UHASSELT

KNOWLEDGE IN ACTION

Faculteit Revalidatiewetenschappen

master in de revalidatiewetenschappen en de kinesitherapie

Masterthesis

Validity of Patient-Reported Outcome Measures for Assessing Dual-Task Performance in Daily Life in People with Multiple Sclerosis: an International Multicenter Study

Robin Breugelmans
Sofie Ponsaers

Scriptie ingediend tot het behalen van de graad van master in de revalidatiewetenschappen en de kinesitherapie, afstudeerrichting revalidatiewetenschappen en kinesitherapie bij kinderen

PROMOTOR :
Prof. dr. Peter FEYS

COPROMOTOR :
Mevrouw Zuhal ABASIYANIK

BEGELEIDER :
Mevrouw Anne-Marie D'HOOGE

2022
2023

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT

KNOWLEDGE IN ACTION

Faculteit Revalidatiewetenschappen

master in de revalidatiewetenschappen en de kinesitherapie

Masterthesis

Validity of Patient-Reported Outcome Measures for Assessing Dual-Task Performance in Daily Life in People with Multiple Sclerosis: an International Multicenter Study

Robin Breugelmans
Sofie Ponsaers

Scriptie ingediend tot het behalen van de graad van master in de revalidatiewetenschappen en de kinesitherapie, afstudeerrichting revalidatiewetenschappen en kinesitherapie bij kinderen

PROMOTOR :
Prof. dr. Peter FEYS

COPROMOTOR :
Mevrouw Zuhal ABASIYANIK

BEGELEIDER :
Mevrouw Anne-Marie D'HOOGE

Preface

This is our Master's thesis on 'Validity of Patient-Reported Outcome Measures for Assessing Dual-Task Performance in Daily Life in People with Multiple Sclerosis: An International Multicenter Study.' This is our end work for graduating in Physical Therapy and Rehabilitation Sciences at the University of Hasselt (UHasselt). Our thesis is part of the following international study: 'Psychometric Properties of Patient-Reported Questionnaires Assessing Dual-Task Difficulties in Daily Life in Persons with MS: an International Multicenter Study' conducted by Dr. Zuhal Abasiyanik, and supervised by Prof. Dr. Peter Feys (UHasselt). B1152021000023 is the Belgian code for the study. Data collection in Belgium was conducted at REVAL UHasselt, Noorderhart Rehabilitation & MS, and National MS Centre Melsbroek. We have been partaking in this study since our scientific internship during our first Master's year and have been writing our thesis from December 2022 till May 2023.

Our interest in neurological disorders (young and old) grew during our Bachelor's in Physical herapy and Rehabilitation Science. This is why Sofie chose to specialize in Neurological Disorders and Robin in Pediatric Disorders during the last year of our Master's degree. Our knowledge of Patient-Reported Outcome Measures and the need for validation grew in the previous year. We learned about the statistical analysis needed for validation, which students do not learn in the standard courses. We both see improvements in our scientific knowledge about Multiple Sclerosis, validity, and dual-task problems.

The study from Dr. Abasiyanik investigates the validity and reliability of patient-reported outcome measures (PROM) for assessing dual-task performance in people with Multiple Sclerosis. For our thesis we have included structural, convergent, and discriminative validity. The study is situated in the following research areas: 'Balance & gait' and 'Motor control, cognition & brain.' These areas are fitting because dual-tasking requires a certain cognitive capacity, and many dual-tasks involve walking.

This study is relevant since no research has been done on the validity of the Dual-Task Screening List, the Dual-Task Questionnaire, and Dual-task Impact on Daily-living Activities Questionnaire within the Multiple Sclerosis population in Belgium, Chile, Israel, Italy, Spain, and Turkey. When validated, these Patient-Reported Outcome Measures can be used in clinical practice to identify and assess dual-task problems in daily life in persons with Multiple Sclerosis.

The research question was decided in consultation with our promotor and co-promotor. We followed the original protocol of the study on 'Psychometric Properties of Patient-Reported Questionnaires Assessing Dual-Task Difficulties in Daily Life in Persons with MS: an International Multicenter Study.' We both participated in translating the PROMs, participant recruitment, data collection, and data analysis over 1.5 years in Belgium.

We thank our supervisor, Dr. Zuhal Abasiyanik, for the support and help during this process, for allowing us to learn about all aspects of this study, and for giving us excellent feedback. We would also like to thank Anne-Marie d'Hooge, who helped us during the data collection process to reach the number of participants necessary in Belgium. Lastly, we would like to thank our promotor Prof. Dr. Peter Feys, for broadening our knowledge, supporting our critical thinking, and ensuring we always ask questions.

Robin Breugelmans and Sofie Ponsaers

Oudsbergen and Tongeren, 18 May 2023

Table of contents

Abstract

Background. People with Multiple Sclerosis (PwMS) often show performance decrement during dual-tasking, which can harm their quality of life. The number of validated Patient-Reported Outcome Measures (PROMs) to assess dual-task difficulties is limited.

Objective. To investigate the structural, convergent, and discriminative validity of the Dual-Task Screening List (DTSL), Dual-Tasking Questionnaire (DTQ), and Dual-task Impact on Daily-living Activities Questionnaire (DIDA-Q) in six countries within the Rehabilitation in MS (RIMS) network (Belgium, Chile, Italy, Israel, Spain, and Turkey).

Methods. The total sample consisted of 341 PwMS (median age: 48; mean EDSS: 4.0) and 174 (median age: 44) healthy controls (HC) from Belgium, Chile, Israel, Italy, Spain, and Turkey. Structural validity was assessed by Exploratory Factor Analysis (EFA). Correlation coefficients (Spearman's ρ) were used to evaluate hypotheses for convergent validity. Discriminative validity was assessed with the Mann-Whitney U test. Effect sizes were calculated using Cohen's d, and cut-off values were defined by receiver-operating characteristics (ROC) analysis.

Results. EFA of the total sample resulted in a two-factor structure including 'Motor-motor' and 'Cognitive-motor' for DTQ, DTSL, and DIDA-Q. For convergent validity, 81.84% of the hypotheses were supported, and high correlations were found between DTQ, DTSL, and DIDA-Q. However, there was a low correlation with dual-task walking performance. For discriminative validity, all DT PROMs were able to differentiate between HC and PwMS with a strong effect size.

Conclusion. DTQ, DTSL, and DIDA-Q show good results for convergent, discriminative, and structural validity.

Keywords. Activities of Daily Living; Patient-Reported Outcome Measure; Dual-Tasking; Multiple Sclerosis; Questionnaire validation

Introduction

During daily living, numerous activities require performing two tasks simultaneously. This is defined as dual-tasking (e.g., walking while conversing, eating while listening) (MacPherson, 2018). Different populations experience dual-task (DT) difficulties, especially the elderly and people with neurological disorders, such as Multiple Sclerosis (MS) (Al-Yahya et al., 2011). MS is a chronic autoimmune disease defined by an inflammation of the central nervous system (Alroughani & Yamout, 2018). The disease is characterized by several motor and cognitive symptoms, such as impaired balance, gait dysfunctions, reduced attention, and worsened executive functions (Cinnera et al., 2021).

To be able to DT, the individual has to coordinate their attention to both tasks while they are being performed (MacPherson, 2018). Limitations in this capacity can cause dual-task interference (DTI) and reduced task performance. DTI is measured and quantified by the Dual Task Cost (DTC) (MacPherson, 2018). People with Multiple Sclerosis (PwMS) have functional impairments and pathologic changes in neural networks, which can lead to DTI (McIsaac et al., 2018). Recent studies showed a correlation between a greater disability (EDSS), higher cognitive dysfunction, slower walking speed, and worse cognitive and motor DT performance (Rooney et al., 2020; Hamilton et al., 2009). Additionally, studies often confirmed that DTI is greater in PwMS compared to healthy controls (HC).

Previous studies evaluated DT deficits in lab-based assessments, and it has been shown that dual-task walking speed measured in a laboratory setting is a more representative measure of the usual walking speed in daily life compared to single walking speed (Shema-Shiratzky et al., 2020). This results in the DT measure being a better reflection of daily life circumstances; therefore, it is considered a more ecological measurement. However, these lab-based measures may not quantify the problems PwMS experience in daily living as they are measured in a standardized and safe environment with little to no distraction. Considering that daily life includes a variety of activities as well as walking, it is necessary to measure the real-life difficulties of DT. Daily life difficulties could be measured by patient-reported outcome measures (PROMs).

However, there is a need for validated PROMs to quantify the DT difficulties PwMS experience in daily life. To our knowledge, there has yet to be a detailed investigation of the validity of DT PROMs in the MS population.

