

SCIENTIFIC REPORTS



OPEN

An ecological measure to screen executive functioning in MS: the Picture Interpretation Test (PIT) 360°

Olivia Realdon¹, Silvia Serino^{2,3}, Federica Savazzi⁴, Federica Rossetto⁴, Pietro Cipresso^{2,3}, Thomas D. Parsons⁵, Giacomo Cappellini⁶, Fabrizia Mantovani¹, Laura Mendozzi⁷, Raffaello Nemni^{8,9}, Giuseppe Riva^{2,3} & Francesca Baglio⁴

Executive functions are crucial for performance of everyday activities. In Multiple Sclerosis (MS), executive dysfunctions can be apparent from the early onset of the disease. Technology-based time-efficient and resource-saving tools for early evaluation of executive functions using an ecological approach are needed to assess functional performance in real-life. The aim was to compare the efficiency of the Picture Interpretation Test 360° (PIT 360°) with traditional measures on executive dysfunction in Persons with Multiple Sclerosis (PwMS) and Healthy Controls (HC). Participants were 31 patients with Relapsing-Remitting MS (mean age = 44.323 ± 13.149; mean Expanded Disability Status Scale = 2) and 39 HC (mean age = 39.538 ± 15.728). All were tested with standard neuropsychological tests of executive functions, PIT 360°, and measures of user experience. While standard neuropsychological tests failed to differentiate between PwMS and HC group, the PIT 360° was successful in detecting executive dysfunction in PwMS. All participants reported the PIT 360° to be an engaging tool and endorsed positive reactions to their experience. Overall, the PIT 360° is a quick, sensitive, and ecological tool that captures real-world executive dysfunction in PwMS. This engaging measure is sensitive for the detection of executive deficits since the early phases of the disease.

Cognitive impairment in Multiple Sclerosis (MS) includes, among other deficits, executive dysfunction, multi-tasking difficulties, verbal fluency declines, and visuo-spatial deficits^{1–3}. Cognitive impairment has been found in all disease subtypes⁴, including one-third of patients with early Relapsing-Remitting (RR) MS⁵. Early onset of difficulties in simultaneous management of everyday activities is often reported in Persons with Multiple Sclerosis (PwMS)⁶. Altogether, these impairments have a disruptive impact on quality of life and the ability of PwMS to actively adapt to the changing demands of the physical and social environment^{7,8}. Although conventional neuropsychological tests exist for assessing cognitive dysfunction in PwMS, they tend to be limited in their capacity for capturing the sorts of executive functioning deficits that are critical for functioning in real-world contexts⁶.

Evaluating functional performance across a range of real-life situations is the core of the function-led approach to the assessment of executive functions⁹. As highlighted by Chan and colleagues¹⁰, this approach, rather than fractionating the executive dysfunction, aims to incorporate the complexity of real-life challenges into tasks able to tap a number of executive domains simultaneously. Using this approach, executive functions can be captured in

¹Università degli Studi di Milano-Bicocca, Department of Human Sciences for Education, Piazza Ateneo Nuovo 1, 20126, Milan, Italy. ²Università Cattolica del Sacro Cuore, Department of Psychology, Largo Gemelli 1, 20123, Milan, Italy. ³IRCCS Istituto Auxologico Italiano, Applied Technology for Neuro-Psychology LAB, via Magnasco 2, 20149, Milan, Italy. ⁴IRCCS, Fondazione Don Carlo Gnocchi, Neurorehabilitation Unit and Imaging in Rehabilitation LAB, Via Capecelatro, 66, 20148, Milan, Italy. ⁵University of North Texas, Computational Neuropsychology and Simulation Laboratory, 1155 Union Circle #311280, Denton, Texas, 76203-5017, USA. ⁶National Research Council of Italy, Institute for the Dynamics of Environmental Processes, Piazza della Scienza, 1, 20126, Milan, Italy. ⁷IRCCS, Fondazione Don Carlo Gnocchi, Multiple Sclerosis Unit, Via Capecelatro, 66, 20148, Milan, Italy. ⁸IRCCS, Fondazione Don Carlo Gnocchi, Neurorehabilitation Unit, Via Capecelatro, 66, 20148, Milan, Italy. ⁹Università degli Studi di Milano, Department of Pathophysiology and Transplantation, Via Francesco Sforza 35, 20122, Milan, Italy. Correspondence and requests for materials should be addressed to F.B. (email: fbaglio@dongnocchi.it)

