Multiple Sclerosis and Related Disorders 67 (2022) 104182

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

journal homepage: www.elsevier.com/locate/msard

Original article

Possible evidence of near transfer effects after adaptive working memory training in persons with multiple sclerosis

Sónia Félix Esbrí, Alba Sebastián Tirado, Carla Sanchis-Segura, Cristina Forn *

Departament de Psicología Bàsica, Clínica i Psicobiología, Universitat Jaume I, Universitat de Ciències de la Salut, Avda. Sos Baynat s/n (Edifici d'investigació II), Castelló de la Plana 12006, Spain

ARTICLE INFO

Keywords:
Multiple sclerosis
Working memory
Information processing
n-back
Transfer effects
cognitive training

ABSTRACT

Background: Cognitive deficits, especially in working memory (WM) and information processing (IP) efficiency, are common in people with multiple sclerosis (PwMS). Few studies have examined the efficacy of n-back training in improving these two cognitive functions in PwMS. In the present study, we examined the effects of an intensive n-back training program by measuring the gains on the trained task (2- and 3-back tasks), but we also studied possible near transfer effects to other tests that assess WM and IP, as well as far transfer effects or improvements in other cognitive functions.

Methods: A sample consisting of 35 PwMS with different cognitive statuses. All the participants underwent an adaptive n-back training for 10 days (60 min/day), and they were neuropsychologically assessed at baseline (D1) and after training (D10). The effectiveness of the training was tested: (1) by using mean-based comparisons and Cohen's d values; (2) by estimating and comparing the quartile values of the D1 and D10 distributions. Two indexes of improvement in individual performance were calculated, the net score improvement index (NSI) and the percent of maximum possible individualized improvement (PMPI).

Results: Repeat practice improves 2- and 3-back performance, showing more correct responses (CR) and lower reaction times (RT) on D10 compared to D1. These results were corroborated by the NSI and PMPI scores, but the gains after training were more statistically significant for the 3-back (observing higher CR and lower RT after training) than for the 2-back (observing gains in CR, but not in RT). We also observed a possible transference of this improvement on the n-back task to other WM/IPS tests. Specifically, statistically significant pre-post training differences were found in the values in three quartiles of the Paced Auditory Serial Addition Test (PASAT; q25, $p < 0.03$; q50, $p < 0.001$; q75, $p < 0.002$) and of the Symbol Digit Modalities Test (SDMT; q25, $p < 0.03$; q50, $p < 0.001$; q75, $p < 0.001$) as well as in two quartiles of the Letter-Number Sequencing Task (LNST; q50, $p < 0.004$; q75 $p < 0.001$), and in one quartile of the Digit Backwards Span Test (DSBT; q75, $p < 0.001$). Reliable change analyses confirmed these performance improvements on the PASAT, SDMT, and LNST.

Conclusions: This study confirmed that the intensive and adaptive n-back training produced improvements in the trained task in PwMS with different cognitive statuses. Furthermore, these gains were not only observed on the trained task, but they seemed to be also transferred to other tests that measured WM and IP functions.

1. Introduction

Cognitive impairment is common in multiple sclerosis (MS), and two of the most frequently affected functions are working memory (WM) and information processing (IP) efficiency. These two basic cognitive functions are mutually related and necessary in other higher-order intellectual operations. Accordingly, it has been proposed that initial WM and IP alterations may underlie the extensive cognitive deficits observed in advanced stages of the disease (Chiaravalloti and DeLuca, 2008), and that enhancing WM and IP could result in general cognitive improvement in people with multiple sclerosis (PwMS; Covey et al., 2018; Turtola and Covey, 2021).

One of the most widely used paradigms to improve WM is the n-back task, where participants are required to recall a sequence of items and determine whether the current item matches the item presented "n" positions before. The n-back tasks engage not only WM processes, but also IP efficiency (Aguirre et al., 2021, 2019; Turtola and Covey, 2021). Consequently, n-back training could be an appropriate strategy to

---

* Corresponding author.
E-mail address: forn@uji.es (C. Forn).

https://doi.org/10.1016/j.msard.2022.104182
Received 25 July 2022; Received in revised form 5 September 2022; Accepted 11 September 2022

improve these functions in PwMS. Effects of training using back tasks have been extensively studied in healthy subjects, and results confirm that repeated practice leads to substantial improvement in task performance (Soveri et al., 2017). These studies differentiate between two types of gains: near and far transfer effects (Soveri et al., 2017; Turtola and Covey, 2021). In the case of near transfer effects, after WM interventions, we can observe an improvement in performance on the same trained task or other "similar" WM tasks. In contrast, far transfer effects refer to improvements in other cognitive domains that are different from the training tasks in nature (Covey et al., 2018; Soveri et al., 2017; von Bastian and Oberauer, 2014). Training effects using back tasks have been extensively studied, with results showing medium-to-small near transfer effects to other WM tasks and to other untrained versions of the n-back task (see the meta-analysis by Soveri et al. 2017). However, evidence about far transfer effects remains inconclusive (Jaeggi et al., 2010; Morrison and Chein, 2011; Soveri et al., 2017).

Likewise, other studies have examined the efficacy of n-back training in PwMS (Aguirre et al., 2021, 2019; Bonzano et al., 2020; Covey et al., 2018; Turtola and Covey, 2021). In this context, the few studies carried out until now provide evidence indicating that, after the n-back training, PwMS show improvements on the trained task that are also accompanied by neuroplasticity processes similar to those observed in healthy controls (Aguirre et al., 2021, 2019; Bonzano et al., 2020; Covey et al., 2018; Turtola and Covey, 2021). In fact, some of these studies (Bonzano et al., 2020; Covey et al., 2018) have also shown that n-back training not only enhances the performance on the trained task, but it also produces near transfer effects to other WM and IP efficiency tasks, as well as far transfer effects (e.g., improved attention and reasoning abilities).

The present study was designed to extend the previous findings by examining the gain and transfer effects after intensive and adaptive n-back training in a sample of PwMS with different cognitive profiles. Based on previous results (Aguirre et al., 2021, 2019; Covey et al., 2018), we anticipated that, after training, participants would exhibit an improvement on the trained task, as well as near transfer effects to other tests assessing WM and IP functions. Considering the inconclusive results related to far transfer effects in healthy subjects (Soveri et al., 2017), we also examined the occurrence of these far transfer effects after n-back training in order to extend the knowledge about these gains to PwMS.

2. Materials and methods

2.1. Participants

The sample in this study was composed of a group of 35 PwMS (n=35) diagnosed according to the revised McDonald criteria (Thompson et al., 2018) and neurologically assessed with the Expanded Disability Status Scale (EDSS; Kurtzke, 1983). Of the enrolled patients, 25 were classified as relapsing-remitting (RR), 5 as having a secondary progressive (SP) form, and 5 as primary progressive (PP). Exclusion criteria in this study included having a degree of motor impairment that interfered with the training and neurological or psychiatric disorders other than MS or substance abuse. None of the participants presented a relapse three months before or during their participation in the study, and none were receiving corticoid treatment in that period.

Participants were neuropsychologically assessed using the Spanish version of the Brief Repeatable Battery of Neuropsychological Tests (BRB-N) (Sepulcre et al., 2006), which includes the Paced Auditory Serial Addition Test (PASAT) to assess WM and IP, the Symbol Digit Modalities test (SDMT) to assess IP efficiency and attention, the Selective Reminding Test (SRT) and the Spatial Recall Test (SPART) to assess verbal and visuospatial learning and long-term memory, respectively, and phonetic and semantic verbal fluency to assess executive functions. Versions A and B of this battery were generally adopted for baseline and post-treatment assessments, respectively. However, because 7 (out of 35) participants had been assessed two years before with Version A of this battery, their first assessment was performed with Version B and the post-treatment assessment with Version A. The cognitive assessment also included other WM tests to assess possible near transfer effects, such as the Digit-Span Backward subtest (DSBT), the Letters and Numbers Sequencing subtest (LNST), and the Digit-Span Forward subtest (DSFT) to measure short verbal memory/attention (WAIS; Wechsler, 2002). We also added executive measures to evaluate possible large transfer effects, including the Stroop Color and Word Test (SCWT; Golden, 2020) and the computerized version of the Tower of London test (Krikorian et al., 1994). In addition, other clinical measures were taken before the training period in order to describe the sample: the Matrix reasoning subtest (WAIS III; Wechsler, 2002) to assess the intelligence quotient (IQ); the Modified Fatigue Impact Scale (MFIS) derived from the original Fatigue Impact Scale (Fisk et al., 1994); and the Spanish version of the Beck Depression Inventory (BDI-II; Sanz et al., 2005). Assessments were carried out by two expert neuropsychologists before training (D1) and after training (D10). All PwMS provided their written informed consent to participate in this study, according to the Declaration of Helsinki, and they were awarded 250 € for their participation. The study was approved by the Ethics Committee of the Universitat Jaume I (UJI). This trial was registered in the ClinicalTrials.gov registry as NCT05270239.

2.2. Training protocol

After the baseline assessment, participants came to the university to complete ten WM training sessions on ten consecutive days. The training sessions have been described in previous studies (Aguirre et al., 2021, 2019). They had a total duration of 60 min and were distributed in two phases: the training phase and the testing phase. During the training phase, participants performed three runs, each composed of eight blocks that varied the WM load (1-back, 2-back, and 3-back). For motivational reasons (Schneiders et al., 2012), all the training sessions started with the least demanding block (1-back), and the subsequent block's WM load depended on the participant's performance on the previous block. Thus, 90% accuracy in their performance led to an increase in the WM load in the following block (e.g., the 2-back block increases to the 3-back block), but an accuracy below 80% led to a decrease in the WM load (e.g., from 2-back to 1-back). In all other cases, the n-back level remained constant. During the testing phase, PwMS performed four 2-back blocks and four 3-back blocks that were not performance dependent but rather randomly displayed, and CRs and RT were registered to observe the patient's daily progress. Subsequently, these data were statistically analyzed in different ways to study the gain effects of the training.

2.3. Statistical analysis

All statistical analyses were conducted in R (R Development Core Team, 2019). Following current recommendations (Wasserstein and Lazar, 2016), statistical significance was tested, but effect sizes (and their 95% bootstrapped confidence intervals, CIs) were also estimated. Moreover, these analyses were conducted using non-parametric, distributionally-robust, and outlier-resistant methods, which offer more statistical power and are informationally richer than more conventional methods (i.e., parametric tests and standardized average differences) (Mair and Wilcox, 2020; Wilcox, 2016).