Different PROMs assessing DT difficulties have been developed for people with neurological disorders (e.g., Parkinson's Disease (PD), stroke, Traumatic Brain Injury (TBI), and Multiple Sclerosis). The systematic review of Abasıyanık et al. (2022) gives an overview of the DT PROMs and their current psychometric properties. The Dual-Task Questionnaire (DTQ) was developed in English for people with TBI and stroke (Evans et al., 2009). The Dual-Task Screening List (DTSL) was initially developed in Dutch for people with PD (Strouwen et al., 2014). The Dual-task Impact on Daily-living Activities Questionnaire (DIDA-Q) was created in Italian for PwMS (Pedullà et al., 2020). Validity has only been investigated for DTQ and DIDA-Q. Structural validity was rated sufficient for DIDA-Q in PwMS, and construct validity was reported as sufficient and intermediate for DIDA-Q and DTQ, respectively (Pedullà et al., 2020; Eldemir et al., 2022). Discriminative validity has only been documented for DIDA-Q, showing significant differences between different disability levels in PwMS (Pedullà et al., 2020). However, the difference between HC and PwMS is unknown and no research has been done concerning the validity of DTSL. Therefore, implementing PROMs in different languages and populations and investigating the psychometric properties with a robust methodology has been suggested (Abasıyanık et al., 2022).

Measurements can be linked to one or more levels of the framework of International Classification of Functioning, Disability and Health (ICF) to better understand its role in describing aspects of health. This consists of body function and structures, activity, participation, and personal and environmental factors. (Pollard et al., 2009).

This study aimed to investigate structural, convergent, and discriminative validity of the DTSL, the DTQ, and the DIDA-Q for PwMS in six different countries within the Rehabilitation in MS (RIMS) Network (Belgium, Chile, Israel, Italy, Spain, and Turkey).

Method

Participants

The study was conducted in six countries within the Rehabilitation in MS (RIMS) network: Belgium, Chile, Italy, Israel, Turkey, and Spain. In Belgium, PwMS were recruited through two MS rehabilitation centers (Noorderhart Rehabilitation & MS and National MS Centre Melsbroek), social media, and flyers in private practices. HC were mainly recruited through flyers and social media.

An overview of inclusion and exclusion criteria for PwMS and HC can be found in **Table 1**.

Table 1. Inclusion and exclusion criteria

	Persons with MS (PwMS)	Healthy controls (HC)
Inclusion criteria	- Age 18 years and above - Confirmed diagnosis of definite MS - Relapse-free for at least 30 days - EDSS-score below 7 - Minimal cognitive functioning to understand the questionnaires according to the clinical observation and judgment by a clinician specialized in MS	- Age 18 years and above
Exclusion criteria	- Diagnosed with neurological disease other than MS - Cognitive decline that renders the participant incapable of performing tests and questionnaires	- Diagnosed with neurological disease - Any other medical condition that can affect walking and balance - Cognitive decline that renders the participant incapable of performing tests and questionnaires

Abbreviations: EDSS = Expanded Disability Status Scale; MS: Multiple Sclerosis

Study Design

A methodological multicenter study design was used. Ethics approval was obtained by the Ethics Committee University Hasselt and local committees of Noorderhart Rehabilitation & MS and National MS Centre Melsbroek. Each country individually also received approval from its local center. The leading center is Hasselt University, with an approval date of 10th October 2021 (B1152021000023).

Procedure

This study investigated DTQ, DTSL, and DIDA-Q for validity, including structural, convergent, and discriminative validity. Translating the PROMs was necessary as the DTQ was only provided in Dutch (Evans et al., 2009), Hebrew (translated for the international study of Veldkamp, Baert, et al., 2019), and Turkish (Translated by Eldemir et al., 2022), the DTSL was only available in Dutch (Strouwen et al., 2014) and Hebrew (Translated for the international study of Veldkamp, Baert, et al., 2019), and the DIDA-Q was only provided in Italian (Pedullà et al., 2020). **Table 2** gives an overview of the availability of DT PROMs in different languages.

Table 2. Availability or need for translation of DT PROMs

DT PROM	Belgium (Dutch)	Chile (Spanish)	Italy (Italian)	Israel (Hebrew)	Turkey (Turkish)	Spain (Spanish)
DTQ	Available	Translation	Translation	Available	Available	Translation
DTSL	Available	Translation	Translation	Available	Translation	Translation
DIDA-Q	Translation	Translation	Available	Translation	Translation	Translation

Abbreviations: DIDA-Q=Dual-Task Impact on Daily-living Activities Questionnaire; DT PROM=Dual-Task Patient-Reported Outcome Measure; DTSL=Dual-Task Screening List; DTQ=Dual-Task Questionnaire

For this translation, the guideline for the process of cross-cultural adaptation of self-report measures was used (Beaton et al., 2000). First, permission from each of the developers of the DT PROMs (Pedullà, Evans, and Strouwen) was obtained. Next, the DT PROMs were translated into the target language. Afterward, the translations were discussed by an expert panel, and finally, the DT PROMs were back-translated into their original language by an independent translator and compared with the original DT PROM. Possible linguistic discrepancies were discussed in a panel.

In Belgium, measurements occurred at the rehabilitation center where the participant was being treated (Noorderhart Rehabilitation & MS and National MS Centre Melsbroek) or at Hasselt University if the participant was not treated at one of these centers or due to their domicile.

According to the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN), each country recruited at least 50 PwMS and 25 age- and sex-matched HC.

A standardized protocol was used in each country to reduce measurement variability. During data collection, participants' demographic data were collected using a form. In this form, participants were asked about their age, gender, mother language, and level of education (in years, starting from elementary school). PwMS were asked additional questions about their type of MS, duration of the disease (in years, starting from the moment of diagnosis), EDSS score, and the use of walking aids.

PwMS attended one session to collect data. During this session, participants received information about the study and filled in the informed consent forms. This session consisted of all measurements, including demographics, performance-based measures, dual-task performance, PROMs for DT difficulties, and other PROMs. The session duration was between 30 and 60 minutes, depending on how quickly participants could complete the questionnaires. HC also attended one session for data collection. The course of this session was identical to the session of PwMS, with the only difference being that HC did not need to fill in MS-specific PROMs.

Measures

Primary outcome measures

Following PROMs were applied to assess perceived DT difficulties in daily life.

Dual-Task Questionnaire (DTQ)

Evans JJ et al. developed the Dual-Task Questionnaire for stroke and traumatic brain injury patients (2009) (**Appendix A**). It consists of 10 items, each with a score between zero (never or not applicable) and four (very often). Higher scores are an indication of more DT problems in daily life (Evans et al., 2009). This PROM has been validated in the Turkisch MS population, concluding that the DTQ is reliable (ICC>0.90) and valid (low-moderate correlations between motor DTCs and cognitive DTCs) to measure DT difficulties (Eldemir et al., 2022).

The percentage of the total score was used in the analysis, which was calculated by adding up the score of all responses of 'Never' to 'Very often' and dividing by that number of items. The responses of 'Not applicable' were left out, which resulted in a more representative total score.

Dual-Task Screening List (DTSL)

The Dual-Task Screening List evaluates dual-task difficulties in daily life in people with Parkinson's disease (**Appendix B**). The questionnaire was developed by Strouwen et al. and consists of 13 items (2014). The possible responses are 'yes', 'no', and 'not applicable.' No research has been conducted on the reliability and validity of this questionnaire (Strouwen et al., 2014). The responses were transformed to calculate the total score. 'Yes' equals one point and 'no' and 'not applicable' both equal zero points. Higher scores are an indication of more DT problems in daily life. The percentage of the total score was used in the analysis, which was calculated by adding up the scores of all responses of 'Yes' or 'No' and dividing by that number of items. This resulted in a more representative total score.

Dual-Task Impact on Daily-life Activities Questionnaire (DIDA-Q)

The Dual-Task Impact on Daily-life Activities Questionnaire was developed in Italy by Pedullà et al. to evaluate the impact of dual-task difficulties on daily life in PwMS (2020) (**Appendix C**). The questionnaire has 19 items with scores between zero (not difficult) and four (extremely difficult). The DIDA-Q has a maximum score of 76 points, higher scores are an indication of more DT problems in daily life. The PROM is successful in discriminating MS from mild-moderate to severe (Pedullà et al., 2020).

Secondary outcome measures

Performance-based measures

Nine-Hole Peg Test (NHPT)

The NHPT was used to assess fine manual dexterity. During the test, participants placed nine pegs into nine holes and then removed those pegs. This must be done as quickly as possible. Participants first performed two consecutive trials on the dominant hand, followed by two consecutive trials on the non-dominant hand.

Based on these trials an average score was calculated for both the dominant and non-dominant hand. Based on these averages a final average score was calculated and this score was used during data analysis. The NHPT has excellent reliability and validity (discriminant, concurrent, and ecological) (Feys et al., 2017).

Timed-25-Foot Walk Test (T25FW)

The T25FW was performed to assess the fastest walking speed. Participants walked as quickly and safely as possible on a flat, obstacle-free 7.62 meters. They performed this task two consecutive times. The time was measured in seconds, and the average of two trials was reported to evaluate the fastest walking speed (Fischer et al., 1999).

The Symbol Digit Modalities Test (SDMT)

The Symbol Digit Modalities Test (SDMT) is a short test to evaluate cognitive processing speed in PwMS. Participants were given the instruction to match as many numbers and symbols as possible in a period of 90 seconds. The total of correct matches was reported. The SDMT has excellent reliability, and good validity (Benedict et al., 2017).

Dual-Task Performance

Single motor task

The single motor task consists of walking for 30 seconds on a 15-meter, free-of-obstacles walkway. Participants were asked to walk as far as possible. The total distance in meters was recorded.