	PwMS group [N = 31]	HC group [N = 39]	Group comparison ^a p_value
Age(years) [mean ± SD]	44.323 ± 13.149	39.538 ± 15.728	0.170
Education (years) [mean ± SD]	15.161 ± 2.922	16.385 ± 2.551	0.066
Gender (males:females)	15:16	15:24	0.405
EDSS [mean ± SD]	2.350 ± 1.750	—	—
Disease Duration (years) [mean ± SD]	12.387 ± 9.583	—	—
MoCA [mean ± SD]	24.902 ± 1.971	26.145 ± 1.497	0.005 ^b
TMT-A [mean ± SD]	32.871 ± 12.099	36.154 ± 13.698	0.298
TMT-B [mean ± SD]	98.032 ± 36.892	100.128 ± 43.808	0.832
TMT-BA [mean ± SD]	64.258 ± 27.922	63.718 ± 33.114	0.942
Verbal Fluency (FAS form) [mean ± SD]	35.229 ± 10.479	38.259 ± 10.361	0.231
PIT 360°, time [mean ± SD]	65.249 ± 55.307	33.048 ± 19.565	0.024
PIT 360°, number [mean ± SD]	8.355 ± 7.153	3.872 ± 3.299	0.063

Table 1. Sociodemographics, neuropsychological assessment and PIT 360° scores for PwMS and HC groups. Abbreviations: PwMS, People with Multiple Sclerosis; HC, Healthy Control; EDSS, Expanded Disability Status Scale; MoCA, Montreal Cognitive Assessment; TMT, Trail Making Test; FAS, administration letters to test Verbal Fluency; PIT, Picture Interpretation Test. ^aSociodemographic and neuropsychological variables were compared at baseline using t test or χ^2 as appropriate. ^bBecause the Levene's test was significant, the t-test reported assumes that variances were not equal and $df = 54.730$.

a context-specific scenario that enhances prediction of everyday functional behaviours. Along this line, Rouaud and colleagues¹¹ showed that ecological tests could detect executive dysfunction in PwMS that was underestimated by conventional neuropsychological assessments. In an update on strategies for assessing cognition in PwMS, Ruet and Brochet¹² pointed out that neuropsychological and ecological tests are indeed complementary tools for assessing cognitive dysfunction in everyday-like conditions.

Several tools for assessing executive functioning using an ecological approach have been developed through the application of novel technologies like Virtual Reality (VR). VR platforms allow for the development of ecologically valid assessment that simulates everyday activities in secure scenarios¹³. Virtual environments for the assessment of cognitive impairments have already been developed and empirically validated with regard to several clinical conditions^{14–16}, including MS, as in the Urban DailyCog task¹⁷.

It is widely recognized that conventional paper and pencil tests for the assessment of cognitive status in MS are time- and resource-consuming, limiting their incorporation in standard MS care¹⁸. There is a wide consensus on the need for short and well validated tools that can be seamlessly incorporated in everyday clinical practice¹⁸. Sumowski and colleagues⁶ therefore advocated the development and validation of technology-enhanced brief and resource-efficient tools – be they computer- or tablet-based – as a key priority in measuring cognitive impairments in this population.

With the PIT 360°¹⁹ we advance a quick and ecological tool for the evaluation of dysexecutive deficits. The PIT 360° - which is a 360° version of the Picture Interpretation Test^{20,21} based on Luria and colleagues' work²² - proved to be effective in the screening of executive dysfunction in Parkinson's Disease (PD) in comparison to healthy controls.

In the present study we apply the PIT 360° to the evaluation of executive dysfunction in PwMS. We predicted that PIT 360° would be able to capture real-world executive dysfunction, in PwMS, in a quick and more sensitive way than conventional assessment tests.

Results

Participants' characteristics and conventional neuropsychological assessment of executive functions. Table 1 exhibits results from the comparison between groups (PwMS vs Healthy Control, HC) on baseline characteristics and neuropsychological assessment scores. No significant difference was detected between the two groups for gender [$\chi^2 = 0.695$; $df = 1$, $p = 0.405$], age [$t(67,804) = 1.386$, $p = 0.170$], or education [$t(68) = -1.869$, $p = 0.066$]. Findings obtained from Independent Student's t-tests indicated no significant differences between PwMS and HC group with respect to the three sub-tests of TMT [TMT-A, $t(68) = -1.048$; TMT-B, $t(68) = -0.213$; TMT-BA, $t(68) = 0.073$] or Verbal Fluency [$t(68) = -1.209$]. When compared with HC, PwMS obtained a significantly lower MoCA score [$t(54,730) = -2.906$].

Performance on PIT 360°. Results in both indices obtained from the PIT 360° (see Table 1) revealed greater performance deficits for PwMS group when compared to the HC group. The PwMS group took more time (mean = 65.249; SD = 55.307) when compared to the HC group (mean = 33.048; SD = 19.565) for interpreting the PIT 360° scene [$F(1,66) = 5.329$; $p = 0.024$; Partial $\eta^2 = 0.075$]. Moreover, PwMS reported a higher number of scene elements in comparison to the HC group [MS, mean = 8.355; SD = 7.153; HC, mean = 3.872, SD = 3.299; $F(1,34) = 4.449$; $p = 0.039$; Partial $\eta^2 = 0.063$].