Therefore, as a first approach, mean-based comparisons and Cohen's d values were calculated. However, deeper, more accurate, and more nuanced understanding of the data was obtained by comparing cumulative distribution functions (CDFs; Callaert, 1999; Handcock and Morris, 1998; Wilcox and Rousselet, 2018) of the 2- and 3-back net scores before (D1) and after (D10) repeated n-back practice. This allowed us to compare these two conditions in three complementary ways: (1) by estimating and comparing the three quartile values of the D1 and D10 distributions. These estimations and comparisons were carried out with the *Dqcomhd* function (bootstrap: 10,000 repetitions) of the WRS2 package (Mair and Wilcox, 2020), which uses the unbiased Harrell-Davis estimator and automatically adjusts p-values for multiple comparisons); (2) by directly contrasting the proportion of PwMS in each condition whose net scores were above three meaningful cutoffs (scores $\geq$ 6, 12, and 18, which translate to accuracies $\geq$ 25, 50, and 75% respectively); (3) by estimating the proportion of subjects with D10 scores equal to or higher than the mean of the D1 distribution (Cohen U3; Cohen, 1962; Grissom and Kim, 2012). In a second step, two indexes of individual performance improvement were calculated: the net score improvement index (NSI) and the percent of maximum possible individualized improvement (PMPI). The NSI was calculated as the raw difference between the D10 and D1 net scores, and, when necessary, this index was translated to the percentual gain in accuracy by using the formula: 100*(NSI/24). The PMPI was designed to provide an alternative measure of performance improvement independent from the baseline (D1) scores. The PMPI was calculated using the formula: 100*{(D10 net score- D1 net score)/(max net score- D1 net score)}. The 95% CIs of these two indexes were calculated using the percentile bootstrap method (10,000 repetitions), and, when necessary, these CIs were inverted to estimate p-values for the null hypothesis that the calculated index was equal to 0. In addition, the robust Spearman's correlation index was used to quantify: (1) the relationship between the NSI and the D1 and D10 net 2/3 back scores; (2) the relationship between the PMPI and the D1 and D10 net 2/3 back scores; (3) the relationship between the NSI and the PMPI. Finally, Silver's test was used to compare the strength of this correlation in the 2-back and 3-back conditions (Diedenhofen and Musch, 2015; Silver et al., 2004).

The same approach was used to analyze the effects of repeated n-back practice on the 2- and 3-back reaction times (RTs) on D1 and D10. However, in this case, no a priori meaningful cutoffs can be established, and so CDF analysis was restricted to the comparison of the quartile values on D1 and D10 and to the calculation of Cohen's U3. Similarly, because there is no *a priori* maximum reduction, no PMPI was calculated, and individual performance improvement was solely assessed in terms of the raw difference between D10 and D1 RTs (RT improvement). To facilitate their comparison with score changes, RT changes were inverted and reported as the net reduction in milliseconds, so that a larger positive RT improvement value denotes a larger RT reduction (that is, a larger negative value in the D10 minus D1 difference).

CDFs and quartile comparisons were also used to compare the performance of MS patients on other cognitive tests that were administered before and after n-back repeated practice. Statistically significant post-pre score changes in these tests were interpreted as suggestive of possible transfer effects of n-back training. These possible effects were further characterized by calculating the number of individuals exhibiting a performance improvement according to three different change indexes: the Sign Discrepancy Score (SDS), the Reliable Change Index (RCI), and the practice-corrected RCI (RCI-Pe).

The SDS just considers the sign of the raw post-pre score difference, and its use has been recommended when the costs associated with Type I and Type II errors are considered to be of similar importance (McAleavey, 2021). The RCI was developed by Jacobson and Truax (1991) to assess individual change while accounting for measurement error, and its being increasingly used to distinguish between spurious and meaningful cognitive variations in PwMS (e.g., Walker et al. 2016). As calculated with the Jacobson and Truax formula, RCIs are considered fairly accurate measures of meaningful change (Strober et al., 2022). However, these RCIs do not account for possible practice effects derived from repeated testing, which are expectable in, at least, some tests (e.g., PASAT, see Tombaugh 2006). Therefore, although the use of the A and B forms of the tests included in the BRB-N battery could have attenuated these possible practice effects (Roar et al., 2016), practice-corrected RCIs (RCI-Pe; Chelune et al.,1993; Walker et al., 2016; Strober et al., 2022) were also calculated.

Because the present study did not include a control (i.e., a non-repeatedly n-back trained) group, the information needed for the calculation of RCI and RCI-Pc individual scores were taken from previously published reports. More specifically: (1) for the PASAT and SDMT tests, we used the RCI values estimated by Sonder et al (2014) in a large sample (n=485) of PwMS, whereas the practice effects were calculated by taking the mean difference between the second and first administration of these tests for 237 Spanish PwMS (López-Gongora et al., 2015); (2) for the DSBT, data obtained in PwMS were not found and the RCI values and the estimated practice effects in 124 Spanish volunteers (the Neuronorma normative sample; Sánchez-Benavides et al., 2016) were employed; (3) for the LNST, we estimated the RCI values and the practice effects from the descriptive statistics and the Pearson's correlation provided in the study of Lemay et al (2004). On the other hand, in agreement with current standards, post-pre score differences exceeding the limits of the 70% confidence interval (Z= 1.036) of the RCI or RCI-Pc values were considered as meaningful. This threshold was chosen because it seems to provide an appropriate compromise within the ample range of cut-off Z-values previously proposed (from 0.6 to 1.96; see McAleavey 2021).

Finally, the robust Spearman's correlation index was used to assess the possible relationship between the different improvement indexes for the 2- and 3-back tasks and those for other tests showing statistically significant post-pre differences. For this correlational analysis, possible performance changes were operationalized as the raw difference between the post n-back training scores (D10) minus the pre n-back training (D1) scores on each of these tests.

3. Results

Table 1 shows the descriptive data for demographic and clinical characteristics of the group of patients.

3.1. *Effects of repeated 2-back practice on 2-back performance*

3.1.1. *2-back scores*

Fig. 1A shows that repeated practice improved 2-back performance in PwMS. Thus, as revealed by Student's t test, there was a statistically significant (t=6.48, $p < 0.001$) and "large" (d=1.1) difference between the means of the net scores after (D10) and before (D1) repeated 2-back training. More robust evidence for this performance improvement was obtained after comparing the CDFs for the 2-back net scores on D1 and D10. Thus, as Fig. 1B illustrates, all quartile values were larger on D10 than on D1 ($p < 0.001$ in all cases), thus revealing that repeated practice leads to a consistent shift towards higher 2-back scores across the entire distribution. This conclusion is further reinforced by additional measures that can be extracted from the same figure. For example, the relative number of PwMS exhibiting net scores $\geq$ 6, 12, and 18 (that is, accuracies $\geq$ 25, 50, and 75%) was 88.6%, 77.1%, and 34.3% on D1, but these percentages increased to 100%, 97.1%, and 68.6% on D10.

Table 1

Demographic and clinical characteristics of the participants.

	N=35 (19F, 16M); multiple sclerosis subtypes: 25 RR, 5SP, 5PP		
	Median	Mean	SD
Age	44	43.5	9.2
IQ	110	108.9	10.8
Disease years	8	10.7	9.5
EDSS	1.5	2.8	2.4
Educational level	4	3.4	0.9
Education years	14	13.6	2.8
FAMS	120	117.6	22.1
MFIS	44	41.5	20.9
BDI-II	11	11.6	9.4

Abbreviations: IQ: Intellectual quotient; EDSS: Expanded disability status scale; FAMS: Functional assessment of multiple Sclerosis; MFIS: Modified fatigue impact scale; BDI-II: Beck depression inventory II; RR: relapsing-remitting; SP: secondary progressive; PP: primary progressive.