Dual-task (DT)

The single motor task was combined with a semantic word list generation task for the DT. This is a cognitive task for verbal fluency. While walking, the participants needed to say as many words as possible in a particular category (fruits). The total distance in meters was recorded, and the walking speed was calculated. Participants were instructed to perform both tasks at their best possible level. This was to avoid task prioritization.

Dual-task cost (DTC)

The dual-task cost (DTC) is the decline in the performance score of a specific task when this activity is concurrently executed with another task. The following formula was used to calculate the DTC (Veldkamp, Baert, et al., 2019).

$$\left(\frac{single\ task\ score - dual\ task\ score}{single\ task\ score}\right) x\ 100$$

Other Patient-Reported Outcome Measurements

Short Fall Efficacy Scale International (Short FES-I)

The Short Fall Efficacy Scale International (Short FES-I) measures the concern for falling. It consists of seven items. Participants filled in the questionnaire and gave a score on a four-point scale (1 = not at all concerned, 4 = very concerned), where a higher score equals a greater fear of falling (Yardley et al., 2005).

Modified Fatigue Impact Scale (MFIS)

This self-reported questionnaire assesses the effects of physical, cognitive, and psychosocial dimensions of fatigue on daily life functioning in PwMS. The questionnaire has 21 items in total. Scores for the statements ranged between 'never' and 'almost always.' Higher scores are indicative of a more significant influence of fatigue on daily living. The total score of MFIS is a valid and reliable tool to measure fatigue in PwMS (Kos et al., 2003; Kos et al., 2005; Larson, 2013).

Multiple Sclerosis Walking Scale (MSWS-12)

This questionnaire assesses perceived walking difficulties in PwMS. It contains 12 items and uses a five-point rating scale ranging from 'not at all,' 'a little,' 'moderately,' 'quite a bit,' to 'extremely.' The sum of each item is used to generate a total score ranging from 0 to 100. Higher scores mean there is an increased walking impairment. MSWS-12 is a reliable measurement to detect perceived walking difficulties (Mokkink et al., 2016, Hobart et al., 2003).

Fall history

Participants were asked about their fall history, specifically if they had fallen in the last six months. A fall is 'an event where the person unintentionally landed on the ground or a lower level' (Lamb et al., 2005). If they had fallen, they were asked how many times.

Data analysis

All data were analyzed using JMP®, Version 16. SAS Institute Inc., Cary, NC, 1989–2023. Descriptives were tested on normality using Shapiro-Wilk. They are presented using mean and standard deviation for normally distributed data, and median and interquartile range for non-normally distributed data. The level of significance was set at a p-value < 0.05 for all analyses.

Structural validity

One part of validity is structural validity. This measures the degree to which scores of PROMs adequately reflect the dimensionality of the construct to be measured (Mokkink et al., 2010). For each country, 50 PwMS needed to be included resulting in a sample size of > 300. An exploratory factor analysis (EFA) with varimax rotation using baseline data to determine the underlying factorial structure of the scales was used. EFA was chosen because this has not been done previously on structural validity for DTQ and DTSL. It was done for DIDA-Q in the original language it has been developed in (Italian). However, there is no data available on EFA for all the different countries included in this study (Orcan, 2018). It was expected to verify the three-factor structure for DIDA-Q and one-factor structures for DTQ and DTSL as defined in the original validation and development studies (Pedullà et al., 2020; Evans et al., 2009; Strouwen et al., 2014). Kaiser-Meyer-Olkin measure (KMO) and Bartlett's Test of Sphericity (BTS) will be used to assess if EFA is appropriate. KMO should have a value > 0.60 and BTS should be significant with a p-value < 0.05 (Cerny & Kaiser, 1977).

Hypotheses testing for construct validity

Construct validity measures the degree to which the scores of PROMs are consistent with the hypotheses based on the assumption that the PROM validly measures the construct to be measured (Mokkink et al., 2010). It can be divided into convergent and discriminative validity.

For each country, at least 50 PwMS and 25 HC were recruited resulting in a total sample size of > 300. Data analysis was conducted for each country separately and the total sample size.

Convergent validity

To assess convergent validity, hypotheses were investigated for the total sample size of PwMS (n=341) and each country separately ($n_{Belgium}$=51, n_{Chile}=50, n_{Israel}=35, n_{Italy}=101, n_{Spain}=50, n_{Turkey}=54). To calculate convergent validity, the following hypotheses were defined at activity and participation levels according to the ICF:

1. A strong correlation is expected between DTQ, DTSL, and DIDA-Q.
2. A moderate correlation is expected between the DTC and DTQ, DTSL, and DIDA-Q.
3. Moderate to strong correlations are expected between the DTQ, DTSL, DIDA-Q, and other self-reported measures (e.g., FES-I, MFIS, MSWS-12, and fall history).
4. Weak to moderate correlations are expected between the DTQ, DTSL, DIDA-Q, and performance-based measures (e.g., NHPT, T25FW, and SDMT).

All hypotheses were defined on both activity and participation level because DTQ, DTSL, and DIDA-Q all contain items that focus on activity level and items that focus on participation level. The performance-based measures and DTC can be classified under the activity level, but since a correlation is calculated with the DT PROMs, those hypotheses are classified under both activity and participation level.

Correlations were calculated for each country and the total sample size using Spearman's ρ. The correlation coefficients were interpreted as weak (0.10), moderate (0.30), and strong (0.50) (Cohen, 2013). Based on the interpretation of correlation coefficients a percentage of confirmation of the hypotheses was calculated.

Discriminative validity (known-group validity)

Another form of construct validity is discriminative validity. This reflects whether PROMs can differentiate between groups (Rodrigues et al., 2019). To assess discriminative validity following hypotheses were defined on activity and participation level of the ICF:

1. A large degree of difference is expected between the mean scores of HC and PwMS on the DIDA-Q.
2. A large degree of difference is expected between the mean scores of HC and PwMS on the DTSL.
3. A large degree of difference is expected between the mean scores of HC and PwMS on the DTQ.

Again, hypotheses were defined on both activity and participation level since DTQ, DTSL, and DIDA-Q all contain items that are related to the activity level and items that relate more to the participation level.

Firstly, the Mann-Whitney U test was used to assess differences between PwMS and HC in mean total scores of the DT PROMs for the full sample size and the countries individually. A p-value of < 0.05 indicated a statistically significant difference between both groups. The Mann-Whitney U test is a rank-sum test that can be used to compare 2 ordinal, independent groups (**Appendix D**). Next, Cohen's d was used to define the magnitude of the difference between the groups. Using the value of Cohen's d the hypotheses were confirmed or rejected. The degree of difference (Cohen's d) of 0.20 to 0.49 was interpreted as small, 0.50 to 0.79 as medium, and above 0.80 as large (Cohen, 2013). Cut-off values for the DT PROMs were calculated by comparing Receiver-Operating Characteristics (ROC) analysis together with sensitivity and specificity values for these cut-off values. The area under the curve (AUC) was calculated, it ranges from 0 to 1 and higher values reflect better discriminability. Values between 0.7 and 0.8 represented good diagnostic accuracy, and 0.6 and 0.7 sufficient. Sensitivity and specificity indicated the true positive and true negative rates, respectively (Zhou et al., 2011).

Floor and Ceiling effects

Floor and ceiling effects were checked for DTQ, DTSL, and DIDA-Q of the total sample size of PwMS (n=341) by visually inspecting histograms. A cut-off value of 15% was used to determine the existence of a floor and/or ceiling effect, meaning ≥ 15% of the sample size scored the lowest or highest score (Hobart, 2000).

Results

Sample characteristics

In this study, a total of 341 PwMS and 175 HC were included. Every country reached the goal of 50 PwMS and 25 HC, except for Israel with only 35 PwMS at the time of the analysis. Of the PwMS, the median age was 48 years (Q1-Q3 = 38-57). The median EDSS score is 4 (Q1-Q3 = 2-6) and disease severity was divided almost equally (Mild = 48.6%, moderate-severe = 51.32%). Most PwMS presented with Relapse-Remitting MS (71.84%). Except for Italy (for age and educational level), all countries and the total sample showed similar age, sex, and educational level. A detailed overview of sample characteristics is depicted in **Table 3**.