User experience assessment. Findings from comparison of the number of self-reported felt emotions and their intensities (in each of the four Geneva Emotion Wheel - GEW - quadrants) between the two groups

Scale	Subscale	PwMS Group [M (SD)]	HC Group [M (SD)]	Z	p	N ^a
GEW	GEW 1 - Positive Valence/High coping potential					
	Number of emotions	2.806 (1.1667)	2.842 (1.326)	-0.137	0.891	68
	Intensity	3.433 (0.685)	3.373 (1.092)	-0.043	0.965	68
	GEW 2 - Positive Valence/Low coping potential					
	Number of emotions	1.000 (0.856)	1.184 (1.159)	-0.470	0.639	68
	Intensity	2.561 (0.972)	2.856 (0.960)	-0.867	0.386	47
	GEW 3 - Negative Valence/Low coping potential					
	Number of emotions	0.097 (0.396)	0.263 (0.601)	-1.444	0.149	68
	Intensity	1.500 (0.707)	1.357 (0.475)	-0.327	0.743	10
	GEW 4 - Negative Valence/High coping potential					
	Number of emotions	0.064 (0.359)	0.526 (1.109)	-2.375	0.018	68
	Intensity	3.00	2.713 (0.985)	/	/	
FSS	Perceived coping skills	3.548 (0.850)	3.769 (0.667)	-1.291	0.197	70
	Perceived challenge	2.548 (0.624)	2.103 (0.598)	-2.873	0.004	70
	Perceived challenge-skill balance	2.323 (0.600)	2.179 (0.644)	-0.914	0.361	70
IMI		3.006 (0.881)	3.369 (0.810)	-1.790	0.078	70
SUS Questionnaire		3.667 (1.510)	3.863 (1.675)	-0.509	0.612	70

Table 2. Scores obtained from the user experience assessment for PwMS and HC. Data are shown as Means (M) and Standard Deviations (SD). Abbreviations: PwMS group, People with Multiple Sclerosis; HC group, Healthy Control; M, Mean; SD, Standard Deviation; GEW, Geneva Emotion Wheel; FSS, Flow Short Scale; IMI, Intrinsic Motivation Inventory; SUS, Slater-Usch-Steed. ^aOne participant in the “HC group” was excluded from the analyses of GEW data because of problems in data collection.

			U	p
GEW 1	3.400 (0.923)	vs. GEW2	-6.711	<0.001
		vs. GEW3	-7.181	<0.001
		vs. GEW4	-6.934	<0.001
GEW 2	2.718 (0.967)	vs. GEW3	-5.89	<0.001
		vs. GEW4	-4.793	<0.001
GEW 3		vs. GEW4	-1.317	0.241
	1.389 (0.486)			
GEW 4				
	2.742 (0.989)			

Table 3. Results obtained from Wilcoxon Test comparisons of different quadrants in the Geneva Emotion Wheel (GEW) - Mean number of self-reported emotions (SD).

(PwMS vs. HC), revealed a significant difference in the number of emotions with negative valence and high coping potential. In this specific quadrant, in comparison to PwMS group, the HC group reported a significantly higher number of self-reported emotions.

No significant differences were observed between groups with respect to the overall number of self-reported felt emotions and their intensities (Table 2).

However, findings obtained from the Friedman Test revealed a significant difference among the four quadrants of GEW in terms of the number of self-reported emotions [$\chi^2(3) = 155.285; p < 0.001$].

Wilcoxon tests (with Bonferroni adjustment) indicated that all participants experienced a higher number of emotions with positive valence and high coping potential (Table 3).

Assessment using the Flow Short Scale (FSS) revealed that the PwMS group perceived a higher level of challenges when confronted with the proposed activity in comparison to HC. However, no significant between-group difference emerged with respect to the perceived level of skills and challenge-skills balance (Table 2).

Both groups endorsed a high appreciation for the activity (Intrinsic Motivation Inventory - IMI) and an intense sense of presence (Slater-Usch-Steed - SUS - Questionnaire) (Table 2).

Classification of PwMS or HC. Tables 4 and 5 show the classification results for discriminating between the HC and the PwMS groups. Naïve Bayes and Support Vector Machine algorithms emerged as the best algorithms for classifying HC and PwMS in their respective groups. Using the scores from conventional executive functions tests as input, the machine learning algorithms showed a classification accuracy between 52.9% and 65.7%. In contrast, the indices from PIT 360° achieved a higher classification accuracy, ranging from 65.7% to 72.9%.

Method	AUC	CA	F1	Precision	Recall
Logistic Regression	0.494	0.529	0.653	0.554	0.795
Random Forest	0.627	0.643	0.706	0.652	0.769
Support Vector Machine (SVM)	0.636	0.657	0.727	0.653	0.821
Naïve Bayes	0.617	0.629	0.683	0.651	0.718

Table 4. Stratified 10-fold Cross validation for the neuropsychological assessment battery. ^aAUC (Area under the ROC curve) is the area under the classic receiver-operating curve. CA (Classification accuracy) represents the proportion of the examples that were classified correctly; F1 represents the weighted harmonic average of the precision and recall (defined below); Precision represents a proportion of true positives among all the instances classified as positive. In our case, the proportion of a condition was identified correctly; Recall represents the proportion of true positives among the positive instances in our data.