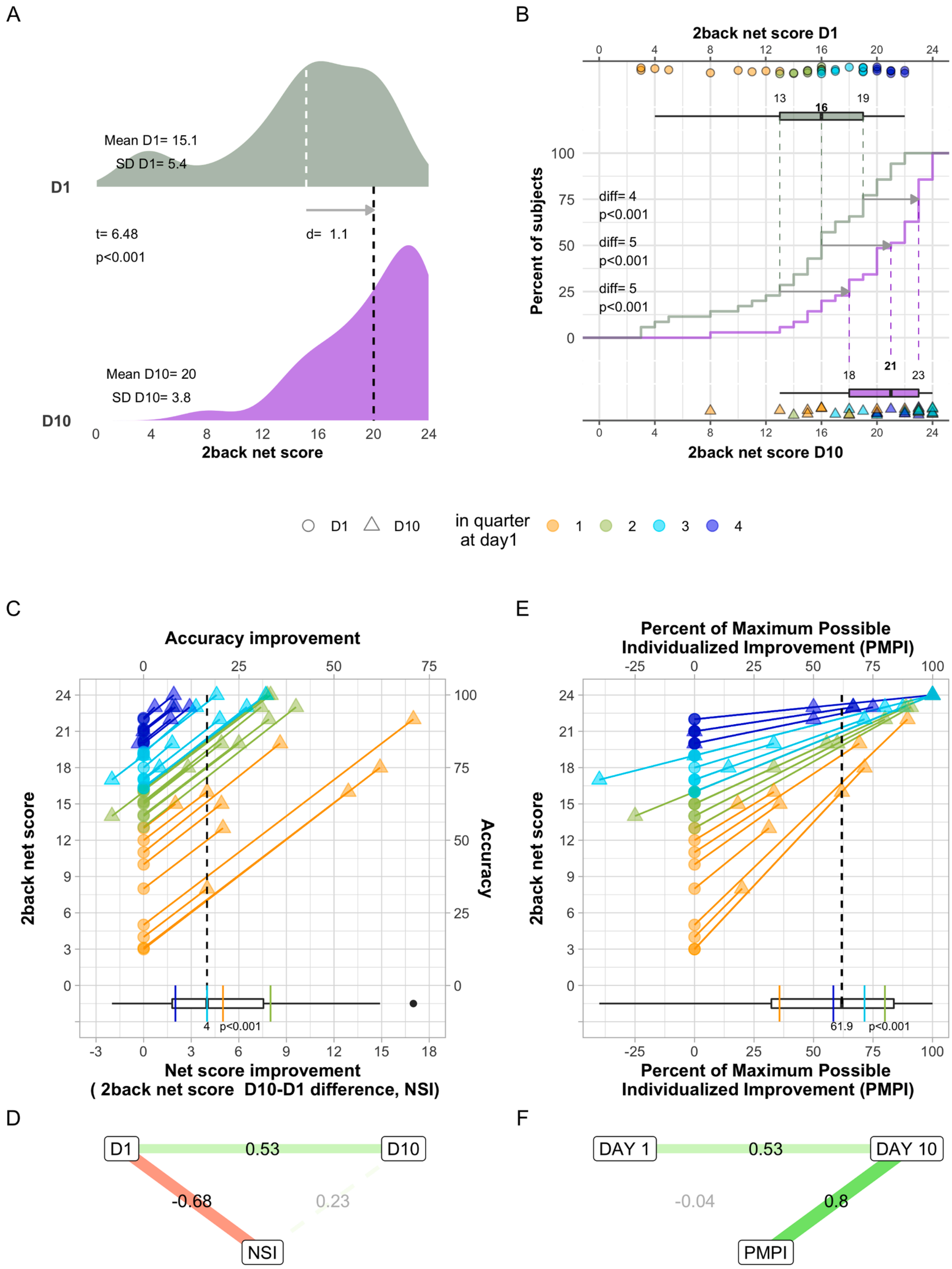


Fig. 1. Repeated n-back practice improves 2-back net scores in persons with multiple sclerosis (PwMS). Panel A depicts the distributions of the 2-back net scores before (D1) and after (D10), their respective means, and the details about this between-means difference. Panel B illustrates the cumulative distributive functions (CDFs) of the same scores, their respective quartile values, and the details about their differences in these quartile values. The same panel also includes the dot plots for the individual 2-back net scores on D1 (circles) and D10 (triangles), where each dot is colored to denote the quarter where each individual was located on D1. The same color scheme is used in Panels C and E, which depict the individual 2-back net scores/ accuracy on D1 and D10 (Y axes) as a function of their score/ accuracy improvement (NSI) or the proportion of maximum possible individualized improvement (PMPI), respectively. The distribution of each of these improvement measures is summarized in a boxplot at the bottom of each panel, where the net score improvement index (NSI) medians for the D1-based quarters are also depicted (colored vertical bars). Panels D and E depict the pattern of correlations (Spearman's rho) between the 2-back scores and the NSI and the PMPI.

Finally, 85.73% ([77.14, 97.14], $p < 0.001$) of participants showed D10 net scores that were higher than the mean D1 score (Cohen's U3).

These distribution-level effects were due to generalized, but probably non-uniform, improvement at the individual level (Fig. 1C-H). As Fig. 1C shows, 30 out of 35 patients (85.7%) enhanced their 2-back performance, and the median net score improvement (NSI) was significantly greater than 0 (4 ([2, 5], accuracy gain=17% [8.3, 21]; $p < 0.001$). The NSI magnitude varied between PwMS (range: -2, 17), and it tended to be larger/ smaller in individuals with small/ large 2-back D1 scores, respectively (Fig. 1C and 1D). However, this seems to occur because the 2-back is a relatively easy task on which repeated training leads to a ceiling effect in 2-back scores that artificially reduces the NSI scores of high-performing individuals. This conclusion is supported by several sources of evidence: First, the maximum possible 2-back score is less than 1.5 standard deviation from the mean of the 2-back D10 net scores ($20+ 1.5*3.8=25.7 > 24$), indicating that these scores are likely to be limited by a ceiling effect (Uttl, 2005). Second, the variance in the 2-back scores on D10 was half of what was observed on D1 ($D10_{var}=14.76$, $D1_{var}=28.95$; $p < 0.02$), and 37.1% of PwMS exhibited scores ≥23 (accuracy >95%) on D10, which means they were probably not able to demonstrate their true level of performance on this task (Uttl, 2005). Third, the NSI magnitude was significantly correlated with the D1 scores (rho=-0.68, $p < 0.001$), but not with the D10 scores (rho=0.23, p=0.19; Fig. 1D). Finally, the median accuracy improvement (17%) was substantially lower than the median PMPI (62% [35.7, 71.0]), which shares the same scale but is independent from D1 scores (Fig. 1E and 1F). The fact that PMPI and NSI show a tight ordinal correlation (rho=0.71, $p < 0.001$) but yield very different numeric values confirms that the 2-back NSI is severely affected by a ceiling effect that obscures the performance gains of PwMS with high scores on D1.

Taken together, these results indicate that repeated 2-back practice results in improved performance (higher net scores). However, there seems to be a ceiling effect that impedes appropriately quantifying this performance enhancement with the NSI, whereas it seems to be more accurately captured by the PMPI.

3.1.2. 2-back reaction time

Fig. 2A shows that the mean reaction time (RT) on D10 was slightly reduced compared to what was observed on D1 (t=-2.51, $p < 0.05$, d=-0.42). As subsequent analyses revealed (Fig. 2B), this effect is not only "small" but also unreliable (e.g., the U3 value was 60%, but its 95%CI [42.9, 82.9] contains the null value of 50%, p=0.19), and it seems to result from a non-uniform shift that only reaches significance at the left tail of the RT distribution (Q25). Moreover, as Fig. 2C shows, only 57.14% of the participants reduced their RTs, and the median RT reduction was not significantly different from zero (28.18 [-23.4, 126], p=0.41). RT changes tended to be larger for PwMS with larger RTs on D1 (Fig. 2C and 2D); however, even for these individuals, the median RT reduction was not statistically different from zero. Taken together, these results suggest that, contrary to what was observed with the 2-back scores, repeated 2-back practice did not substantially or consistently improve 2-back RTs.

3.2. Effects of repeated 3-back practice on 3-back performance

3.2.1. 3-back scores

Fig. 3A shows that repeated practice resulted in a "large" enhancement of 3-back scores (t=8.27, $p < 0.001$; d=1.4). More reliable grounds for the same conclusion were obtained by comparing the CDFs of the 3-back scores on D1 and D10 (Fig. 3B). Thus, all the quartile values were significantly larger ($p < 0.001$) on D10 than on D1. Moreover, 80, 28.6, and 2.9% of patients exhibited accuracies ≥ 25, 50, and 75% on D1, but these numbers rose to 96.3, 80, and 34.3% on D10. In a similar vein, 31 subjects showed 3-back D10 scores that were higher or equal to the D1 mean (U3=91.4% [80, 100], $p < 0.001$). Therefore, we can conclude that repeated 3-back practice leads to a generalized and consistent improvement in 3-back performance.

At the individual level, 32 (91.42%) subjects enhanced their 3-back performance, and the typical NSI was statistically different from 0 (median=7 [3, 9], accuracy gain= 29% [12.5, 37.5]; $p < 0.001$; Fig. 3C). The size of this increase tended to be larger for individuals with lower D1 3-back net scores. However, probably due to the greater difficulty of this task, this effect was less pronounced than on the 2-back tasks, and it did not seem to produce a ceiling effect on NSI scores. In fact, contrary to what was observed on the 2-back task: (1) The mean of the 3-back scores was more than 1.5 deviations from the maximum possible score. (2) The variances in the 3-back scores on D10 and D1 did not differ ($D10_{var}=27.3$, $D1_{var}=21.63$; p=0.52), and only 5.7% of the PwMS exhibited an accuracy >95%. (3) The 3-back NSI was significant and similarly related to the 3-back scores on D1 and D10 (Fig. 2E). (4) Finally, although the median of the accuracy improvement (29%) was still smaller than the median PMPI (45% [33.3, 53.8]), the size of this difference was less than half of the one observed on the 2-back task, whereas the NSI-PMPI correlation (rho=0.89, $p < 0.001$) was significantly higher than the one observed on the 2-back (Z=2.71, p=0.006).

Taken together, these results suggest that repeated 3-back practice leads to a "large" improvement in performance that seems to be appropriately captured by the PMPI, but also by the NSI scores.

3.2.2. 3-back reaction time

Fig. 4A shows that the 3-back mean RT on D10 exhibits a "moderate to large" reduction compared to what was observed on D1 (t=-4.68, $p < 0.001$, d=-0.79). Confirming and extending this initial observation, 82.9% ([68.6, 91.4], $p < 0.001$) of the PwMS exhibited D10 RTs that were lower than the mean RT on D1 (U3), and all the bootstrap estimated quartile values for the D10 scores were significantly smaller than those estimated on D1 (Fig. 4B). The size of these differences varied among the quartiles (Q25< Q50< Q75), suggesting that repeated 3-back practice reduces 3-back RTs, especially in individuals with greater RTs.

Moreover, as Fig. 4C shows, 27 out of 35 (77.14%) PwMS reduced their RTs and, contrary to what was observed on the 2-back task, the typical RT reduction was significantly different from 0 (113.3 [51.1, 186.8], $p < 0.001$). This reduction tended to be larger in individuals with longer RTs on D1 (Fig. 4C and 4D); however, as in the case of the 3-back scores, the 3-back RT improvement did not seem to be preconditioned by these initial values. Thus, RT variances on D1 and D10 did not significantly differ ($D10_{var}= 30421.3$, $D1_{var}= 24448.8$; p=0.49), and RT reductions were similarly correlated with D1 and D10 RTs (Fig. 4D). Therefore, we can conclude that repeated 3-back practice resulted in a generalized, but not uniform, reduction in 3-back RTs, and this RT improvement was unbiasedly captured by the calculated D10-D1 RT differences.

3.3. Transference of the effects of repeated n-back practice to untrained tasks

Fig. 5 summarizes the PwMS performance on tasks assessing executive functions (Stroop, London Tower, Semantic Fluency, Phonetic Fluency), attention (DSFT), IPS/ attention (SDMT), WM (PASAT, LNST, DSBT), and learning and long-term memory (SRT and SPART) before and after repeated n-back practice. As can be observed, statistically significant pre-post differences (adjusted $p < 0.05$) were obtained for: (1) three quartile values of the PASAT and the SMDT; (2) two quartiles of the LNST; and (3) one quartile of the DSBT. The use of a more liberal criterion (unadjusted $p < 0.1$) suggested possible effects in the second quartile of the DSBT, DSFT, and the London Tower (total time), as well as in the third quartile of the London Tower (score). Taken together, these results suggest that repeated n-back training leads to enhanced cognitive performance in the WM and IPS/attention domains.

