the benefits provided by the digital tool is substantial and comparable to that of mood domain. Correlation between Ease of use and educational level is in line with Marrie et al. [51] noted that an higher educational level have been associated with greater likelihood of mHealth application use. Since the app was rated as having overall "best imaginable" levels of usability, findings suggest that such positive experience could increase willingness to use DIGICOG-MS®, leading to a greater adherence to future interventions proposed by clinicians. It is important to acknowledge that involving end users in different phases of an app development (i.e., conception, design and testing), working closely with them to learn what their needs are and formulate how an app may even be of use, is a useful approach to increase usability and engagement..

Limitations and Strengths

We acknowledge some limitations of this study and suggest avenues for future research.

Firstly, although digital tests from DIGICOG-MS® were developed to be as similar as possible to the original paper-based versions, some changes in digital test development could affect participants' performance. In this study, a fully automated modified version of the SDMT has been proposed, requiring PwMS to tap the correct answer through a number keyboard. However, we believe that an oral version of such a digital test will minimize the influence of hand disability on task performance [52,53] and thus accommodate more PwMS' characteristics [54]. Future developments of DIGICOG-MS® will include an integration with a speech recognition software, allowing for a fully automated use. This implementation will further provide an automatic score for those tests that required a speech registration (i.e., VSM and VM test). To ensure proper clinical implementation of DIGICOG-MS® as screening tool, a sample of healthy subjects should be enrolled in a further study to obtain norms for each digital test enabling interpretation in a clinical setting (e.g. indicate normal functioning or the presence of mild, moderate, or severe impairment).

Secondly, the digital assessment was conducted on-site under the guidance of a trained neuropsychologist. While this approach may have some limitations in terms of generalizability to real-life scenarios, it was a deliberate choice. Our sample lacked prior experience with digital tools for remote cognitive assessment. Thus, we opted to conduct validation and usability tests in a controlled clinical setting. This allowed participants to familiarize themselves with the novel technology and provided us with valuable feedback to enhance the app's quality.

Moving forward, our future research endeavours will involve deploying DIGICOG-MS® in a more realistic environment, such as the homes of PwMS. This approach will enable us to explore the tool's potentialities in a setting that more closely mirrors everyday life, thus addressing concerns regarding generalizability.

Thirdly, as a single-center study, the heterogeneity of personal and clinical characteristics of our sample, ranging from younger, active, and autonomous individuals to older, sedentary, and dependent people, leads to a wide spectrum of symptoms and different needs [55]. Participants' characteristics may limit the interpretation of our results in terms of both validity and usability. Study sample may be considered representative of those clinic-attending PwMS followed as outpatients in rehabilitation centers (i.e., middle-age/older adults and with a longer disease duration) [56]. Thus, results may not generalize to other populations of individuals with MS (e.g., young and neo-diagnosed people).

Fourthly, it is known that previous experience with technologies increases acceptance for technological approaches and some older adults could have less experience and/or interest in using their smartphone regularly than younger MS individuals. This aspect should be taken into account when designing and development such a digital tool for a wide variety of PwMS' needs and characteristics [51].

As a digital and remote communication-technology app, DIGICOG-MS® was developed for clinical monitoring and self-evaluation, as an alternative or complementary to traditional in-clinic traditional assessment that could be implemented rapidly and routinely to address CI issue in MS. To our knowledge, this is the first novel mHealth app developed for tablet and smartphone in Italian language and validated in a sample of PwMS (available for download from Google Play store or Apple Store soon). Ultimately, although this version will be susceptible of changes, it could be carried out at PwMS' home, helping those people who are unable to access clinics easily, for various reasons such as mobility restrictions, travel costs, consultation and treatment time constraints, and a lack of locally available MS expert services [57]. This will reduce effectively the distance between patients and their professionals.