Method	AUC	CA	F1	Precision	Recall
Logistic Regression	0.639	0.657	0.721	0.660	0.795
Random Forest	0.707	0.729	0.787	0.700	0.897
Support Vector Machine (SVM)	0.668	0.700	0.779	0.661	0.949
Naïve Bayes	0.678	0.700	0.764	0.680	0.872

Table 5. Stratified 10-fold Cross validation for the indexes of PIT360°. ^aAUC (Area under the ROC curve) is the area under the classic receiver-operating curve. CA (Classification accuracy) represents the proportion of the examples that were classified correctly; F1 represents the weighted harmonic average of the precision and recall (defined below); Precision represents a proportion of true positives among all the instances classified as positive. In our case, the proportion of a condition was identified correctly; Recall represents the proportion of true positives among the positive instances in our data.

Figure 1 shows the confusion matrix of all classifiers used for classifying individuals into PwMS Group and HC Group. Results revealed that indices from the PIT 360° had a higher capability for correctly classifying PwMS in their group.

Discussion

We aimed to evaluate the efficacy of a 360° version of the PIT for detecting executive dysfunction in PwMS through a function-led approach that combined experimental control with a real-world engaging background. In line with research findings on MS that reveal cognitive impairments that can be characterized as executive dysfunction⁶, the PwMS performed significantly worse on the PIT 360° than did persons in the HC group.

While the mean global cognitive level of PwMS in our study was lower than that of HC, it was still in a non-pathological range. This suggests initial subclinical global dysfunction in PwMS. This initial dysfunction detected with a renowned test sensitive to MS-related cognitive impairment is in line with the prior work detecting cognitive impairment in PwMS²³. It is important to note that although verbal fluency is a sensitive tool for assessing executive impairment in PwMS and is part of the minimal assessment of cognitive function in MS (MACFIMS battery)²⁴, the assessment with this test and the TMT failed in showing differences in executive functions between groups.

Different from standard neuropsychological tests used, the PIT 360° differentiated successfully between the pathological and the control conditions both in terms of time to give an answer and in number of elements in the scene. This result showed that PIT 360° is an ecological tool that is highly sensitive to MS pathology—even in its initial phases (EDSS, range 1–3). These findings were also confirmed by the higher accuracy in the Random Forest classification of participants to the clinical or non-clinical conditions (when using indices from PIT 360°) with respect to those from neuropsychological assessment. These robust findings demonstrated the efficacy of PIT 360° for detecting impairment of executive functions at an early clinical stage of MS. Moreover, they suggest that this ecological tool can be used for prompt diagnosis and early enrollment of PwMS in targeted rehabilitation⁴. The importance of an early management of cognitive impairment in MS is highlighted by the fact that it can pre-date the onset of physical disability and slow cognitive decline^{7,25}.

Although the findings of the present study are promising, in the comparison to standard neuropsychological assessment, PIT 360° is only a very sensitive screening tool not covering the need for a full and analytical examination of executive functions. In addition, it is a technology-based test implying the use of a VR headset with potential side-effects (e.g. nausea) in some patients.

Considering findings related to users' experiences, we found that PIT 360° was considered to be an engaging tool both by the HC and the PwMS groups. Firstly, all participants reported a good sense of presence in the 360° scene (SUS Questionnaire) showing that they actively experienced the task in a context perceived as a real-life place. Both groups rated the challenge of the PIT 360° task as feasible, in the sense that it was considered balanced with respect to their skills (FSS scale). Furthermore, PwMS and HC positively assessed participant appreciation for and interest in performing PIT 360° task (IMI scale).

All participants endorsed positive reactions to the task, showing that their experience of the PIT 360° was highly pleasant and under control. This was apparent in the higher number of self reported emotions in the first quadrant of the GEW, which includes interest, joy, happiness, satisfaction, elation and pride. Interestingly, HC