Table 3. Sample Characteristics (HC and PwMS)

	Total Sample PwMS = 341 HC = 175	**Belgium** PwMS = 51 HC = 42	**Chile** PwMS = 50 HC = 25	**Israel** PwMS = 35 HC = 25	**Italy** PwMS = 101 HC = 25	**Spain** PwMS = 50 HC = 31	**Turkey** PwMS = 54 HC = 27
Age							
HC *Median (*Q1-Q3*)*	44 (32-54)	50.5 (43.75-55)	35 (31.5-44.5)	44 (31.5-56)	41 (30.5-50.5)	44 (35-58)	41 (31-52)
PwMS *Median (*Q1-Q3*)*	48 (38-57)	50 (43-60)	35.5 (29.75-43.5)	52 (45-60)	52 (45.5-59)	50.5 (41.74-57.75)	39.50 (32.75-49.5)
Sex							
HC *n (% female)*	112 (64%)	23 (54.76%)	15 (60%)	16 (64%)	19 (76%)	23 (74.19%)	16 (59.26%)
PwMS *n (% female)*	229 (67.16%)	34 (66.67%)	38 (76%)	28 (80%)	68 (67.33%)	29 (58%)	32 (59.26%)
Educational level (Years)							
HC *Median (*Q1-Q3*)*	16 (14-17)	15 (13.75-16)	17 (16-18.5)	15 (12.75-16.5)	17 (13.5-18.5)	15 (13-16)	16 (16-16)
PwMS *Median (*Q1-Q3*)*	15 (12-17)	15 (12-15)	17 (16-17.25)	15 (12-18)	13 (11-16)	14.5 (12-17)	16 (12-16)
EDSS (0-10)							
PwMS *Median (*Q1-Q3*)*	4 (2-6)	4 (2.5-6)	2 (1.38-3)	4.5 (3-6)	4 (2-6)	4.5 (3-6)	3 (1.5-4.63)

Disease severity (PwMS)							
Mildly disabled PwMS (EDSS ≤ 4) *n (%)*	166 (48.68%)	21 (41.18%)	41 (82%)	12 (34.43%)	45 (44.55%)	18 (36%)	29 (53.70%)
Moderate-severe disabled PwMS (EDSS > 4) *n (%)*	175 (51.32%)	30 (58.82%)	9 (18%)	23 (65.71%)	56 (55.45%)	32 (64%)	25 (46.3%)
Disease duration (years)							
PwMs *Median (*Q1-Q3*)*	10 (5-19)	12 (7-20)	5 (2-8.25)	15 (8.5-17)	12 (5-20)	18.5 (7.75-23.25)	10 (5-13.25)
MS Type (PwMS)							
Relapse-Remitting *n (%)*	245 (71.85%)	34 (66/67%)	41 (82%)	30 (85.71%)	66 (65.35%)	30 (60%)	44 (81.48%)
Secondary progressive *n (%)*	58 (17.01%)	9 (17.65%)	3 (6%)	1 (2.86%)	25 (24.75%)	15 (30%)	5 (9.26%)
Primary Progressive *n (%)*	38 (11.14%)	8 (15.69%)	6 (12%)	4 (11.43%)	10 (9.9%)	5 (10%)	5 (9.26%)
Walking aid							
PwMS *n (% Yes)*	117 (34.31%)	28 (54.9%)	4 (8%)	16 (45.71%)	30 (29.7%)	28 (56%)	11 (20.37%)
Fall history (PwMS)							
Fallers *n (%)*	148 (43.4%)	21 (41.18%)	19 (38%)	18 (51.43%)	40 (39.6%)	28 (56%)	22 (40.74%)
Non-fallers *n (%)*	193 (56.6%)	30 (58.82%)	31 (62%)	17 (48.57%)	61 (60.4%)	22 (44%)	32 (59.26%)

Abbreviations: EDSS = Expanded Disability Status Scale; HC = Healthy Controls; MS = Multiple Sclerosis; n = number; PwMS = People with Multiple Sclerosis; Q1-Q3 = Quartile 1-Quartile 3

Descriptives of all primary and secondary outcome measures are depicted in **Table 4**. HC scored lower on the DT PROMs compared to PwMS, meaning HC experienced fewer DT difficulties. The mean DTC of Belgium and Israel is higher for HC compared with PwMS, resulting in worse DT performance. However, for the total sample size, PwMS have higher scores on DTC. HC score better on all performance-based measures.

Table 4. Descriptives of outcome measures

Outcome measure		Total Sample PwMS = 341 HC=175	Belgium PwMS = 51 HC = 42	Chile PwMS = 50 HC = 25	Israel PwMS = 35 HC = 25	Italy PwMS = 101 HC = 25	Spain PwMS = 50 HC = 31	Turkey PwMS = 54 HC = 27
Dual-Task Patient-Reported Outcome measures								
DIDA-Q Median (Q1-Q3)	*HC*	2.00 (0.00-5.00)^	2.00 (0.00-7.00)^	0.00 (0.00-2.00)^	3.00 (0.0-10.0)^	3.00 (1.50-7.00)^	2.00 (0.00-5.00)^	2.00 (0.00-6.00)^
	PwMS	23.00 (10.0-39.0)	24.00 (12.0-40.0)	9.00 (5.0-22.5)	30.00 (10.0-46.0)	24.00 (11.0-38.0)	31.50 (23.8-47.3)	22.50 (7.8-36.0)
DTSL* Median (Q1-Q3)	*HC*	0.00 (0.0-7.69)^	0.00 (0.00-8.65)^	0.00 (0.00-0.00)^	0.00 (0.00-7.69)^	0.00 (0.0-11.5)^	0.00 (0.00-0.00)^	0.00 (0.00-7.69)^
	PwMS	38.46 (15.0-69.0)	53.85 (23.1-76.9)	23.08 (0.0-46.2)	38.46 (23.1-76.9)	38.46 (8.0-61.5)	53.85 (38.5-70.7)	38.46 (7.7-61.5)
DTQ* Median (Q1-Q3)	*HC*	15.00 (5.0-27.5)^	18.75 (11.9-33.1)^	2.50 (0.00-8.75)^	20.00 (12.5-31.3)^	17.50 (15.0-28.6)^	12.50 (5.0-27.5)^	10.00 (0.0-25.0)^
	PwMS	40.00 (22.5-57.5)	45.00 (32.5-70.0)	30.00 (17.5-42.5)	42.50 (22.5-65.0)	34.38 (20.0-52.5)	51.25 (17.8-68.1)	33.75 (22.5-57.5)
Dual-task performance								
DTC Mean (SD)	*HC*	12.13 (9.51)^	10.62 (7.86)	19.35 (6.75)	11.16 (8.88)	14.02 (12.58)	12.11 (9.73)^	6.93 (7.21)^
	PwMS	14.65 (11.22)	10.13 (9.15)	17.32 (10.47)	10.18 (15.16)	17.38 (10.35)	17.88 (11.60)	11.28 (8.82)
Other self-reported measures								
FES-I Median (Q1-Q3)	*PwMS*	13.00 (9.0-17.0)	13.00 (10.0-18.0)	10.00 (7.8-13.0)	16.00 (10.0-20.0)	12.00 (9.0-16.0)	13.50 (10.8-18.0)	13.00 (10.0-17.0)
MFIS Median (Q1-Q3)	*PwMS*	41.00 (24-55)	43.50 (34.8-55.0)	30.00 (21.0-45.0)	43.00 (24.0-58.0)	39.00 (23.5-56.0)	48.00 (28.8-55.0)	43.00 (19.5-55.0)
MSWS-12* Median (Q1-Q3)	*PwMS*	60.00 (35-80)	70.00 (41.7-86.7)	27.50 (21.7-45.4)	75.00 (51.7-90.0)	51.67 (30.8-75.0)	72.50 (61.7-86.7)	58.33 (39.4-79.1)
Fall history Median (Q1-Q3)	*PwMS*	0.00 (0.00-2.00)	0.00 (0.00-1.25)	0.00 (0.00-1.00)	0.00 (0.00-1.00)	0.00 (0.00-1.00)	1.00 (0.00-3.25)	0.00 (0.00-1.00)
Performance-based measures								
NHPT Median (Q1-Q3)	*HC*	18.82 (17.6-20.2)^	18.46 (17.8-20.1)^	17.60 (16.3-18.4)^	No data	19.89 (18.8-21.3)^	19.36 (18.5-21.0)^	18.75 (17.6-20.3)^
	PwMS	24.00 (20.5-30.0)	21.76 (18.7-26.8)	20.25 (18.9-22.8)	26.03 (22.1-31.8)	25.70 (22.2-31.9)	24.61 (22.6-31.3)	23.85 (19.5-30.5)
T25FW Median (Q1-Q3)	*HC*	3.86 (3.39-4.36)^	3.47 (3.10-4.05)^	3.70 (3.42-3.93)^	3.95 (3.53-4.32)^	3.56 (3.36-3.98)^	4.54 (4.22-4.80)^	3.83 (3.39-5.43)^
	PwMS	5.93 (4.62-8.33)	5.66 (4.18-8.43)	4.99 (4.24-5.83)	7.50 (5.1-10.5)	6.19 (5.03-8.47)	6.95 (5.5-10.6)	5.43 (4.45-7.66)
SDMT	*HC*	57.39	59.38	61.76	47.12	65.20	52.94	57.68

Mean (SD)		(12.16)^	(10.21)^	(13.54)^	(8.10)^	(10.12)^	(10.98)^	(12.37)^
	PwMS	47.68 (14.20)	50.92 (12.10)	54.82 (12.50)	37.11 (12.46)	49.17 (15.18)	41.96 (11.18)	48.00 (13.94)

Abbreviations: DIDA-Q = Dual-Task Impact on Daily-Life Activities Questionnaire; DTSL = Dual-Task Screening List; DTC = Dual-Task Cost; DTQ = Dual-Task Questionnaire; FES-I = Falls Efficacy Scale-I; HC = Healthy Controls; MFIS = Modified Fatigue Impact Scale; MSWS-12 = Multiple Sclerosis Walking Scale; n = number; NHPT = Nine Hole Peg Test; PwMS = People with Multiple Sclerosis; Q1-Q3 = Quartile 1-Quartile 3; SD = Standard Deviation; SDMT = Symbol Digit Modalities Test; T25FWT = Timed 25 Foot Walk Test
** Scores of DTQ, DTSL, and MSWS-12 are in %.*
^ Represents a significant difference between HC and PwMS ($p<0.05$).