A self-evaluation cognitive tool, like DIGICOG-MS®, brings potential advantages. Firstly, digital administration procedures are more standardized than in the paper-based version, reducing inter-rater differences and/or assessor dependent errors in stimuli presentation (e.g., a pre-recorded audio with an appropriate pace and tone could avoid prosody inflections that would suggest the position of the word within the list in the VM test), thus improving accuracy. Secondly, an app could reduce high stress levels and demands for clinicians resulted to repeat over time administration procedures. This aligns results from previous studies that highlighted the potentialities of a digital tool for cognitive assessment in MS [23,58,59]. As indicated by Tacchino et al. [60], digital cognitive assessments, expected to occur more frequently, at home or in an unsupervised setting, could also support the integration with electronic PROs of other dimensions significant for MS (e.g., mood). Brichetto et al. (2020) [61] found that PROs and clinician-assessed outcomes (CAOs) could be used to build accurate models of MS disease course prediction (i.e., transition between a relapsing-remitting form to a secondary progressive form). In addition, PROs and CAOs, alone or integrated with other indexes such as MRI outcomes and biomarkers, could help the decision-making process of clinicians in their daily practice. The availability of longitudinal multi-domain big-data could allow the application of revolutionary technology, as digital twin, to MS cognitive phenotyping [62–64], where analysis of big data through artificial intelligence enables visualization of a virtual copy (twin) of the patient at different stages of the disease and supports further therapeutic decisions [65]. In conclusion, the possibility to better define the clinical complexity levels of each individual with MS and have sufficient and adequate predictive criteria for MS evolution through a novel digital solution like DIGICOG-MS® could be pivotal for the construction of a more fruitful therapeutic pact between patient and clinician based on better perspective knowledge, increased disease consciousness and engagement in rehabilitative treatments.

Conclusion

Since CI poses major limitations to PwMS, the current results open up new paths to deploying digital cognitive tests for MS and further support the use of DIGICOG-MS® for cognitive self-assessment in PwMS into clinical practice. Follow-up measurements will be easier to implement and could lead to timely identification of cognitive decline in PwMS and subsequently allow for adequate counseling. Deeper knowledge on cognitive phenotypes to detect CI early, more accessible, and tailored interventions, and better understanding of multitasking deficits in everyday life activities should be considered the main goals by current and future research in MS. Thus, a mHealth app as DIGICOG-MS® could be one of the available technological solutions necessary to address these goals.

Acknowledgments

We would like to thank Maria Madera and Giulia Bignone for their help in participants' enrollment, Salvatore Mulè and Francesco Armellino for their effort in the digital fundraising initiative #Stream4AISM that helped us to promote the study, and Ct Solution for the support in the app development. Finally, a special thanks to all the people with Multiple Sclerosis that took part to our study. Jessica Podda, PhD won the 2022 edition of MAKEtoCARE award with DIGICOG-MS®. This contest promoted by Sanofi, that seeks and rewards innovative, practical solutions that are useful for responding to the real needs of people with any form of disability, serious and/or chronic diseases, or victims of traumatic events, which result in a temporary or permanent change in the quality of life [https://www.maketocare.it/]. This work has been funded as Progetto Ricerca Speciale by Italian Multiple Sclerosis Foundation (2021/Special/005).

Data Availability

All data produced in this study are available upon reasonable request to the corresponding author.

Authors' Contributions

JP conceived the study, designed DIGICOG-MS®, interpreted results, created figures, and drafted the manuscript; AT contributed to the definition of experimental hypotheses and revised the manuscript; MP and FDA performed the data analysis, created tables and contributed to the initial draft of the manuscript; AS performed data collection; LP revised the manuscript. All authors read and approved the final manuscript. The manuscript was written entirely by humans. All authors are responsible and accountable for the originality, accuracy, and integrity of the work.

Conflicts of interest

All authors declare that there is no conflict of interest.

Reference

1. Giovannoni G, Butzkueven H, Dhib-Jalbut S *et al.* Brain health: time matters in multiple sclerosis. *Mult Scler Relat Disord* 2016;**9**:S5–48.