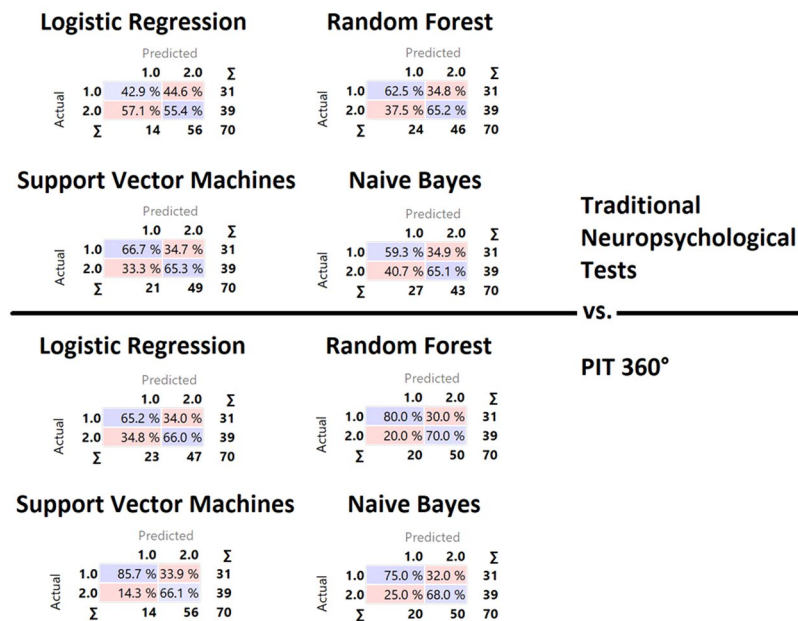


Figure 1. Confusion matrix. The confusion matrix of all classifiers computed for the classification into “HC Group” and “PwMS Group”. The values on the diagonal (i.e., purple values) indicate the elements for which the predicted group is equal to the true group, while off-diagonal values are those that are mislabelled by the classifiers. Results showed that PIT 360° had a higher capability for correctly classifying PwMS in their group with respect to traditional neuropsychological tests of executive functioning.

vs. PwMS reported a higher number of self-reported emotions with high coping potential and negative valence. This finding can be related to the higher level of challenges perceived by the PwMS group when faced with the proposed activity in comparison to HC (FSS scale). A possible interpretation of these results is that PwMS exerted greater attentional effort when attempting to complete the task which, most probably, was higher than that of HC.

Interestingly, the efficiency of PIT 360° in detecting executive dysfunctioning was observed also in PD¹⁹ but was lower compared to SM. This is not surprising because executive function disorders are defined in functional terms and not as a topographic syndrome²⁶ and are assessed by functional-led approach in PIT 360°. The difference in the accuracy of classification in the two clinical conditions with respect to HC may be due to several reasons. The aging can represent a factor: people with PD were older than PwMS for the natural history of the disease. Moreover, it is well known that aging is associated with decline in executive function and in the two studies were included different HC groups due to the age-related demographics of the two neurological conditions. Then, it is reasonable to expect the different sensitivity in classification accuracy in a middle-aged vs. older adults sample. Furthermore, the overall brain profile vary in the two conditions involving fronto-subcortical degeneration in PD and white matter frontal pathway disconnection in MS. Therefore, the degree of severity of the brain damage can impact executive functioning differently in the two diseases.

Finally, PIT 360° offers a promise for answering the need for time-efficient and resource-saving tools that can screen PwMS for executive deficits. This reduces patient stress at the first evaluation and orients clinicians to perform subsequent clinical investigations using longer neuropsychological assessment batteries and the prompt inclusion in targeted rehabilitation programs.

Future studies should examine PIT 360° efficacy in detecting executive dysfunctioning with a larger cohort and with other clinical populations. Moreover, it will be important that the PIT 360° be investigated using neuroimaging to establish neural correlates. Additionally, it will be of major importance to proceed with the validation of PIT 360° parallel forms to make possible a short-term re-evaluation of executive functions.

In conclusion, the PIT 360° is a quick and ecological measure that demonstrated effective and sensitive screening of real-world deficits related to executive functioning in the early stages of MS. These findings support, within Parsons’ theoretical framework¹³ for the assessment of executive functions, the methodological note advanced by Sumowski and colleagues⁶ on the need of advancing effective, evidence-based, clinically feasible understanding and measurement of dysexecutive functioning.

Materials and Methods

Participants. Seventy participants took part in the study: thirty-one PwMS (51.6% female; mean age = 44.323 ± 13.149; mean years of education = 15.161 ± 2.922; PwMS group) and thirty-nine healthy controls (61.5% female; mean age = 39.538 ± 15.728; mean years of education = 16.385 ± 2.551; HC group).

Outpatients meeting the diagnostic criteria for clinically definite MS²⁷ with a RR disease course were consecutively recruited from the MS Unit of Don Carlo Gnocchi Foundation, IRCCS. All patients were at a mild stage of the disease, scoring between 1 and 3 of the Expanded Disability Status Scale (EDSS).

Exclusion criteria were as follows: less than 6 months from diagnosis, documented relapses within the last 3 months, severe psychiatric and neurological disorders other than MS.

The study was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013 and approved by the Local Ethics Committee (IRCCS Don Carlo Gnocchi Foundation). Written informed consent was obtained for all participants before study initiation.

Procedure of the study and measures. The study was carried out in three subsequent steps, as in Serino and colleagues²¹. After conventional neuropsychological assessment, we administered the PIT 360° session. Next, we evaluated participants' experiences relative to their subjective feelings, intrinsic motivations, balance between resources and demands while performing the task. Additionally, their sense of presence in the 360° environment was assessed.

Pre-task evaluation: neuropsychological measures. We used the same battery as in Serino et al.¹⁹: global cognitive level was assessed with the Montreal Cognitive Assessment (MoCA)²⁸ – which has been shown to be sensitive in identifying MS-related cognitive impairment²⁹; executive functioning was assessed using the Trail Making Test³⁰ (in two specific sub-tests: TMT-A and TMT-B) as a visuo-spatial examination with an index of time; PIT 360°; and a measure of phonemic verbal fluency, the controlled oral word association test (FAS form)²³.