Structural validity

The sample size was adequate according to the Kaiser-Meyer-Olkin measure (KMO), which has a value of 0.894, 0.892, and 0.958 for DTQ, DTSL, and DIDA-Q respectively. Using Bartlett's Test of Sphericity, it could be concluded that there is at least one common factor ($p < 0.0001$) for all three DT PROMs (**Table 5**). This means there was a high correlation between variables, and an EFA was useful to compress the data. Eigenvalues were used to determine how many factors needed to be extracted. They are visually represented in the Scree Plot (**Figure 1**).

Table 5. KMO and Bartlett's Test (PwMS)

		DTQ	DTSL	DIDA-Q
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.894	0.892	0.958
Bartlett's Test of Sphericity	**Approx. Chi-Square**	1972.234	1438.205	1972.234
	df	45	78	45
	sig.	<0.0001	<0.0001	<0.0001

Abbreviations: Approx. Chi-Square = Approximate Chi Square; df = Degree of Freedom; DIDA-Q = Dual-Task Impact on Daily-life Activities Questionnaire; DTQ = Dual-Task Questionnaire; DTSL = Dual-Task Screening List; Sig. = Significance

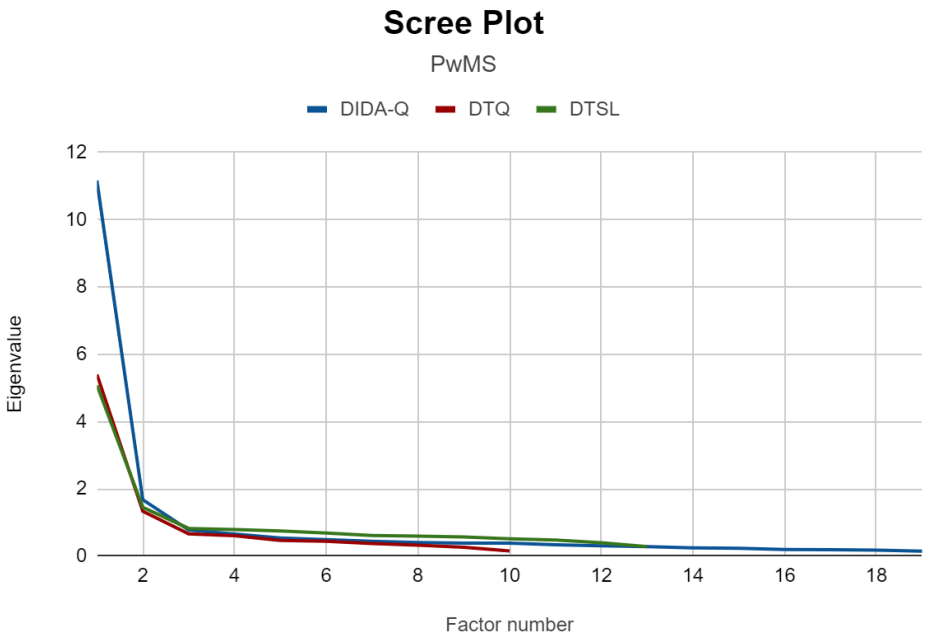

Figure 1. Scree plot of DIDA-Q, DTQ, and DTSL
Abbreviations: DIDA-Q = Dual-Task Impact on Daily-life Activities Questionnaire; DTQ = Dual-Task Questionnaire; DTSL = Dual-Task Screening List

Dual-Task Questionnaire (DTQ)

Two factors with Eigenvalues larger or equal to 1.0 were extracted, explaining 67.273% of the total variance. The item-factor loading needed to be greater than 0.500 in order to attribute the variable to that factor. The results of the varimax rotation final factor solution are represented in **Table 6**. Factors one and two were labeled 'Cognitive-motor' and 'Motor-motor' respectively.

The first factor consisted of seven items on a five-point Likert scale that explained 54% of the variance with factor loadings ranging from 0.570 to 0.775. Factor two consisted of three items on a five-point Likert scale that explained 13% of the variance with factor loadings ranging from 0.593 to 0.926. Item 9 had a factor loading of 0.583 for factor two, reaching the minimum of 0.500. However, there was only a small difference between the factor loading on the first and second factors. This suggested that item 9 correlates to both unobserved variables: 'Cognitive-motor' and 'Motor-motor'.

Table 6. Exploratory Factor Analysis with Varimax rotation (DTQ)

		Factors	
Item	**Description**	**Cognitive-motor**	**Motor-motor**
1	Attention	**0.685**	0.235
2	Talking	**0.775**	0.264
3	Talking	**0.763**	0.259
4	Converse	**0.764**	0.144
5	Listening/talking	**0.580**	0.433
6	Thinking	**0.665**	0.251
7	Spilling cup + walking	0.238	**0.855**
8	Spilling cup + talking	0.208	**0.927**
9	Bump into someone/carrying things	0.467	**0.583**
10	Eating + listening to music	**0.570**	0.195
Eigenvalue		**5.397**	**1.330**
% Variance explained		**53.974**	**13.299**

Abbreviation: DTQ = Dual-Task Questionnaire

Dual-Task Screening List (DTSL)

Two factors with Eigenvalues larger or equal to 1.0 were extracted, explaining 50.199% of the total variance. Items with a factor loading larger than 0.500 were attributed to that factor. Items 6, 7, and 12 did not reach this minimum. Therefore, they did not correlate sufficiently with either of the factors, resulting in a weak link with factor one. Item 2 did reach the minimum of a factor loading of 0.500, but there was only a small difference between the factor loadings on factors one and two. Thus, item 2 could be correlated to both factors, with a larger correlation to factor one. The results of the varimax rotation final factor solution are represented in **Table 7**.

Factors one and two were labeled ‘Motor-motor’ and ‘Cognitive-motor’ respectively. The first factor was comprised of seven items on a 3-point scale that explained 39% of the variance with factor loadings from 0.406 to 0.804. Factor two was comprised of six items on a three-point scale that explained 11% of the variance with factor loadings from 0.500 to 0.580.

Table 7. Exploratory Factor Analysis with Varimax rotation (DTSL)

		Factors	
Item	**Description**	**Motor-motor**	**Cognitive-motor**
1	Walking + talking	0.277	**0.549**
2	Walking + telephoning	**0.527**	0.473
3	Walking + carrying	**0.549**	0.251
4	Walking + carrying	**0.804**	0.163
5	Walking + carrying	**0.778**	0.096
6	Walking + avoiding obstacles	**0.407**	0.366
7	Walking + taking something out pocket	**0.497**	0.457
8	Walking + attention (traffic)	0.286	**0.565**
9	Walking + memory	0.0585	**0.580**
10	Walking + thinking	0.174	**0.551**
11	Walking + finding groceries	0.228	**0.501**
12	Walking + zipping jacket	**0.485**	0.424
13	Walking + navigating (airport)	0.154	**0.531**
Eigenvalue		**5.077**	**1.449**
% Variance explained		**39.053**	**11.146**

Abbreviation: DTSL = Dual-Task Screening List

Dual-Task Impact on Daily-life Activities Questionnaire (DIDA-Q)

Two factors with Eigenvalues larger or equal to 1.0 were extracted, explaining 67.562% of the total variance. The item-factor loading needed to be greater than 0.500 in order to attribute the variable to that factor. All factor loadings reached this minimum. Item 12 had a similar factor loading for factors one and two, 0.587 and 0.516 respectively. Both of these values were above the minimum of 0.500, meaning item 12 correlated to both factors. The results of the varimax rotation final factor solution are represented in **Table 8**.

Factors one and two were labeled ‘Motor-motor’ and ‘Cognitive-motor’ respectively. The first factor was comprised of eleven items on a five-point Likert scale that explained 58% of the variance with factor loadings ranging from 0.586 to 0.783. Factor two was comprised of

eight items on a five-point Likert scale that explained 9% of the variance with factor loadings from 0.640 to 0.775.