2. Thompson AJ, Baranzini SE, Geurts J *et al.* Multiple sclerosis. *Lancet* 2018;**391**:1622–36.

3. Ehde DM, Gibbons LE, Chwastiak L *et al.* Chronic pain in a large community sample of persons with multiple sclerosis. *Mult Scler J* 2003;**9**:605–11.

4. Nortvedt MW, Riise T, Frugaård J *et al.* Prevalence of bladder, bowel and sexual problems among

multiple sclerosis patients two to five years after diagnosis. *Mult Scler J* 2007;**13**:106–12.

5. Sosnoff JJ, Socie MJ, Boes MK *et al.* Mobility , Balance and Falls in Persons with Multiple Sclerosis. 2011;**6**:2–6.

6. Amato MP, Prestipino E, Bellinvia A. Identifying risk factors for cognitive issues in Multiple Sclerosis. *Expert Rev Neurother* 2019:333–47.

7. DeLuca J, Chiaravalloti ND, Sandroff BM. Treatment and management of cognitive dysfunction in patients with multiple sclerosis. *Nat Rev Neurol* 2020;**16**:319–32.

8. Chiaravalloti ND, Genova HM, Deluca J. Cognitive rehabilitation in multiple sclerosis : the role of plasticity. *Front Neurol* 2015;**6**:1–10.

9. Podda J, Tacchino A, Pedullà L *et al.* Focus on neglected features of cognitive rehabilitation in MS: Setting and mode of the treatment. *Mult Scler J* 2020, DOI: 10.1177/1352458520966300.

10. Mitolo M, Venneri A, Wilkinson ID *et al.* Cognitive rehabilitation in multiple sclerosis: A systematic review. *J Neurol Sci* 2015;**354**:1–9.

11. 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**:1–6.

12. Benedict RHB, Amato MP, Boringa J *et al.* Brief International Cognitive Assessment for MS (BICAMS): international standards for validation. *BMC Neurol* 2012;**12**, DOI: 10.1186/1471-2377-12-55.

13. Rao SM. A manual for the brief repeatable battery of neuropsychological tests in multiple sclerosis. *Milwaukee Med Coll Wisconsin* 1990;**1696**.

14. Benedict RHB, Fischer JS, Archibald CJ *et al.* Minimal neuropsychological assessment of MS patients: A consensus approach. *Clin Neuropsychol* 2002;**16**:381–97.

15. Pless S, Woelfle T, Naegelin Y *et al.* Assessment of cognitive performance in multiple sclerosis using smartphone-based training games: a feasibility study. *J Neurol* 2023;**270**:3451–63.

16. Lavorgna L, Brigo F, Moccia M *et al.* e-Health and multiple sclerosis: An update. *Mult Scler J* 2018;**24**:1657–64.

17. Salimzadeh Z, Damanabi S, Kalankesh LR *et al.* Mobile applications for multiple sclerosis: A focus on self-management. *Acta Inform Medica* 2019;**27**:12–8.

18. Beier M, Alschuler K, Amtmann D *et al.* iCAMS: Assessing the reliability of a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS) tablet application. *Int J MS Care* 2020;**22**:67–74.

19. 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* 2020;**38**:101479.

20. Montalban X, Graves J, Midaglia L *et al.* A smartphone sensor-based digital outcome assessment of multiple sclerosis. *Mult Scler J* 2021:654–64.

21. Hochstrasser C, Rieder S, Jufer-Riedi U *et al.* Computerized Symbol Digit Modalities Test in a Swiss Pediatric Cohort Part 1: Validation. *Front Psychol* 2021;**12**:1–10.

22. Costabile T, Signoriello E, Lauro F *et al.* Validation of an iPad version of the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). *Mult Scler Relat Disord* 2023;**74**:1–11.