PIT 360° session. The PIT 360°¹⁹ is the 360° version of the Picture Interpretation Test (PIT)^{20,21}. The PIT 360° environment consists of a scene in a contemporary real-world room with three frightened girls standing on chairs and a boy who is searching for something on the floor. Although not visible, it is apparent that there is a mouse (or some other small animal) hidden behind a piece of furniture. This scene is a present-day adaptation of the painting “Il Sorcio” (“The Mouse”, 1878, by Giacomo Favretto). Participants undergo a visual exploration task in which they are asked to interpret what is happening in a limited time frame. Time to correct interpretation of the scene (“There is a mouse/small animal”) and number of scene elements before correct interpretation are the outcome metrics. Session components and their unfolding over time are illustrated in Fig. 2.

Post-task evaluation: user experience assessment. After task completion, we evaluated a) self-reported subjective feelings through the Geneva Emotion Wheel (GEW)³¹. This tool provides a wheel shaped arrangement of 20 emotion words. Emotion labels are considered as indices “reflecting a unique experience of mental and bodily changes in the context of being confronted with a particular event”³². The wheel is displayed on a space formed by the underlying dimensions of valence (negative to positive) and control/coping potential (low to high). The orthogonal combination of these dimensions generates four quadrants: negative valence - low control; negative valence - high control; positive valence - low control; positive valence - high control. Subjective feelings about performing the task were rated through the mean number of emotion labels (range 0–5) and the respective reported intensity (range 1–5) within each quadrant; b) we also evaluated the skill-demands compatibility through the Flow Short Scale (FSS)^{33,34} assessing the perceived level of skills in coping with the task (“Perceived coping skills”), the perceived level of challenges of the task (“Perceived challenge”), and the perceived challenge-skill balance (“Perceived challenge- skill balance”) on a 5-points Likert scale; c) intrinsic motivation in performing the task was measured using the Interest/Enjoyment subscale of the Intrinsic Motivation Inventory (IMI; Deci)³⁵. The mean of the item scores (N = 5, 7-points Likert scale) was considered; d) finally, we measured the sense of presence experienced in the 360° environment through the Slater-Usuh-Steed Questionnaire (SUS Questionnaire)³⁶. The scale evaluated participants' sense of being present in the 360° scene, and the extent to which experiencing the scene using the PIT 360° became the dominant reality and recall as a place, through three items on 7-point scale.

Data analyses. First, the Kolmogorov-Smirnov test was used to check for the normality of data distribution for all the variables. Independent Student's t-tests and chi-square tests were used to compare group baseline characteristics. Then, independent Student's t-tests were carried out to explore between-group differences in the conventional assessment of executive functions (i.e., MoCA, TMT and phonemic fluency task). Two univariate analyses of variance with age and education as covariates (ANCOVA) were used to investigate PIT 360° differences in performances between HC group and PwMS group on the two performance indices (i.e., Correct Interpretation and Number of Scene Element). Since the distribution of these two variables differed moderately from normal, a square root transformation was tried. With this procedure, data were closer to the normal distribution as assessed with the Kolmogorov-Smirnov test.

Next, differences in conventional tests of executive functions between the two groups were evaluated using non-parametric tests (Wilcoxon tests). A univariate analysis of covariance (ANCOVA) with age and education as covariates was carried out to investigate differences between HC and PwMS groups in the indexes of PIT 360° (i.e., Correct Interpretation and Number of Scene Element). To investigate potential differences between the HC group and the PwMS group in user experience variables (i.e., GEW, FSS, IMI, and SUS Questionnaire), we performed independent Student's t-tests (for normal variables) and Mann-Whitney U tests (for not normal variables). As specifically concerns the number of self-reported emotions, the Friedman test was used to explore differences within the four quadrants of the GEW. A series of Bonferroni adjusted Wilcoxon tests were subsequently computed to explore significant effects. All these statistical analyses were conducted using the Statistical Package for the Social Sciences for Windows (SPSS Inc., Chicago, IL, USA), version 23.