To better characterize and quantify these possible effects of repeated n-back training on other tests measuring WM/ IPS capabilities, the number of individuals exhibiting a performance improvement in the

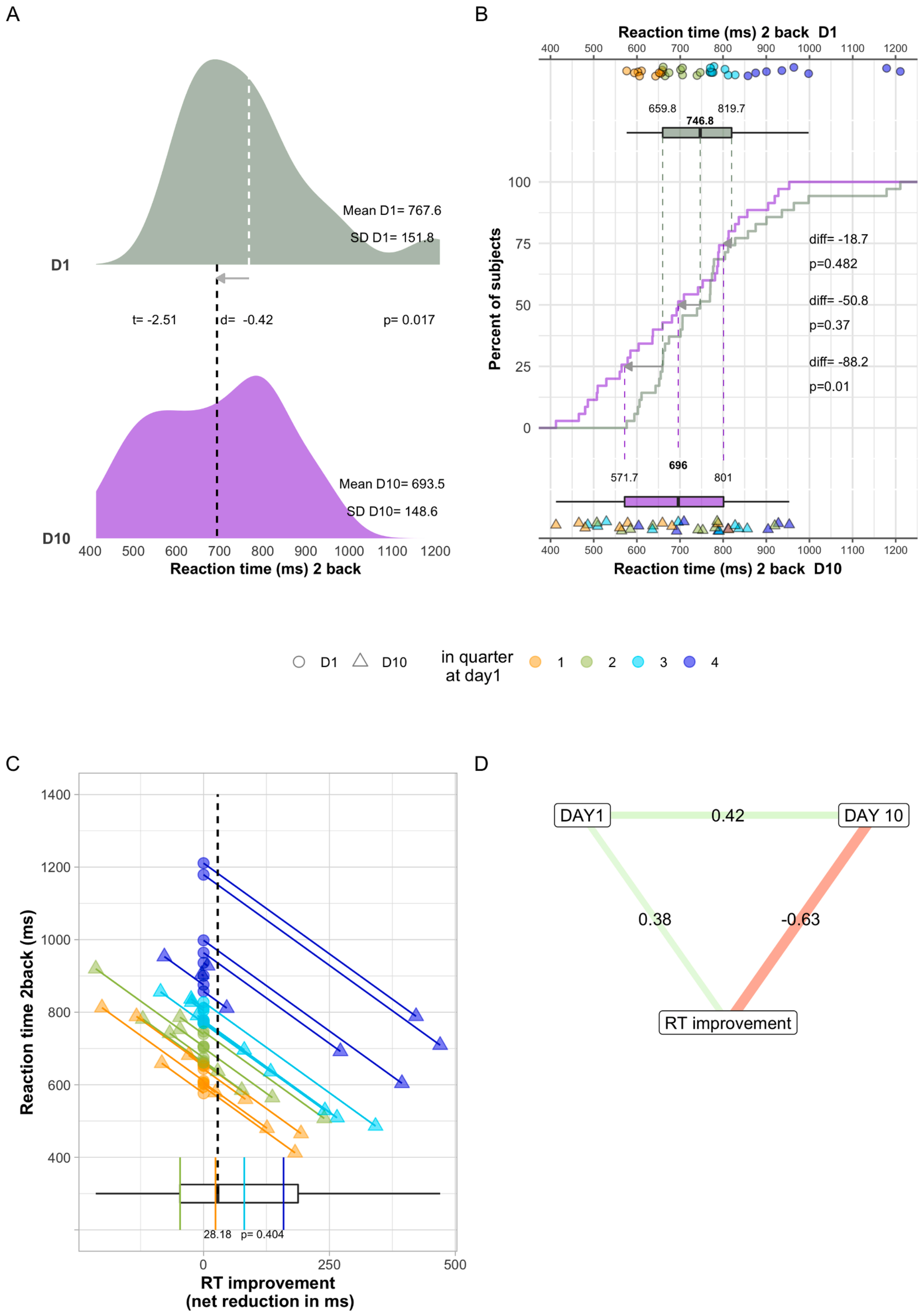


Fig. 2. Repeated n-back practice does not seem to improve 2-back RT in persons with multiple sclerosis (PwMS). Panel A depicts the distributions of the 2-back RTs before (D1) and after (D10), their respective means, and the details about their means' difference. Panel B illustrates the cumulative distributive functions of these RTs, their respective quartile values, and their differences. Dot plots for the individual 2-back RTs on D1 (circles) and D10 (triangles) are also included, with each dot colored to denote the quarter in which each individual was located on D1. The same color scheme is used in Panel C, which depicts the individual 2-back RTs on D1 and D10 (Y axis) as a function of RT improvement. The RT improvement distribution is summarized in a boxplot at the bottom of the panel, where the median RT improvement values of the subjects belonging to each D1-based quarter are also depicted (colored vertical bars). Panel D depicts the pattern of correlations (Spearman's rho) between the 2-back RTs on D1 and D10 and the RT improvement.

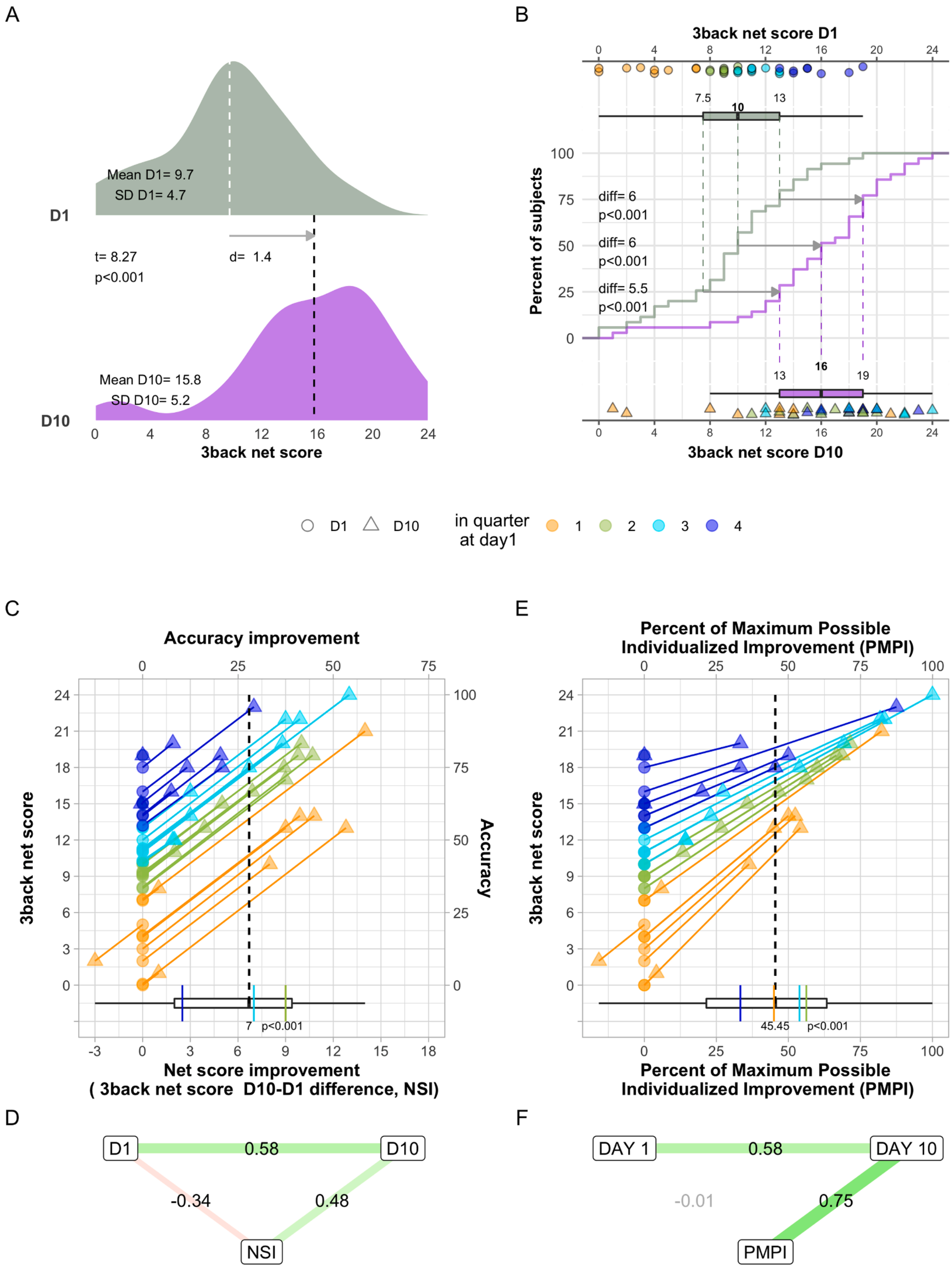


Fig. 3. Repeated n-back practice improves 3-back net scores in persons with multiple sclerosis (PwMS). Panel A depicts the distributions of the 3-back net scores before (D1) and after (D10), their respective means, as well as the details about this between-averages difference. Panel B illustrates the cumulative distributive functions of the same scores, their respective quartile values, and the details about the differences in these values. The same panel also includes the dot plots for the individual 3-back net scores on D1 (circles) and D10 (triangles), in which each dot is colored to denote the quarter where each individual was located on D1. The same color scheme is used in Panels C and E, which depict the individual 3-back net scores/ accuracy on D1 and D10 (Y axes) as a function of their score/ accuracy improvement (NSI) or the percent of maximum possible individualized improvement (PMPI), respectively. The distribution of each of these improvement measures is summarized in a boxplot at the bottom of each panel, where the NSI medians for the D1-based quarters are also depicted (colored vertical bars). Panels D and E depict the pattern of correlations (Spearman's rho) between the 3-back scores and the NSI and the PMPI.