23. Hsu WY, Rowles W, Anguera JA *et al.* Assessing Cognitive Function in Multiple Sclerosis With Digital Tools: Observational Study. *J Med Internet Res* 2021;**23**, DOI: 10.2196/25748.

24. Gromisch ES, Turner AP, Haselkorn JK *et al.* Mobile health (mHealth) usage, barriers, and technological considerations in persons with multiple sclerosis: A literature review. *JAMIA Open* 2021;**4**:1–10.

25. Bennett A V., Dueck AC, Mitchell SA *et al.* Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO. *Health Qual Life Outcomes* 2016;**14**:1–12.

26. Polman CH, Reingold SC, Banwell B *et al.* Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011;**69**:292–302.

27. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;**33**:1444.

28. Honarmand K, Feinstein A. Validation of the Hospital Anxiety and Depression Scale for use with multiple sclerosis patients. *Mult Scler* 2009;**15**:1518–24.

29. Tacchino A, Ponzio M, Pedullà L *et al.* Italian validation of the Arm Function in Multiple Sclerosis Questionnaire (AMSQ). *Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol* 2020;**41**:3273–81.

30. Hobart J, Lamping D, Fitzpatrick R *et al.* The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain* 2001;**124**:962–73.

31. Brooke J. Sus: a “quick and dirty’usability. *Usability Eval Ind* 1996;**189**.

32. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Intl J Human–Computer Interact* 2008;**24**:574–94.

33. Zhou L, Bao J, Setiawan IMA *et al.* The mhealth app usability questionnaire (MAUQ): Development and validation study. *JMIR mHealth uHealth* 2019;**7**:1–15.

34. Frechette M, Fanning J, Hsieh K *et al.* The Usability of a Smartphone-Based Fall Risk Assessment App for Adult Wheelchair Users: Observational Study. *JMIR Form Res* 2022;**6**:1–13.

35. Goretti B, Patti F, Cilia S *et al.* The Rao ’ s Brief Repeatable Battery version B : normative values with age , education and gender corrections in an Italian population. 2013, DOI: 10.1007/s10072-013-1558-7.

36. Amato MP, Portaccio E, Goretti B *et al.* The Rao’s Brief Repeatable Battery and Stroop test: normative values with age, education and gender corrections in an Italian population. *Mult Scler J* 2006;**12**:787–93.

37. Carlesimo GA, Caltagirone C, Gainotti G. The Mental Deterioration Battery: normative data,

diagnostic reliability and qualitative analyses of cognitive impairment. The Group for the Standardization of the Mental Deterioration Battery. *Eur Neurol* 1996;**36**:378–84.

38. Cohen J. Set correlation and contingency tables. *Appl Psychol Meas* 1988;**12**:425–34.

39. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Arch Orofac Sci* 2017;**12**:1–11.

40. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;**15**:155–63.

41. World Medical Association General Assembly. Declaration of Helsinki. Ethical principles for medical research involving human subjects. *World Med J* 2008;**54**.

42. 5 Ways to Interpret a SUS Score. MeasuringU.

43. van Oirschot P, Heerings M, Wendrich K *et al.* Symbol Digit Modalities Test Variant in a Smartphone App for Persons With Multiple Sclerosis: Validation Study. *JMIR mHealth uHealth* 2020;**8**:e18160.

44. Kalb R, Beier M, Benedict RHB *et al.* Recommendations for cognitive screening and management in multiple sclerosis care. *Mult Scler J* 2018;**24**:1665–80.

45. Lee M, Mahmood ABS Bin, Lee ES *et al.* Smartphone and Mobile App Use Among Physicians in Clinical Practice: Scoping Review. *JMIR mHealth uHealth* 2023;**11**, DOI: 10.2196/44765.

46. Ly KH, Janni E, Wrede R *et al.* Experiences of a guided smartphone-based behavioral activation therapy for depression: A qualitative study. *Internet Interv* 2015;**2**:60–8.