To compare the classification accuracy of traditional tests of executive functions and indices from PIT 360°, nonlinear stochastic approximation (i.e., machine learning) methods were employed. In particular, a leave-one-out cross-validation was carried out with the following methods (as in our previous study)¹⁹: (a) a Logistic Regression classification algorithm with ridge regularization; (b) a Random Forest classification to classify features using an ensemble of decision trees; (c) a Support Vector Machine (SVM) to map inputs to

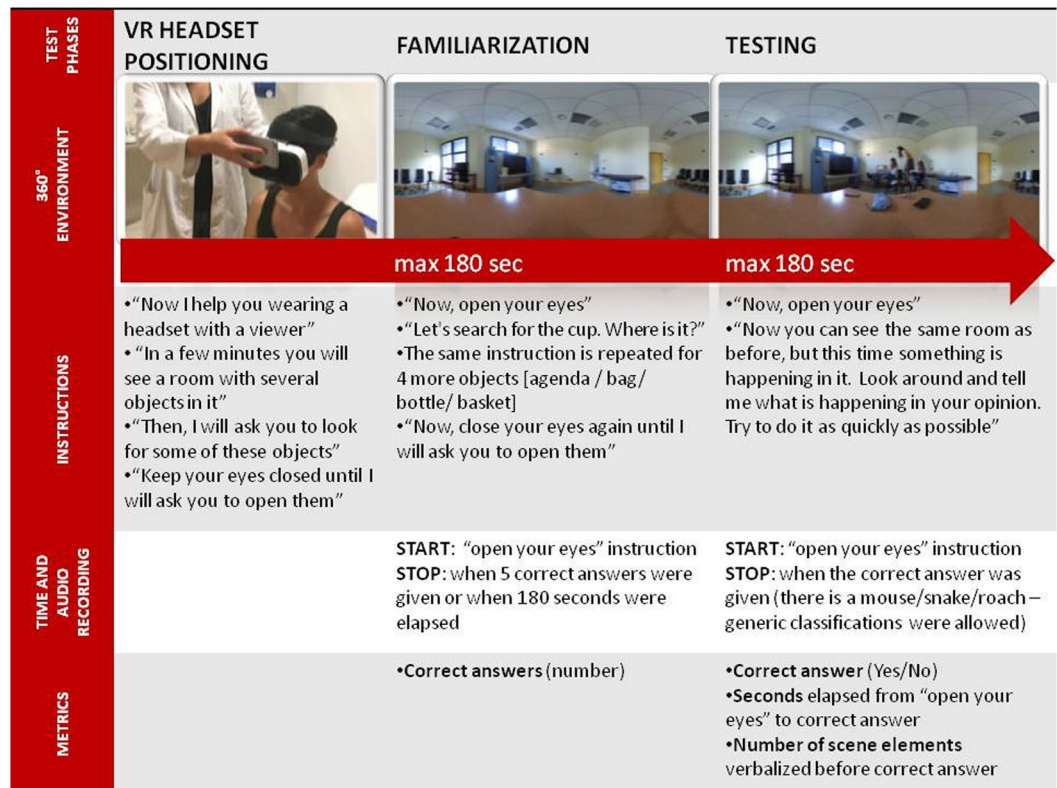


Figure 2. PIT 360° session. *In case of presbyopia, participants were asked to wear their own glasses. **The familiarization phase also allowed control for potential side effects (e.g., dizziness, nausea). The examiner followed a cessation rule in which experimental sessions should be stopped if severe side effects occurred. ***All persons depicted in the pictures were experimenters and signed an informed consent for publication of identifying images in an online open-access publication.

higher-dimensional feature spaces that best separate different features; (d) a naïve Bayes classification. All these analyses were computed using Python 3.4 with the Orange 3.3.5 data mining suite, which was available free in open source code (<https://github.com/biolab/orange3>).

Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- Benedict, R. H. *et al.* Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society* **12**, 549–558 (2006).
- Deloire, M. S. *et al.* MRI predictors of cognitive outcome in early multiple sclerosis. *Neurology* **76**, 1161–1167 (2011).
- Rao, S. M., Leo, G. J., Bernardin, L. & Unverzagt, F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology* **41**, 685–691 (1991).
- Amato, M. P. *et al.* Cognitive impairment in early stages of multiple sclerosis. *Neurological sciences* **31**, 211–214 (2010).
- Amato, M. P. *et al.* Relevance of cognitive deterioration in early relapsing-remitting MS: a 3-year follow-up study. *Multiple Sclerosis Journal* **16**, 1474–1482 (2010).
- Sumowski, J. F. *et al.* Cognition in multiple sclerosis: State of the field and priorities for the future. *Neurology* **90**, 278–288 (2018).
- Amato, M. P. *et al.* Treatment of cognitive impairment in multiple sclerosis: position paper. *J. Neurol.* **260**, 1452–1468 (2013).
- Mitolo, M., Venneri, A., Wilkinson, I. D. & Sharrack, B. Cognitive rehabilitation in multiple sclerosis: a systematic review. *J. Neurol. Sci.* **354**, 1–9 (2015).
- Parsons, T. D., Carlew, A. R., Magtoto, J. & Stonecipher, K. The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychological rehabilitation*, 1–31 (2015).
- Chan, R. C., Shum, D., Touloupoulou, T. & Chen, E. Y. Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of clinical neuropsychology* **23**, 201–216 (2008).
- Rouaud, O. *et al.* Contribution of ecological evaluation of executive disorders in multiple sclerosis. *Rev. Neurol. (Paris)* **162**, 964–969 (2006).
- Ruet, A. & Brochet, B. Cognitive assessment in patients with multiple sclerosis: From neuropsychological batteries to ecological tools. *Annals of physical and rehabilitation medicine* (2018).
- Parsons, T. D. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience* **9**, 660 (2015).
- Cipresso, P. *et al.* Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson’s disease. *Frontiers in behavioral neuroscience* **8**, 405 (2014).