SDMT, PASAT, LNST, and DSBT tests according to three different change indexes (the Sign Discrepancy Score, the Reliable Change Index, and the practice-corrected RCI) was calculated. The obtained results (Table 2) confirmed and extended those obtained in the previously performed quantile comparisons (Fig. 5). More specifically: (1) The number of PwMS exhibiting a SDS>0 (that is, a larger score after than before n-back training) differed among tests and followed the same ordering (SDMT=PASAT > LNST > DSBT) than the number of quantile values

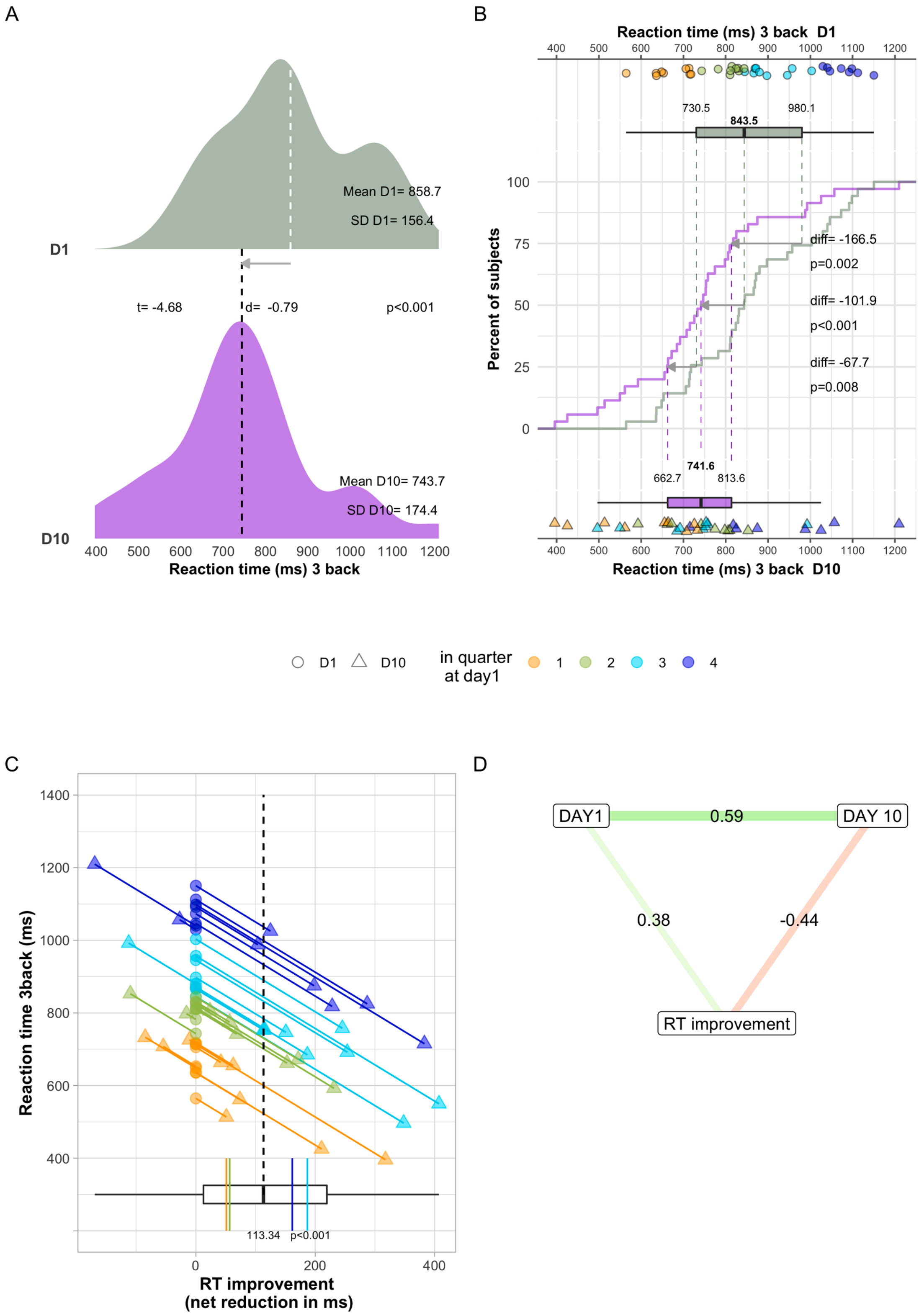
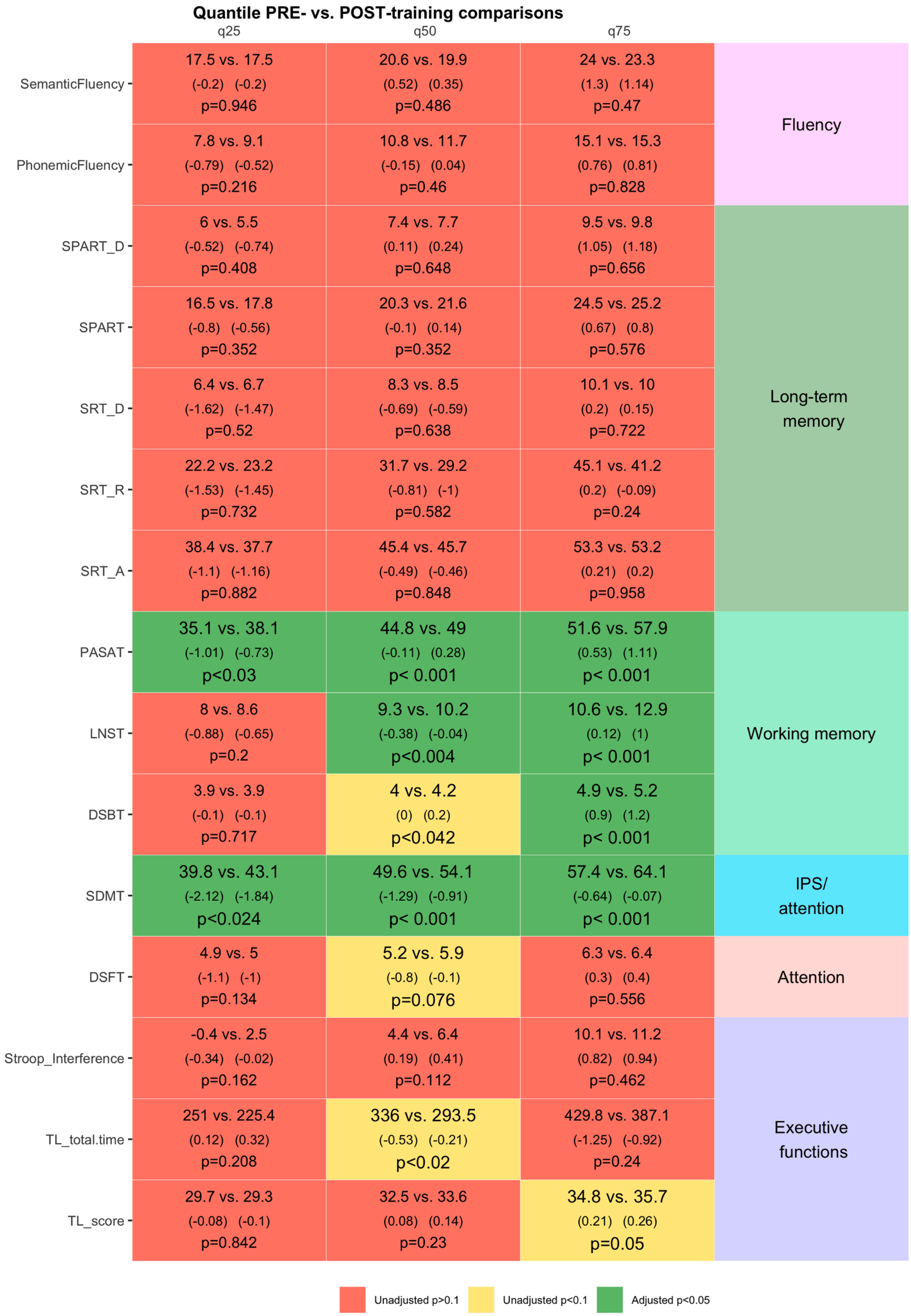


Fig. 4. Repeated n-back practice improves 3-back RT in persons with multiple sclerosis (PwMS).. Panel A depicts the distributions of the 3-back RTs before (D1) and after (D10), their averages, and the details about the observed means' difference. Panel B illustrates the cumulative distributive functions of these RTs, their respective quartile values, and their differences. Dot plots for the individual 3-back RTs on D1 (circles) and D10 (triangles) are also included, with each dot colored to denote the quarter where each individual was located on D1. The same color scheme is used in Panel C, which depicts the individual 3-back RTs on D1 and D10 (Y axis) as a function of RT improvement. The RT improvement distribution is summarized in a boxplot at the bottom of the panel, where the median RT improvement values of the subjects belonging to each D1-based quarter are also depicted (colored vertical bars). Panel D depicts the pattern of correlations (Spearman's rho) between the 3-back RTs on D1 and D10 and the RT improvement.



Quantile PRE- vs. POST-training comparisons

	q25	q50	q75	
SemanticFluency	🟥 17.5 vs. 17.5 (-0.2) (-0.2) p=0.946	🟥 20.6 vs. 19.9 (0.52) (0.35) p=0.486	🟥 24 vs. 23.3 (1.3) (1.14) p=0.47	Fluency
PhonemicFluency	🟥 7.8 vs. 9.1 (-0.79) (-0.52) p=0.216	🟥 10.8 vs. 11.7 (-0.15) (0.04) p=0.46	🟥 15.1 vs. 15.3 (0.76) (0.81) p=0.828	
SPART_D	🟥 6 vs. 5.5 (-0.52) (-0.74) p=0.408	🟥 7.4 vs. 7.7 (0.11) (0.24) p=0.648	🟥 9.5 vs. 9.8 (1.05) (1.18) p=0.656	Long-term memory
SPART	🟥 16.5 vs. 17.8 (-0.8) (-0.56) p=0.352	🟥 20.3 vs. 21.6 (-0.1) (0.14) p=0.352	🟥 24.5 vs. 25.2 (0.67) (0.8) p=0.576	
SRT_D	🟥 6.4 vs. 6.7 (-1.62) (-1.47) p=0.52	🟥 8.3 vs. 8.5 (-0.69) (-0.59) p=0.638	🟥 10.1 vs. 10 (0.2) (0.15) p=0.722	
SRT_R	🟥 22.2 vs. 23.2 (-1.53) (-1.45) p=0.732	🟥 31.7 vs. 29.2 (-0.81) (-1) p=0.582	🟥 45.1 vs. 41.2 (0.2) (-0.09) p=0.24	
SRT_A	🟥 38.4 vs. 37.7 (-1.1) (-1.16) p=0.882	🟥 45.4 vs. 45.7 (-0.49) (-0.46) p=0.848	🟥 53.3 vs. 53.2 (0.21) (0.2) p=0.958	
PASAT	🟩 35.1 vs. 38.1 (-1.01) (-0.73) p<0.03	🟩 44.8 vs. 49 (-0.11) (0.28) p< 0.001	🟩 51.6 vs. 57.9 (0.53) (1.11) p< 0.001	Working memory
LNST	🟥 8 vs. 8.6 (-0.88) (-0.65) p=0.2	🟩 9.3 vs. 10.2 (-0.38) (-0.04) p<0.004	🟩 10.6 vs. 12.9 (0.12) (1) p< 0.001	
DSBT	🟥 3.9 vs. 3.9 (-0.1) (-0.1) p=0.717	🟨 4 vs. 4.2 (0) (0.2) p<0.042	🟩 4.9 vs. 5.2 (0.9) (1.2) p< 0.001	
SDMT	🟩 39.8 vs. 43.1 (-2.12) (-1.84) p<0.024	🟩 49.6 vs. 54.1 (-1.29) (-0.91) p< 0.001	🟩 57.4 vs. 64.1 (-0.64) (-0.07) p< 0.001	IPS/ attention
DSFT	🟥 4.9 vs. 5 (-1.1) (-1) p=0.134	🟨 5.2 vs. 5.9 (-0.8) (-0.1) p=0.076	🟥 6.3 vs. 6.4 (0.3) (0.4) p=0.556	Attention
Stroop_Interference	🟥 -0.4 vs. 2.5 (-0.34) (-0.02) p=0.162	🟥 4.4 vs. 6.4 (0.19) (0.41) p=0.112	🟥 10.1 vs. 11.2 (0.82) (0.94) p=0.462	Executive functions
TL_total.time	🟥 251 vs. 225.4 (0.12) (0.32) p=0.208	🟨 336 vs. 293.5 (-0.53) (-0.21) p<0.02	🟥 429.8 vs. 387.1 (-1.25) (-0.92) p=0.24	
TL_score	🟥 29.7 vs. 29.3 (-0.08) (-0.1) p=0.842	🟥 32.5 vs. 33.6 (0.08) (0.14) p=0.23	🟨 34.8 vs. 35.7 (0.21) (0.26) p=0.05	