Table 8. Exploratory Factor Analysis with Varimax rotation (DIDA-Q)

		Factors	
Item	**Description**	**Motor-motor**	**Cognitive-motor**
1	Walking + carrying a plate	**0.752**	0.266
2	Walking + drinking	**0.764**	0.225
3	Walking + memory	0.212	**0.705**
4	Walking + listening	0.328	**0.776**
5	Walking + planning	0.312	**0.758**
6	Walking + zipping jacket	**0.777**	0.334
7	Talking + walking on curvilinear paths	**0.619**	0.552
8	Walking + reacting to visual input	0.343	**0.653**
9	Talking + maintaining balance	**0.619**	0.461
10	Talking + walking fast	**0.758**	0.357
11	Talking + changing walking direction	**0.715**	0.406
12	Talking + walking	**0.587**	0.516
13	Talking + attention (traffic)	0.314	**0.724**
14	Walking + taking something out pocket	**0.720**	0.428
15	Walking + calculating	0.371	**0.676**
16	Walking + using telephone	**0.784**	0.339
17	Stairs + carrying	**0.688**	0.316
18	Walking + talking	0.399	**0.709**
19	Walking + listening (music)	0.311	**0.640**
Eigenvalue		**11.164**	**1.673**
% Variance explained		**58.575**	**8.807**

Abbreviation: DIDA-Q = Dual-Task Impact on Daily-Life Activities Questionnaire

Convergent validity

Correlation Coefficients were calculated and reported in **Table 9**, which shows a detailed heatmap of Spearman's correlation coefficient (ρ). A strong correlation was found between DTQ, DTSL, and DIDA-Q (hypothesis one) for all countries (ρ 0.613 to 0.882), except for Spain showing a moderate correlation between DTQ and DIDA-Q ($\rho = 0.493$; $p < 0.001$) and DTSL ($\rho = 0.425$; $p < 0.001$). Weak correlations were found for DTQ, DTSL, DIDA-Q, and DTC, however, only three of those were significant ($p < 0.05$). These findings resulted in the rejection of hypothesis two. Correlations between the PROMs (DTQ, DTSL, and DIDA-Q) and the other self-reported measures (hypothesis three) ranged from -0.048 to 0.818, with most outcomes having moderate to strong correlations. Weak correlations were mainly found for fall history across all countries. Weak to moderate correlations (-0.512 to 0.657) were found between the PROMs and the performance-based measures (hypothesis four), resulting in the acceptance of this hypothesis. Correlations between DIDA-Q and T25FW were better than expected for the total sample size and for Italy, Spain, and Turkey separately.

Table 9. Heatmap of correlation coefficients (Spearman's ρ)

	DIDA-Q	DTSL	DTQ	DTC	FES-I	MFIS	MSWS-12	Fall history	NHPT	T25FW	SDMT
All countries											
DIDA-Q	1.000	0.819**	0.737**	0.125*	0.702**	0.671**	0.632**	0.366**	0.467**	0.577**	-0.416**
DTSL	0.819**	1.000	0.702**	0.053	0.639**	0.627**	0.561**	0.369**	0.398**	0.491**	-0.361**
DTQ	0.737**	0.702**	1.000	0.144*	0.551**	0.711**	0.444**	0.270**	0.255**	0.272**	-0.302*
Belgium											
DIDA-Q	1.000	0.853**	0.785**	-0.017	0.658**	0.616**	0.631**	0.367*	0.290*	0.463*	-0.383*
DTSL	0.853**	1.000	0.700**	-0.044	0.593**	0.540**	0.438*	0.284*	0.322*	0.409*	-0.296*
DTQ	0.785**	0.700**	1.000	0.096	0.603**	0.665**	0.401*	0.099	0.030	0.128	-0.287*
Chile											
DIDA-Q	1.000	0.756**	0.749**	-0.089	0.491*	0.665**	0.688**	0.510*	0.451*	0.307*	-0.392*
DTSL	0.756**	1.000	0.613**	-0.144	0.456*	0.713**	0.622**	0.516**	0.310*	0.332*	-0.324*
DTQ	0.749**	0.613**	1.000	-0.020	0.483*	0.745**	0.502*	0.447*	0.433*	0.154	-0.276*
Israel											
DIDA-Q	1.000	0.882**	0.810**	0.207	0.663**	0.763**	0.544*	-0.023	0.355*	0.300	-0.123
DTSL	0.882**	1.000	0.823**	0.296	0.528*	0.725**	0.394*	0.062	0.420*	0.275	-0.154
DTQ	0.810**	0.823**	1.000	0.332	0.542*	0.825**	0.445*	-0.048	0.257	0.181	-0.051
Italy											
DIDA-Q	1.000	0.826**	0.688**	0.122	0.756**	0.724**	0.717**	0.378**	0.465**	0.661**	-0.315*
DTSL	0.826**	1.000	0.661**	0.078	0.741**	0.670**	0.704**	0.509**	0.476**	0.624**	-0.281*
DTQ	0.688**	0.661**	1.000	0.192	0.538**	0.664**	0.453**	0.355*	0.356*	0.385**	-0.208*
Spain											
DIDA-Q	1.000	0.662**	0.493*	0.458*	0.529*	0.506*	0.567**	0.300*	0.539**	0.583**	-0.336*
DTSL	0.662**	1.000	0.425*	0.194	0.292*	0.409*	0.241	0.228	0.242	0.303*	-0.213
DTQ	0.493*	0.425*	1.000	0.333*	0.139	0.698**	0.075	0.244	0.095	0.096	-0.290*
Turkey											
DIDA-Q	1.000	0.867**	0.797**	0.195	0.818**	0.645**	0.202	0.408*	0.334*	0.657**	-0.475*
DTSL	0.867**	1.000	0.772**	0.236	0.775**	0.575**	0.327*	0.301*	0.357*	0.649**	-0.521**
DTQ	0.797**	0.772**	1.000	0.254	0.744**	0.728**	0.229	0.173	0.200	0.365*	-0.392*

Abbreviations: *DIDA-Q = Dual-Task Impact on Daily-Life Activities Questionnaire; DTSL = Dual-Task Screening List; DTC = Dual-Task Cost; DTQ = Dual-Task Questionnaire; FES-I = Falls Efficacy Scale-I; MFIS = Modified Fatigue Impact Scale; MSWS-12 = Multiple Sclerosis Walking Scale; NHPT = Nine Hole Peg Test; SDMT = Symbol Digit Modalities Test; T25FWT = Timed 25 Foot Walk Test;*
*Spearman's ρ with a p-value < 0.05
** Spearman's ρ with a p-value < 0.001
0.10: Weak correlation
0.30: Moderate correlation
0.50: Strong correlation

The four hypotheses above were divided into 189 smaller hypotheses in order to find correlations between the DT PROMs and other outcome measures for each country. Based on these correlations, 154 out of 189 (81.84%) hypotheses were supported. Percentages of confirmation were calculated for each country separately as well. A comprehensive summary of which hypotheses are confirmed can be found in **Table 10**.

Table 10. Confirmation of hypotheses about correlations between DTQ, DTSL, DIDA-Q, and other measurements

	Confirmed (yes/no)						
	All countries	Belgium	Chile	Israel	Italy	Spain	Turkey
Hypotheses about correlations between DT PROMs							
1. A strong correlation is expected between DTQ and DTSL.	Yes	Yes	Yes	Yes	Yes	No	Yes
2. A strong correlation is expected between DTSL and DIDA-Q.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3. A strong correlation is expected between DIDA-Q and DTQ.	Yes	Yes	Yes	Yes	Yes	No	Yes

Hypotheses about correlations between DT PROMs and DTC							
4. A moderate correlation is expected between the DTC and DTQ.	No	No	No	Yes	No	Yes	No
5. A moderate correlation is expected between the DTC and DTSL.	No	No	No	No	No	No	No
6. A moderate correlation is expected between the DTC and DIDA-Q.	No	No	No	No	No	Yes	No
Hypotheses about correlations between DT PROMs and other self-reported measures							
7. Moderate to strong correlations are expected between the DTQ and FES-I.	Yes	Yes	Yes	Yes	Yes	No	Yes
8. Moderate to strong correlations are expected between the DTQ and MFIS.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
9. Moderate to strong correlations are expected between the DTQ and MWSW-12.	Yes	Yes	Yes	Yes	Yes	No	No
10. Moderate to strong correlations are expected between the DTQ and fall history.	No	No	Yes	No	Yes	No	No
11. Moderate to strong correlations are expected between the DTSL and FES-I.	Yes	Yes	Yes	Yes	Yes	No	Yes
12. Moderate to strong correlations are expected between the DTSL and MFIS.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
13. Moderate to strong correlations are expected between the DTSL and MWSW-12.	Yes	Yes	Yes	Yes	Yes	No	Yes
14. Moderate to strong correlations are expected between the DTSL and fall history.	Yes	No	Yes	No	Yes	No	Yes
15. Moderate to strong correlations are expected between the DIDA-Q and FES-I.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
16. Moderate to strong correlations are expected between the DIDA-Q and MFIS.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17. Moderate to strong correlations are expected between the DIDA-Q and MWSW-12.	Yes	Yes	Yes	Yes	Yes	Yes	No
18. Moderate to strong correlations are expected between the DIDA-Q and fall history.	Yes	Yes	Yes	No	Yes	Yes	Yes
Hypotheses about correlations between DT PROMs and performance-based measures							
19. Weak to moderate correlations are expected between the DTQ and NHPT.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
20. Weak to moderate correlations are expected between the DTQ and T25FW.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
21. Weak to moderate correlations are expected between the DTQ and SDMT.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
22. Weak to moderate correlations are expected between the DTSL and NHPT.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
23. Weak to moderate correlations are expected between the DTSL and T25FW.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
24. Weak to moderate correlations are expected between the DTSL and SDMT.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
25. Weak to moderate correlations are expected between the DIDA-Q and NHPT.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
26. Weak to moderate correlations are expected between the DIDA-Q and T25FW.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
27. Weak to moderate correlations are expected between the DIDA-Q and SDMT.	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All countries: 154/189 hypotheses confirmed = 81.84%
Belgium: 22/27 hypotheses confirmed = 81.48%
Chile: 24/27 hypotheses confirmed = 88.89%
Israel: 22/27 hypotheses confirmed = 81.48%
Italy: 24/27 hypotheses confirmed = 88.89%
Spain: 18/27 hypotheses confirmed = 66.67%
Turkey: 21/27 hypotheses confirmed = 77.78%