15. Cipresso, P. *et al.* Break in volition: A virtual reality study in patients with obsessive-compulsive disorder. *Experimental brain research* **229**, 443–449 (2013).
16. Grewe, P. *et al.* Real-life memory and spatial navigation in patients with focal epilepsy: ecological validity of a virtual reality supermarket task. *Epilepsy & Behavior* **31**, 57–66 (2014).
17. Lamargue-Hamel, D. *et al.* Cognitive evaluation by tasks in a virtual reality environment in multiple sclerosis. *J. Neurol. Sci.* **359**, 94–99 (2015).
18. Langdon, D. *et al.* Recommendations for a brief international cognitive assessment for multiple sclerosis (BICAMS). *Multiple Sclerosis. Journal* **18**, 891–898 (2012).
19. Serino, S. *et al.* Picture Interpretation Test (PIT) 360: An Innovative Measure of Executive Functions. *Scientific reports* **7**, 16000 (2017).
20. Bisiach, E., Cappa, S. & Vallar, G. *Guida all'esame neuropsicologico* (R. Cortina, 1983).
21. Rosci, C., Sacco, D., Laiacona, M. & Capitani, E. Interpretation of a complex picture and its sensitivity to frontal damage: a reappraisal. *Neurological Sciences* **25**, 322–330 (2005).
22. Luria, A., Karpov, B. & Yarbuss, A. Disturbances of active visual perception with lesions of the frontal lobes. *Cortex* **2**, 202–212 (1966).
23. Carlesimo, G. A., 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.* **36**, 378–384 (1996).
24. Gromisch, E. S., Portnoy, J. G. & Foley, F. W. Comparison of the abbreviated minimal assessment of cognitive function in multiple sclerosis (aMACFIMS) and the brief international cognitive assessment for multiple sclerosis (BICAMS). *J. Neurol. Sci.* **388**, 70–75 (2018).
25. Nabavi, S. M. & Sangelaji, B. Cognitive dysfunction in multiple sclerosis: Usually forgotten in the clinical assessment of MS patients. *J. Res. Med. Sci.* **20**, 533–534 (2015).
26. Godefroy, O. *et al.* Dysexecutive disorders and their diagnosis: A position paper. *Cortex*. **109**, 322–335 (2018).
27. Polman, C. H. *et al.* Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann. Neurol.* **69**, 292–302 (2011).
28. Conti, S., Bonazzi, S., Laiacona, M., Masina, M. & Coralli, M. V. Montreal Cognitive Assessment (MoCA)–Italian version: regression based norms and equivalent scores. *Neurological Sciences* **36**, 209–214 (2015).
29. Freitas, S. *et al.* The Montreal Cognitive Assessment (MoCA) as a screening test for cognitive dysfunction in multiple sclerosis. *Applied Neuropsychology: Adult* **25**, 57–70 (2018).
30. Giovagnoli, A. R. *et al.* Trail making test: normative values from 287 normal adult controls. *The Italian journal of neurological sciences* **17**, 305–309 (1996).
31. Scherer, K. R., Shuman, V., Fontaine, J. R. J. & Soriano, C. The GRID meets the Wheel: Assessing emotional feeling via self-report. *Components of emotional meaning: A sourcebook*, 281–298 (2013).
32. Scherer, K. R. What are emotions? And how can they be measured? *Social science information* **44**, 695–729 (2005).
33. Rheinberg, F., Vollmeyer, R. & Engeser, S. In *Diagnostik von Motivation und Selbstkonzept* (eds Stiensmeier-Pelster, J. & Rheinberg, F.) 261–279, (2003).
34. Engeser, S. & Rheinberg, F. Flow, performance and moderators of challenge-skill balance. *Motiv. Emotion* **32**, 158–172 (2008).
35. Deci, E. L., Eghrari, H., Patrick, B. C. & Leone, D. R. Facilitating internalization: The self-determination theory perspective. *J. Pers.* **62**, 119–142 (1994).
36. Usho, M., Catena, E., Arman, S. & Slater, M. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments* **9**, 497–503 (2000).

Acknowledgements

The authors wish to thank all the participants for taking part in the study.

Author Contributions

O.R., S.S., F.S., G.R. and F.B. developed the study concept. All authors contributed to the study design. F.R. was in charge of the data collection. G.C. was responsible for the technical development of the PIT. O.R., S.S. and F.S. performed the data analysis and interpretation under the supervision of F.B., F.M., R.N., L.M., T.D.P. P.C. performed computational data analysis. F.S., F.B., O.R. and S.S. wrote the first draft of the manuscript. All authors were involved in a critical revision of the manuscript for important intellectual content. All the authors approved the final version of the manuscript for submission.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019