🟥 Unadjusted p>0.1 🟨 Unadjusted p<0.1 🟩 Adjusted p<0.05

Fig. 5. Repeated n-back practice improves performance on untrained cognitive tasks. The figure shows the values of the quartiles (columns) of the patients' scores on a set of tasks (rows) used to assess the executive functions (London Tower, Semantic Fluency, Phonetic Fluency), attention (DSFT), IPS/ attention (SDMT), working memory (PASAT, LNST, DSBT), and long-term memory capabilities (SRT and SPART) before and after repeated n-back practice. The figure also includes (between brackets) the estimated Z-scores for these quartile values according to normative data as well as the p-values associated with the observed pre-post differences. Cells in which statistically significant differences ($p < 0.05$ after adjustment for multiple comparisons) were found are colored in green, and the reported values are displayed using a larger font size. Cells in which a trend towards a significant effect (unadjusted $p < 0.1$) was found are filled in yellow. Red colored cells include comparisons that showed no effects (unadjusted p>0.1). (PASAT, Paced Auditory Serial Task; SDMT, Symbol Digit Modalities Test; LNST, Letters-Numbers Sequencing task; DSBT, Digit Span Backwards Test).

Table 2
Number (and percent) of participants exhibiting performance gains in tests measuring working memory/ information processing efficiency.

	SDS	RCI	RCI-Pc
SDMT	29 (82.85%)	17 (48.57%)	17 (48.57%)
PASAT	30 (85.71%)	19 (54.29%)	10 (25.71%)
LNST	19 (54.29%)	15 (42.86%)	15 (42.86%)
DSBT	8 (22.86%)	1 (2.86%)	1 (2.86%)

Abbreviations: SDS: Sign discrepancy score; RCI: Reliable change index; RCI-Pc: Practice-corrected reliable change index; SDMT: Symbol digits modalities test; PASAT: Paced auditory sequential test; LNST: Letter-numbers sequencing test; DSBT: Digits span backwards test.

showing a statistically significant difference before and after repeated n-back practice in these tests (3, 3, 2, and 1, respectively); (2) Around half of the PwMS exhibited a meaningful change in the SDMT, PASAT, and LNST (but not in the DSBT) scores after correcting their post-pre difference scores for measurement error (RCI values); (3) As it could be expected (e.g., Tombaugh 2006; Beglinger et al. 2005; Drake et al. 2010), correcting post-pre difference scores for measurement error and possible practice effects, reduced the number of PwMS at which a meaningful change in the PASAT (but not in the SDMT, LNST, or DSBT tests) could be confirmed. Taken together, these results seem to suggest that repeated n-back practice resulted in near transfer effects that reliably enhanced the performance of a relevant proportion of PwMS in the SDMT, the LNST, and the PASAT tests. Conversely, n-back training did not seem to reliably affect DSBT performance.

Finally, to assess whether the confirmed improvements in the SDMT, PASAT, and LNST and the improvement in the 2- and 3-back tasks were mutually related, a correlational analysis was performed. Thus, Fig. 6A depicts the correlations between the 2-back performance improvement indexes (NSI, PMPI, and RT improvement) and the performance

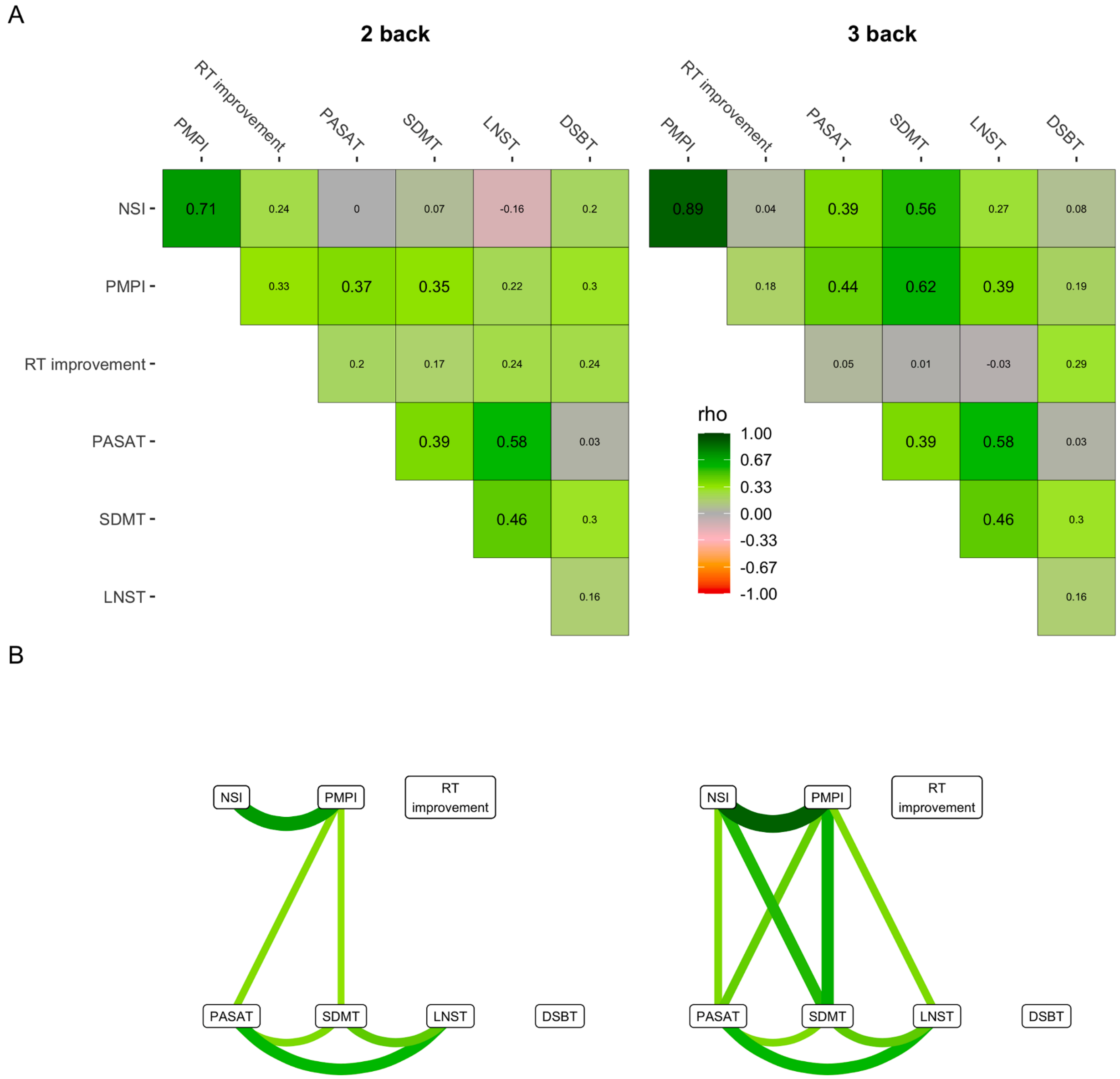


Fig. 6. Correlational analysis of the improvement observed on the 2- or 3-back tasks and the improvement observed on the untrained working memory and IPS/attention tasks. The upper part of the figure contains the correlation matrices for the 2-back (left) and 3-back (right) data. The lower part of the figure schematically depicts the pattern of statistically significant associations found in each of these correlation matrices. Correlations were calculated with Spearman’s rho correlation index (NSI, net score improvement; PMPI, percent of maximum possible individualized improvement; RT, reaction time, PASAT, Paced Auditory Serial Task; SDMT, Symbol Digit Modalities Test; LNST, Letters-Numbers Sequencing Task; DSBT, Digit Span Backwards Test).

improvement (post-pre differences) on the untrained WM and IPS/ attention tasks (PASAT, SDMT, LNST, and DSBT). The main results of this correlational analysis can be summarized as follows: First, the performance improvement (post-pre differences) observed on the PASAT, SDMT, and LNST, but not those observed on the DSBT, were significant and directly correlated with each other. Second, the post-pre differences in the PASAT and the SDMT were also significant and directly correlated with the 2-back PMPI, but not with the 2-back NSI. Finally, RT improvement was not significantly correlated with any pre-post difference on these WM and IPS/attention tasks.