Abbreviations: DIDA-Q = Dual-Task Impact on Daily-Life Activities Questionnaire; DTC = Dual-Task Cost; DT PROMs = Dual-Task Patient-Reported Outcome Measures; DTSL = Dual-Task Screening List; DTQ = Dual-Task Questionnaire; FES-I = Falls Efficacy Scale-I; MFIS = Modified Fatigue Impact Scale; MSWS-12 = Multiple Sclerosis Walking Scale; NHPT = Nine Hole Peg Test; SDMT = Symbol Digit Modalities Test; T25FWT = Timed 25 Foot Walk Test
Green: Higher correlation than expected
Red: Hypothesis not confirmed

Discriminative validity

For discriminative validity 21 out of 21 (100.00%) hypotheses were confirmed. For the DIDA-Q (hypothesis one), Cohen's d ranged from 1.296 to 2.725. With the largest degree of difference in Spain (Cohen's d = 2.725), and the smallest in Chile (Cohen's d = 1.296). For the DTSL (hypothesis two), a large degree of difference between HC and PwMS was also expected. Cohen's d ranged from 1.283 to 2.706, which confirmed the second hypothesis. The hypothesis for the DTQ stated a large degree of difference was expected between the mean scores of HC and PwMS (hypothesis 3). Here, Cohen's d ranged from 0.946 to 1.887, resulting in this hypothesis being confirmed as well. More detailed information about the hypotheses for discriminative validity for all countries can be found in **Table 11**.

Table 11. Hypotheses testing for discriminative validity for DIDA-Q, DTSL, and DTQ

Country	Expected effect size	Cohen's d	P-value	Confirmed
Hypothesis 1: A large degree of difference is expected between the mean scores of HC and PwMS on the DIDA-Q.				
All countries	≥ 0.80	1.668	< 0.0001	Yes
Belgium	≥ 0.80	1.723	< 0.0001	Yes
Chile	≥ 0.80	1.296	< 0.0001	Yes
Israel	≥ 0.80	1.613	< 0.0001	Yes
Italy	≥ 0.80	1.604	< 0.0001	Yes
Spain	≥ 0.80	2.725	< 0.0001	Yes
Turkey	≥ 0.80	1.661	< 0.0001	Yes
Hypothesis 2: A large degree of difference is expected between the mean scores of HC and PwMS on the DTSL.				
All countries	≥ 0.80	1.643	< 0.0001	Yes
Belgium	≥ 0.80	1.700	< 0.0001	Yes
Chile	≥ 0.80	1.283	< 0.0001	Yes
Israel	≥ 0.80	1.763	< 0.0001	Yes
Italy	≥ 0.80	1.580	< 0.0001	Yes
Spain	≥ 0.80	2.706	< 0.0001	Yes
Turkey	≥ 0.80	1.537	< 0.0001	Yes

Hypothesis 3: A large degree of difference is expected between the mean scores of HC and PwMS on the DTQ.				
All countries	≥ 0.80	1.209	< 0.0001	Yes
Belgium	≥ 0.80	1.244	< 0.0001	Yes
Chile	≥ 0.80	1.590	< 0.0001	Yes
Israel	≥ 0.80	1.168	0.0003	Yes
Italy	≥ 0.80	0.946	0.0004	Yes
Spain	≥ 0.80	1.887	< 0.0001	Yes
Turkey	≥ 0.80	1.321	< 0.0001	Yes
Total: 21/21 hypotheses confirmed = 100.00%				

Abbreviations: DIDA-Q = Dual-Task Impact on Daily-living Activities Questionnaire; DTQ = Dual-Task Questionnaire; DTSL = Dual-Task Screening List; HC = Healthy Controls; PwMS = Persons with Multiple Sclerosis

Cut-off scores to determinate between PwMS and HC were calculated for all the DT PROMs. The cut-off scores to successfully discriminate were 8.0 for DIDA-Q, 23.09% for DTSL, and 32.50% for DTQ. AUC values were calculated as well and all of them represented good diagnostic accuracy (> 0.700). An overview of the cut-off scores can be found in **Table 12** together with their corresponding values for sensitivity, specificity, and the AUC.

Table 12. Cut-off values for DIDA-Q, DTSL, and DTQ (for all countries together)

PROM	Cut-off score (ROC analysis)*	Sensitivity for cut-off score (%)	Specificity for cut-off score (%)	AUC
DIDA-Q	8.00	80.88	85.55	0.901
DTSL	23.08	71.47	89.71	0.864
DTQ	32.50	62.65	84.00	0.800

*Abbreviations: AUC = Area Under the Curve; DIDA-Q = Dual-Task Impact on Daily-living Activities Questionnaire; DTQ = Dual-Task Questionnaire; DTSL = Dual-Task Screening List; PROM = Patient-Reported Outcome Measure; ROC = Receiver Operating Characteristics; * Cut-off scores for DTSL and DTQ are in %*

Floor and Ceiling Effects

Eight (2.35%) and seven (2.05%) PwMS achieved the lowest score of 0 on DIDA-Q and DTQ respectively, while only one (0.29%) participant achieved the highest score on DIDA-Q and DTQ. This resulted in no floor or ceiling effects detected for DIDA-Q and DTQ. 15.84% of PwMS achieved the lowest score on DTSL and 3.52% achieved the highest score (100.00%), therefore, a floor effect was detected for the DTSL.

Discussion

The aim of this study was to investigate the structural and construct validity of PROMs and to compare the perceived difficulties of DT in daily life between PwMS and HC. To our knowledge, this was the first cross-cultural multicenter study investigating the validity of DIDA-Q, DTQ, and DTSL. The findings of this study showed that DIDA-Q, DTQ, and DTSL are valid questionnaires to explore DT difficulties in daily life in PwMS. They could discriminate between PwMS and HC and showed moderate-to-strong correlation with other self-reported tools and low-to-moderate correlations with performance-based measures.

DIDA-Q showed three subscales in a previous study (Pedullà et al., 2020), consisting of an upper-limb ability (5 items), balance and mobility (6 items), and cognitive (8 items) subscale. Our analysis did not confirm this and found two subscales: 'Cognitive-motor' and 'Motor-motor'. Items that were allocated to the subscale of upper-limb ability in the study of Pedullà et al. (2020) were allocated to the motor subscale of the EFA of this study. This may be due to all items of the subscale 'upper limb' also consisting of walking tasks, therefore also fitting in the 'Motor-motor' subscale. For DTQ and DTSL, it was expected to find only one factor as the validation and development studies (Evans et al., 2009; Eldemir et al., 2022b; Strouwen et al., 2014) report no subscales, but in this study, two subscales were found for both DTQ and DTSL: 'Cognitive-motor' and 'Motor-motor'. The 'Cognitive-motor' subscales consist of 7 items, 6 items, and 8 items for DTQ, DTSL, and DIDA-Q respectively. Items primarily include walking while performing a cognitive DT (e.g. Walking and navigating traffic, walking and having a conversation with someone). The 'Motor-motor' subscales include primarily walking in combination with other motor tasks, thus resulting in motor-motor DTs. The 'Motor-motor' subscales consist of 3 items, 7 items, and 11 items for DTQ, DTSL, and DIDA-Q respectively. This means both subscales are appropriate for measurement for all three PROMs, as each subscale needs to consist of a minimum of three items according to Comrey and Lee (2013).

For convergent validity, 189 hypotheses were formulated, of which 81.84% could be supported. This suggests convergent validity for the DIDA-Q, DTSL, and DTQ is good. The low correlation of DIDA-Q, DTSL, and DTQ with DTC is an exception. This can be explained by the

study by Eldemir et al. (2022) concluding that DTC is less reliable than absolute dual-task performance. Additionally, Shiratzky et al. (2020) find that absolute dual-task walking in a lab was similar to typical daily life walking, hence it may reflect daily life conditions better than DTC. Another explanation might be that DT PROMs measure a wider range of DTs that individuals may encounter daily compared to one DT condition used to calculate the DTC in our study. DTC is also a lab-based measurement, while PROMs evaluate DT difficulties in daily life.

As expected, strong correlations were found among DT PROMs. It supports that all tools measure the same construct. Moderate to high correlations between DT PROMs and self-reported measurements were reported (fatigue and mobility) as anticipated. Our findings on fatigue were in line with the findings of Pedullà et al. (2020), they found a strong positive correlation between DIDA-Q and fatigue assessed by MFIS. Since subjective fatigue is a very common symptom in PwMS and it affects daily life activities, it is not surprising to find a strong correlation with DT difficulties. DIDA-Q showed the strongest correlation with MSWS-12 in the total sample and each country separately. This may have resulted from the DIDA-Q containing more items that included walking and balance compared with DTQ and DTSL, thus relating more to the MSWS-12 which measures walking difficulties in PwMS.

A weak to moderate correlation was found between DT PROMs and cognitive processing speed evaluated by SDMT, in line with the findings of Pedullà et al. (2020). Since divided attention and concentration are needed for DT activities, the findings support our hypothesis (Benedict et al., 2017). Fall risk has been moderately to highly correlated with DTC (Rooney et al., 2020); however in this study, only low to moderate correlations were found between the DT PROMs and fall history.