47. Thielsch MT, Thielsch C. Depressive symptoms and web user experience. *PeerJ* 2018;**6**:e4439.

48. Torous J, Nicholas J, Larsen ME *et al.* Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. *Evid Based Ment Health* 2018;**21**:116–9.

49. Sarkar U, Gourley GI, Lyles CR *et al.* Usability of Commercially Available Mobile Applications for Diverse Patients. *J Gen Intern Med* 2016;**31**:1417–26.

50. Kampmeijer R, Pavlova M, Tambor M *et al.* The use of e-health and m-health tools in health

promotion and primary prevention among older adults: A systematic literature review. *BMC Health Serv Res* 2016;**16**, DOI: 10.1186/s12913-016-1522-3.

51. Marrie RA, Leung S, Tyry T *et al.* Use of eHealth and mHealth technology by persons with multiple sclerosis. *Mult Scler Relat Disord* 2019;**27**:13–9.

52. Pham L, Harris T, Varosanec M *et al.* Smartphone-based symbol-digit modalities test reliably captures brain damage in multiple sclerosis. *npj Digit Med* 2021;**4**:1–13.

53. Benedict RHB, DeLuca J, Phillips G *et al.* Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler J* 2017;**23**:721–33.

54. Leavitt VM. The SDMT is not information processing speed. *Mult Scler J* 2021;**27**:1806–7.

55. Brichetto G, Pedullà L, Podda J *et al.* Beyond center-based testing: Understanding and improving functioning with wearable technology in MS. *Mult Scler J* 2019;**25**, DOI: 10.1177/1352458519857075.

56. Tacchino A, Brichetto G, Ponzio M *et al.* Multiple sclerosis and rehabilitation : an overview of the different rehabilitation settings. *Neurol Sci* 2017;**38**:2131–8.

57. Marziniak M, Brichetto G, Feys P *et al.* The use of digital and remote communication technologies as a tool for multiple sclerosis management: Narrative review. *J Med Internet Res* 2018;**20**:1–21.

58. Lam KH, 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**:300–8.

59. Settle JR, Robinson SA, Kane R *et al.* Remote cognitive assessments for patients with multiple sclerosis: A feasibility study. *Mult Scler* 2015;**21**:1072–9.

60. Tacchino A, Podda J, Bergamaschi V *et al.* Cognitive rehabilitation in multiple sclerosis: Three digital ingredients to address current and future priorities. *Front Hum Neurosci* 2023;**17**:1–7.

61. Brichetto G, Monti Bragadin M, Fiorini S *et al.* The hidden information in patient-reported outcomes and clinician-assessed outcomes: multiple sclerosis as a proof of concept of a machine

learning approach. *Neurol Sci* 2020;**41**:459–62.

62. Ziccardi S, Fuchs T, Dwyer MG *et al.* Cognitive phenotypes predict response to restorative cognitive rehabilitation in multiple sclerosis. :1–5.

63. Podda J, Ponzio M, Messmer Uccelli M *et al.* Predictors of clinically significant anxiety in people with multiple sclerosis: A one-year follow-up study. *Mult Scler Relat Disord* 2020;**45**:102417.

64. Hancock LM, Galioto R, Samsonov A *et al.* A proposed new taxonomy of cognitive phenotypes in multiple sclerosis: The International Classification of Cognitive Disorders in MS (IC-CoDiMS). *Mult Scler J* 2023;**29**:615–27.