As Fig. 6B shows, very similar results were observed when analyzing the correlations between the 3-back performance improvement indexes (NSI, PMPI, and RT improvement) and the improvement (pre-post differences) on untrained WM and IPS/attention tasks. However, in this case, both the PMPI and the NSI were significantly correlated with the post-pre differences observed on the PASAT and the SDMT. This observation agrees with our previous conclusion indicating that the 3-back, but not the 2-back, NSI unrestrictedly captures the cognitive changes promoted by repeated n-back practice. However, in contrast to what was observed on the 2-back task, a significant correlation was observed between the 3-back PMPI and the pre-post differences on the LNST. Again, this result seems to indicate that, due to its greater difficulty, improvement indexes obtained from 3-back performance are more sensitive and better capture the cognitive changes promoted by repeated practice on n-back tasks.

4. Discussion

In this study, we evaluated the potential benefits of adaptative and intensive n-back training in a cohort of PwMS with different cognitive statuses. The results support that this training program is effective and not only seems to improve performance on the trained tasks but probably also on other tests that measure WM and IP processes. However, we did not find evidence that these beneficial effects are transferred to other cognitive domains.

Thus, in agreement with our previous studies (Aguirre et al., 2021, 2019), we found that intensive training significantly improved the performance of PwMS on 2- and 3-back tasks. These effects were more pronounced when considering task accuracy or net scores rather than RTs, when the WM load was high (3-back), and when they were assessed through individualized improvement scores (PMPI). In this regard, it should be noted that, although statistically significant improvements in task accuracy were evident on the 2- and 3-back tasks, the lower difficulty of the 2-back resulted in a ceiling effect that partially masked the effects of training. More specifically, because some participants already presented a high performance on the 2-back task before starting their training, and given that the maximum number of correct responses on this task is limited to 24, the NSI scores of these individuals probably did not accurately reflect their actual progression on the task and appeared to be lower than those of patients with initially worse 2-back performance. These individual distortions also resulted in a reduction in the median NSI, which was paradoxically lower on the 2-back than on the 3-back (4 and 7, corresponding to a 17% and 29% increase in accuracy, respectively). In contrast, when improvements were calculated in terms of the individual possible maximums (PMPI), training effects could be unrestrictedly estimated, not only generally becoming larger, but also, as would be expected, more prominent on the 2-back than on the 3-back (medians: 62 vs. 45%, respectively). Taken together, these findings replicate those of previous studies (Aguirre et al., 2021, 2019; Covey et al., 2018; Turtola and Covey, 2021) and reinforce the evidence suggesting that PwMS benefit from intensive n-back training to restore their WM and IP efficiency, two of the most affected cognitive domains in this population. Moreover, the individualized analyses conducted in the present study also reveal that at least some degree of improvement occurs in the majority of the participants (85.7 and 91.4% on the 2- and 3-back tasks, respectively). Therefore, it seems that this training program can potentially improve the WM/ IP capabilities of PwMS with different cognitive statuses.

The second aim of this study was to examine the possible near and far transfer effects after the intervention. As mentioned above, we only found evidence for possible near transfer effects because the only tests that showed statistically significant performance improvements after n-back training were those measuring WM and IP capabilities (Forn et al., 2008). More specifically, performance improvements were more pronounced on the PASAT and SDMT, whose post-training scores were significantly enhanced in the three quartile values. We also observed an enhancement of the LNST scores, although in this case, the improvement only reached statistical significance for intermediate and high scores (i.e., Q50 and Q75). Finally, a smaller and even more restricted (Q75) effect of training was also observed on another WM task, the DSBT. The improvements in the PASAT, SDMT, and LNST scores (but not in the DSBT) were confirmed by the obtained RCI individual values in these tests, which also revealed that not all, but many PwMS exhibit reliable and meaningful improvements in WM/ IPS capabilities after n-back training. Of note, the performance improvements in the PASAT, SDMT, and LNST scores (but not in the DSBT) were directly correlated with the accuracy improvements observed on the 2- and 3-back tasks. Again, these correlations were larger and more consistent when the WM load was high (3-back) and/ or when measured with the PMPI. Taken together, these results seem to suggest that the benefits of n-back training are extended to other tasks that engage WM and IP processes.

In this regard, previous studies had shown near-transfer effects similar to those observed in the present study. Thus, Vogt et al. (2009) compared the effectiveness of two kinds of WM training (one intensive and another distributed) in a group of PwMS using the BrainStim software. The authors described improvements in all participants (regardless of the training carried out) on WM/IP tasks such as the PASAT, DSBT, and 2-back tasks. Hancock et al. (2015) also used n-back to improve WM and IP in a group of PwMS. They observed potential gains on the n-back after training that was also translated to improvements in performance on the PASAT task. Nonetheless, other studies have found that WM training results in both near and far transfer effects in PwMS. Thus, Covey et al. (2018) also used n-back training and observed similar results to those found in the present study (i.e., enhanced CR and decreased RT) on the 3-back task, as well as enhanced SDMT average performance in a group of PwMS. However, in contrast to our results, Covey et al. also found a significant improvement in cognitive skills not related to WM and IP, such as fluid reasoning and concept formation. In the same vein, Bonzano et al. (2020) reported that adaptive training in n-back and a personalized visuospatial WM task enhanced the average performance of a group of PwMS on the PASAT, the SDMT, and all the other tests included in the BRB-NT. Finally, improved average performance on two WM tests (the LNST and the DSBT) and a modified Stroop task (but not on the PASAT or the SDMT) was observed in a recent study conducted in 11 PwMS trained with a specific WM training program (Blair et al., 2021). Taken together, these findings suggest that WM interventions in PwMS are useful to improve not only WM but also IP capabilities. These improvements are not limited to the task specifically trained, but (at least in most cases) they also extend to other WM and IP tests, confirming an actual improvement in these two cognitive functions (as opposed to a mere effect of task practice). In addition, far transfer effects are found in some studies, but not in others (including ours). The reasons for this discrepancy between studies remain unclear, although it is worth noting that all the studies that observed far transfer effects used longer training programs than the one in the present study, and two of them were conducted in PwMS preselected according to their cognitive status (i.e. patients exhibiting z scores $\leq$ -1.5 on the PASAT and or the SMDT tests; Blair et al., 2021; Bonzano et al., 2020).

Before concluding, the main limitations of the present study should be mentioned. First, the sample was relatively small, which may potentially result in reduced statistical power and a higher rate of type II errors. However, it should be noted that our sample was larger than those used in all the preceding studies that assessed the effects of WM training and its possible near/ far transference to other tasks. Moreover, we used robust statistical procedures (Wilcox and Rousselet, 2018) that do not assume normality or homoscedasticity, that properly handle skewness or outlier values, and that achieve higher statistical power than traditional parametric methods (see Wilcox 2012). A second (and most relevant) limitation of the present study is the absence of a control group. In this regard, we decided to perform comparisons within one sample, hence increasing statistical power and, consequently, reducing the chance of type II errors. However, this methodological decision forced us to draw upon previously published data when trying to evaluate whether the benefits of the n-back training program used in the present study are extended to other non-trained tasks.

In summary, our results confirm and extend previous studies showing that restorative n-back training improves WM and IP benefits in a heterogeneous sample of PwMS. Moreover, this computerized cognitive training program could easily be adapted to be conducted remotely (hence facilitating its implementation across longer training periods) and combined with other tools in more extensive cognitive rehabilitation programs. However, despite these promising results, more studies are needed to identify and determine the role of possible moderator variables whose optimization could increase the effectiveness of this restorative program.

Funding

This work was supported by a grant funded by MCI/AEI 10.13039/501100011033 provided to CF and CSS and a grant (UJI B2020-02) awarded to CF and CSS. S.F.E A was supported by an FPI grant from UJI (PREDOC/2020/22). These funding sources did not play any role in designing the study or in the collection, analysis, and interpretation of the data.

CRediT authorship contribution statement

Sónia Félix Esbrí: Formal analysis, Conceptualization, Writing – review & editing. **Alba Sebastián Tirado:** Formal analysis, Writing – review & editing. **Carla Sanchis-Segura:** Conceptualization, Project administration, Formal analysis, Writing – review & editing. **Cristina Forn:** Conceptualization, Project administration, Writing – review & editing.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest

Acknowledgments

We appreciate the collaboration of the Associació Esclerosis Múltiple de Castelló (AEMC).

References

Aguirre, N., Cruz-Gómez, Á.J., Esbrí, S.F., Miró-Padilla, A., Bueichekú, E., Broseta, Torres, R., Ávila, C., Sanchis-Segura, C., Forn, C., 2021. Enhanced frontoparietal connectivity in multiple sclerosis patients and healthy controls in response to an intensive computerized training focused on working memory. Mult. Scler. Relat. Disord. 52 https://doi.org/10.1016/j.msard.2021.102976.

Aguirre, N., Cruz-Gómez, Á.J., Miró-Padilla, A., Bueichekú, E., Broseta Torres, R., Ávila, C., Sanchis-Segura, C., Forn, C., 2019. Repeated working memory training improves task performance and neural efficiency in multiple sclerosis patients and healthy controls. Mult. Scler. Int. 2019 https://doi.org/10.1155/2019/2657902.

Beglinger, L.J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D.A., Crawford, J., Fastenau, P.S., Siemers, E.R., 2005. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch. Clin. Neuropsychol. 20 https://doi.org/10.1016/j.acn.2004.12.003.

Blair, M., Goveas, D., Safi, A., Marshall, C., Rosehart, H., Orenczuk, S., Morrow, S.A., 2021. Does cognitive training improve attention/working memory in persons with MS? A pilot study using the cogmed working memory training program. Mult. Scler. Relat. Disord. 49 https://doi.org/10.1016/j.msard.2021.102770.

Bonzano, L., Pedullà, L., Pardini, M., Tacchino, A., Zaratin, P., Battaglia, M.A., Brichetto, G., Bove, M., 2020. Brain activity pattern changes after adaptive working memory training in multiple sclerosis. Brain Imaging Behav. 14, 142–154. https://doi.org/10.1007/s11682-018-9984-z.

Callaert, H., 1999. Nonparametric hypotheses for the two-sample location problem. J. Stat. Educ. 2, 7–7.

Chelune, G.J., Naugle, R.I., Lüders, H., Sedlak, J., Awad, I.A., 1993. Individual change after epilepsy surgery: practice effects and base-rate information. Neuropsychology 7. https://doi.org/10.1037/0894-4105.7.1.41.

Chiaravalloti, N.D., DeLuca, J., 2008. Cognitive impairment in multiple sclerosis. Lancet Neurol. https://doi.org/10.1016/S1474-4422(08)70259-X.