65. Voigt I, Inojosa H, Dillenseger A *et al.* Digital Twins for Multiple Sclerosis. *Front Immunol* 2021;**12**:1–17.

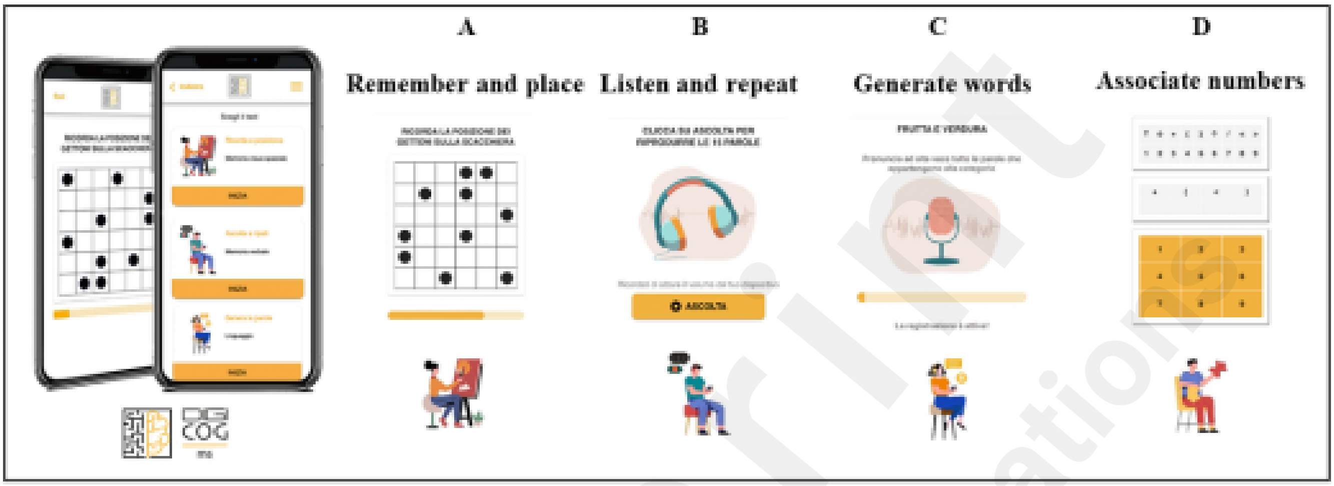
Abbreviation

AMSQ: Arm Function in Multiple Sclerosis
CI: cognitive impairment
DIGICOG-MS®: DIGItal assessment of COGnitive impairment in Multiple Sclerosis
EDSS: Expanded Disability Status Scale
HADS-a: Hospital Anxiety and Depression Scale - anxiety subscale
HADS-d: Hospital Anxiety and Depression Scale - depression subscale
ICC: Intraclass Correlation Coefficient
IPS: Information Processing Speed
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
MS: Multiple Sclerosis
MSIS29: Multiple Sclerosis Impact Scale
PPMS: Primary Progressive Multiple Sclerosis
PwMS: People with Multiple Sclerosis
RAVLT: Rey Verbal Learning Test
RRMS: Relapsing Remitting Multiple Sclerosis
SDMT: Symbol Digit Modalities Test
SF: Semantic Fluency
SPART: 10/36 Spatial Recall Test
SPMS: Secondary Progressive Multiple Sclerosis
SUS: System Usability Scale
VM: verbal memory
VSM: visuospatial memory
WLG: Word List Generation

Supplementary Files

Figures

Visualization of DIGICOG-MS®. The four digital tests implemented in DIGICOG-MS® that assess visuospatial memory (A), verbal memory (B), semantic fluency (C), and information processing speed (D).



Scatter and Bland–Altman plots of digital and traditional assessment tests for visuospatial memory (VSM), verbal memory (VM), semantic fluency (SF) and information processing speed (IPS). Correlations between digital (DIG) on y-axis and traditional (TRAD) on x-axis are shown (first column). In the Bland-Altman plots (second column), agreement between DIG and TRAD is shown. The dashed line represents the mean difference, that is expected to align with y = 0 for an optimal level of agreement; the solid lines represent the 95% limits of agreement, within which all points are expected to fall. Following digital and traditional tests were used to measure cognitive domains frequently altered in MS: Remember and place and 10/36 Spatial Recall Test (VSM); Listen and repeat and Rey Verbal Learning Test (VM); Generate words and Word List Generation (SF); Associate numbers and Symbol Digit Modalities Test (IPS).