Cohen, J., 1962. The statistical power of abnormal-social psychological research: a review. J. Abnorm. Soc. Psychol. 65 https://doi.org/10.1037/h0045186.

Covey, T.J., Shucard, J.L., Benedict, R.H.B., Weinstock-Guttman, B., Shucard, D.W., 2018. Improved cognitive performance and event-related potential changes following working memory training in patients with multiple sclerosis. Mult. Scler. J. Exp. Transl. Clin. 4 https://doi.org/10.1177/2055217317747626.

Diedenhofen, B., Musch, J., 2015. Cocor: a comprehensive solution for the statistical comparison of correlations. PLoS One 10. https://doi.org/10.1371/journal.pone.0121945.

Drake, A.S., Weinstock-Guttman, B., Morrow, S.A., Hojnacki, D., Munschauer, F.E., Benedict, R.H.B., 2010. Psychometrics and normative data for the Multiple Sclerosis Functional Composite: replacing the PASAT with the Symbol Digit Modalities Test. Multiple Sclerosis Journal 16, 228–237. https://doi.org/10.1177/1352458509354552.

Fisk, J.D., Pontefract, A., Ritvo, P.G., Archibald, C.J., Murray, T.J., 1994. The impact of fatigue on patients with multiple sclerosis. Can. J. Neurol. Sci. 21 https://doi.org/10.1017/S0317167100048691. J. Can. des Sci. Neurol.

Forn, C., Belenguer, A., Parcet-Ibars, M.A., Avila, C., 2008. Information-processing speed is the primary deficit underlying the poor performance of multiple sclerosis patients in the paced auditory serial addition test (PASAT). J. Clin. Exp. Neuropsychol. 30 https://doi.org/10.1080/13803390701779560.

Golden, C.J., 2020. STROOP. Test de Colores y Palabras – Edición Revisada (B. RuizFernández, T. Luque y F. Sánchez-Sánchez, adaptadores), TEA Ediciones.

Grissom, R.J., Kim, J.J., 2012. Effect sizes for research: univariate and multivariate applications. second edition Effect Sizes for Research, 2nd ed. Univariate and Multivariate Applications. https://doi.org/10.4324/9780203803233.

Hancock, L.M., Bruce, J.M., Bruce, A.S., Lynch, S.G., 2015. Processing speed and working memory training in multiple sclerosis: a double-blind randomized controlled pilot study. J. Clin. Exp. Neuropsychol. 37 https://doi.org/10.1080/13803395.2014.989818.

Handcock, M.S., Morris, M., 1998. Relative distribution methods. Sociol. Methodol. 28 https://doi.org/10.1111/0081-1750.00042.

Jacobson, N.S., Truax, P., 1991. Clinical Significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol 59. https://doi.org/10.1037/0022-006X.59.1.12.

Jaeggi, S.M., Studer-Luethi, B., Buschkuehl, M., Su, Y.F., Jonides, J., Perrig, W.J., 2010. The relationship between n-back performance and matrix reasoning - implications for training and transfer. Intelligence 38, 625–635. https://doi.org/10.1016/j.intell.2010.09.001.

Krikorian, R., Bartok, J., Gay, N., 1994. Tower of london procedure: a standard method and developmental data. J. Clin. Exp. Neuropsychol. 16 https://doi.org/10.1080/01688639408402697.

Kurtzke, J.F., 1983. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 33. https://doi.org/10.1212/wnl.33.11.1444.

Lemay, S., Bédard, M.A., Rouleau, I., Tremblay, P.L.G., 2004. Practice effect and testretest reliability of attentional and executive tests in middle-aged to elderly subjects. Clin. Neuropsychol. 18 https://doi.org/10.1080/13854040490501718.

López-Góngora, M., Querol, L., Escartín, A., 2015. A one-year follow-up study of the Symbol Digit Modalities Test (SDMT) and the Paced Auditory Serial Addition Test (PASAT) in relapsing-remitting multiple sclerosis: An appraisal of comparative longitudinal sensitivity. BMC Neurol. 15 https://doi.org/10.1186/s12883-015-0296-2.

Mair, P., Wilcox, R., 2020. Robust statistical methods in R using the WRS2 package. Behav. Res. Methods 52. https://doi.org/10.3758/s13428-019-01246-w.

McAleavey, A.A., 2021. When (Not) to rely on the reliable change index [WWW Document]. doi:10.31219/osf.io/3kthg.

Morrison, A.B., Chein, J.M., 2011. Does working memory training work? the promise and challenges of enhancing cognition by training working memory. Psychon. Bull. Rev. 18, 46–60. https://doi.org/10.3758/s13423-010-0034-0.

R Development Core Team, 2019. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. R Found. Stat. Comput.

Roar, M., Illes, Z., Sejbaek, T., 2016. Practice effect in symbol digit modalities test in multiple sclerosis patients treated with natalizumab. Mult. Scler. Relat. Disord. 10 https://doi.org/10.1016/j.msard.2016.09.009.

Sánchez-Benavides, G., Pena-Casanova, J., Casals-Coll, M., Gramunt, N., Manero, R.M., Puig-Pijoan, A., Aguilar, M., Robles, A., Antúnez, C., Frank-García, A., Fernández-Martínez, M., Blesa, R., 2016. One-year reference norms of cognitive change in Spanish old adults: data from the NEURONORMA sample. Arch. Clin. Neuropsychol. 31 https://doi.org/10.1093/arclin/acw018.

Sanz, J., García-Vera, M.P., Espinosa, R., Fortún, M., Vázquez, C., 2005. Spanish adaptation of the beck depression inventory-II (BDI-II): 3. Psychometric features in patiens with psychological disorders. Clín. Salud 16.

Schneiders, J.A., Opitz, B., Tang, H., Deng, Y., Xie, C., Li, H., Mecklinger, A., 2012. The impact of auditory working memory training on the fronto-parietal working memory network. Front. Hum. Neurosci. https://doi.org/10.3389/fnhum.2012.00173.

Sepulcre, J., Vannotti, S., Hernández, R., Sandoval, G., Cáceres, F., Garcea, O., Villoslada, P., 2006. Cognitive impairment in patients with multiple sclerosis using the brief repeatable battery-neuropsychology test. Mult. Scler. 12 https://doi.org/10.1191/1352458506ms1258oa.

Silver, N.C., Hittner, J.B., May, K., 2004. Testing dependent correlations with nonoverlapping variables: a monte carlo simulation. J. Exp. Educ. 73 https://doi.org/10.3200/JEXE.71.1.53-70.

Sonder, J.M., Burggraaff, J., Knol, D.L., Polman, C.H., Uitdehaag, B.M.J., 2014. Comparing long-term results of PASAT and SDMT scores in relation to neuropsychological testing in multiple sclerosis. Mult. Scler. J. 20 https://doi.org/10.1177/1352458513501570.

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., Laine, M., 2017. Working memory training revisited: a multi-level meta-analysis of n-back training studies. Psychon. Bull. Rev. 24, 1077–1096. https://doi.org/10.3758/s13423-016-1217-0.

Strober, L.B., Bruce, J.M., Arnett, P.A., Alschuler, K.N., DeLuca, J., Chiaravalloti, N., Lebkuecher, A., Di Benedetto, M., Cozart, J., Thelen, J., Cadden, M., Guty, E., Román, C.A.F., 2022. A much needed metric: defining reliable and statistically meaningful change of the oral version symbol digit modalities test (SDMT). Mult. Scler. Relat. Disord. 57 https://doi.org/10.1016/j.msard.2021.103405.

Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., Fujihara, K., Galetta, S.L., Hartung, H.P., Kappos, L., Lublin, F.D., Marrie, R.A., Miller, A.E., Miller, D.H., Montalban, X., Mowry, E.M., Sorensen, P.S., Tintoré, M., Traboulsee, A.L., Trojano, M., Uitdehaag, B.M.J., Vukusic, S., Waubant, E., Weinshenker, B.G., Reingold, S.C., Cohen, J.A., 2018. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. https://doi.org/10.1016/S1474-4422(17)30470-2.

Tombaugh, T.N., 2006. A comprehensive review of the paced auditory serial addition test (PASAT). Arch. Clin. Neuropsychol. 21 https://doi.org/10.1016/j.acn.2005.07.006.

Turtola, Z.P., Covey, T.J., 2021. Working memory training impacts neural activity during untrained cognitive tasks in people with multiple sclerosis. Exp. Neurol. 335 https://doi.org/10.1016/j.expneurol.2020.113487.

Uttl, B., 2005. Measurement of individual differences: Lessons from memory assessment in research and clinical practice. Psychol. Sci. 16 https://doi.org/10.1111/j.0956-7976.2005.01557.x.

Vogt, A., Kappos, L., Calabrese, P., Stöcklin, M., Gschwind, L., Opwis, K., Penner, I.K., 2009. Working memory training in patients with multiple sclerosis - Comparison of two different training schedules. Restor. Neurol. Neurosci. 27, 225–235. https://doi.org/10.3233/RNN-2009-0473.

von Bastian, C.C., Oberauer, K., 2014. Effects and mechanisms of working memory training: a review. Psychol. Res. https://doi.org/10.1007/s00426-013-0524-6.

Walker, L.A.S., Berard, J.A., Bowman, M., Atkins, H.L., Lee, H., Arnold, D., Freedman, M.S., 2016. Longitudinal change in paced auditory serial addition test (PASAT) performance following immunoablative therapy and haematopoietic stem cell transplant in multiple sclerosis. Mult. Scler. Demyelinating Disord. 1 https://doi.org/10.1186/s40893-016-0014-y.

Wasserstein, R.L., Lazar, N.A., 2016. The ASA's statement on p-values: context, process, and purpose. Am. Stat. https://doi.org/10.1080/00031305.2016.1154108.

Wechsler, D., 2002. WAIS-III : test de inteligencia para adultos de Wechsler, Test de inteligencia para adultos: WAIS-III.

Wilcox, R.R., 2016. Introduction to robust estimation and hypothesis testing: 4th edition, Introduction to Robust Estimation and Hypothesis Testing: 4th Edition.

Wilcox, R.R., 2012. Introduction to robust estimation and hypothesis testing, introduction to robust estimation and hypothesis testing. 10.1016/C2010-0-67044-1.

Wilcox, R.R., Rousselet, G.A., 2018. A guide to robust statistical methods in neuroscience. Curr. Protoc. Neurosci. 82 https://doi.org/10.1002/cpns.41.