As expected for discriminative validity, large differences (Cohen's $d \geq 0.80$) were found for the difference of mean scores between HC and PwMS for all the DT PROMs. This indicates 100.00% of hypotheses were supported. Therefore, it can be concluded that DIDA-Q, DTSL, and DTQ have good discriminative validity for HC and PwMS. There is no study investigating

the discriminative validity of DTSL and DTQ. However, for DIDA-Q, Pedullà et al. (2020) showed discriminative validity between PwMS with mild disability (EDSS < 4.0) and moderate-to-severe disability (EDSS ≥ 4.0). Learmonth et al. (2017) reported minimal non-significant differences in DTC between PwMS and HC. Our findings suggest that PwMS report more difficulty with DT activities, even though it is not found in performance-based tests.

No previous studies have calculated cut-off scores for the DT PROMs. Based on our findings, the recommended cut-off score is 8/23 for DIDA-Q, 23/100 for DTSL, and 32/100 for DTQ. These cut-off values can be used in clinical practice to discriminate between PwMS and HC.

No floor or ceiling effects were detected for DIDA-Q and DTQ, however, DTSL did show a floor effect with 15.84% of PwMS achieving the lowest score (0). This may be due to the Three-point scale ('Yes', 'No', and 'Not Applicable'), which can result in less diverse scoring answers. Therefore, PwMS could often achieve a score of 0.

Comparing all DT PROMs, DIDA-Q shows the most promising results for structural, convergent, and discriminative validity. For structural validity, the two included factors 'Motor-motor' and 'Cognitive-motor' explained the highest percentage (67.562%) of total variance for DIDA-Q, with DTQ following closely with 67.273%. Factor loadings were > 0.500 for all items of both DIDA-Q and DTQ. DIDA-Q reported the strongest correlations for all secondary outcome measures, resulting in a better convergent validity. DTQ showed the weakest correlations with secondary outcome measures comparing all DT PROMs. DIDA-Q and DTSL also correlate higher with each other compared to the correlations of DTQ with DIDA-Q and DTSL. Discriminative validity is excellent for all DT PROMs, with DIDA-Q showing the highest discriminative ability.

Strengths and limitations

The total sample was large (n=341) as for most data analyses a total of 100 participants results in a very good quality according to COSMIN (Mokkink et al., 2019). However, not every country reached the goal of including 50 PwMS and 25 HC at the time of analysis due to insufficient time management. The smaller sample size of PwMS (n=35) for Israel resulted in a weaker quality of the analysis, however adequate to very good quality was obtained in all other countries. To evaluate the discriminative validity of the DT PROMs, a sample size of 100 PwMS for each country was needed in order to have results with very good quality (Mokkink et al., 2019). This criterium was not met for almost all countries (except for Italy with n=101), but adequate quality was reached in almost all countries. The criterium for very good quality was met when combining countries resulting in a total sample size of 341 PwMS.

Due to practical reasons, six researchers participated in data collection. This could have resulted in differences in measurement/instructions and therefore have an impact on the data. To reduce this impact, a standardized instruction paper was written out, which everyone had to follow. Researchers also taught each other how to measure, to minimize differences.

Performance bias could have occurred as a result of assessors knowing to which group participants belonged. Therefore they could give more encouragement to HC in order to achieve better results. This risk was, however, minimized by using standardized test instructions for both PwMS and HC. Selection bias could also have occurred as HC were age- and sex-matched to PwMS, thus they could be handpicked to achieve better results in data analysis. In order to combat this type of bias, the scores of primary and secondary outcome measures of HC were not viewed when matching HC to PwMS.

Conclusion

In conclusion, EFA showed two factors ('Motor-motor' and 'Cognitive-motor') for each of the PROMs (DIDA-Q, DTSL, and DTQ). All subscales consist of at least 3 items resulting in good structural validity. Convergent validity for DIDA-Q, DTSL, and DTQ is in line with the hypotheses on activity and participation level of the ICF, as 81.84% of all hypotheses were supported. Discriminative validity for DIDA-Q, DTSL, and DTQ is good. The PROMs can successfully differentiate between HC and PwMS. Several differences were notable between the countries, therefore further research about cross-cultural validity is necessary.

References

Abasıyanık, Z., Veldkamp, R., Fostier, A., Van Goubergen, C., Kalron, A., & Feys, P. (2022). Patient-Reported Outcome Measures for Assessing Dual-Task Performance in Daily Life: A Review of Current Instruments, Use, and Measurement Properties. *International Journal of Environmental Research and Public Health*, *19*(22), 15029. https://doi.org/10.3390/ijerph192215029

Alroughani, R., & Yamout, B. (2018). Multiple Sclerosis. *Seminars in Neurology*, *38*(02), 212–225. https://doi.org/10.1055/s-0038-1649502

Al-Yahya, E., Dawes, H., Smith, L., Dennis, A., Howells, K., & Cockburn, J. (2011). Cognitive motor interference while walking: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 715–728. https://doi.org/10.1016/j.neubiorev.2010.08.008

Argento, O., Spanò, B., Pisani, V., Incerti, C. C., Bozzali, M., Foti, C., Caltagirone, C., & Nocentini, U. (2021). Dual-Task Performance in Multiple Sclerosis' Patients: Cerebellum Matters? *Archives of Clinical Neuropsychology*, *36*(4), 517–526. https://doi.org/10.1093/arclin/acaa089

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Benedict, R. H., DeLuca, J., Phillips, G., LaRocca, N. G., Hudson, L. D., & Rudick, R. A. (2017). Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Multiple Sclerosis Journal*, *23*(5), 721–733. https://doi.org/10.1177/1352458517690821

Cerny, B. A., & Kaiser, H. F. (1977). A Study Of A Measure Of Sampling Adequacy For Factor-Analytic Correlation Matrices. *Multivariate Behavioral Research*, *12*(1), 43–47. https://doi.org/10.1207/s15327906mbr1201_3

Chen, A. P., Kirkland, M. C., Wadden, K. P., Wallack, E. M., & Ploughman, M. (2020). Reliability of gait and dual-task measures in multiple sclerosis. *Gait & Posture*, *78*, 19–25. https://doi.org/10.1016/j.gaitpost.2020.03.004

Cinnera, A. M., Bisirri, A., Leone, E., Morone, G., & Gaeta, A. (2021). Effect of dual-task training on balance in patients with multiple sclerosis: A systematic review and meta-analysis. *Clinical Rehabilitation*, *35*(10), 1399–1412. https://doi.org/10.1177/02692155211010372

Cohen, J. (2013). Statistical Power Analysis for the Behavioral Sciences. In *Routledge eBooks*. https://doi.org/10.4324/9780203771587

Comrey, A. L., & Lee, H. (2013). A First Course in Factor Analysis. In *Psychology Press eBooks*. https://doi.org/10.4324/9781315827506

Eldemir, S., Ozkul, C., Eldemir, K., Saygili, F., Guclu-Gunduz, A., & Irkec, C. (2022a). The Reliability and Validity of the Turkish Version of the Dual-Task Questionnaire in Patients with Multiple Sclerosis. *Multiple Sclerosis and Related Disorders*, *64*, 103942. https://doi.org/10.1016/j.msard.2022.103942

Evans, J., Greenfield, E., Wilson, B. A., & Bateman, A. (2009). Walking and talking therapy: Improving cognitive–motor dual-tasking in neurological illness. *Journal of The International Neuropsychological Society*, *15*(1), 112–120. https://doi.org/10.1017/s1355617708090152

Feys, P., Lamers, I., Francis, G. A., Benedict, R. H., Phillips, G., LaRocca, N. G., Hudson, L. D., & Rudick, R. A. (2017). The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Multiple Sclerosis Journal*, *23*(5), 711–720. https://doi.org/10.1177/1352458517690824

Fischer, J., Rudick, R. A., Cutter, G., & Reingold, S. C. (1999). The Multiple Sclerosis Functional Composite measure (MSFC): an integrated approach to MS clinical outcome assessment. *Multiple Sclerosis Journal*, *5*(4), 244–250. https://doi.org/10.1177/135245859900500409

Hamilton, F., Rochester, L., Paul, L., Rafferty, D., O'Leary, C. P., & Evans, J. (2009). Walking and talking: an investigation of cognitive—motor dual tasking in multiple sclerosis. *Multiple Sclerosis Journal*, *15*(10), 1215–1227. https://doi.org/10.1177/1352458509106712

Hobart, J. (2000). Kurtzke scales revisited: the application of psychometric methods to clinical intuition. *Brain*, *123*(5), 1027–1040. https://doi.org/10.1093/brain/123.5.1027

Hobart, J., Riazi, A., Lamping, D. L., Fitzpatrick, R., & Thompson, A. S. (2003). Measuring the impact of MS on walking ability: The 12-Item MS Walking Scale (MSWS-12). *Neurology*, *60*(1), 31–36. https://doi.org/10.1212/wnl.60.1.31

Kos, D., Kerckhofs, E., Carrea, I., Verza, R., Ramos, M., & Jansa, J. (2005). Evaluation of the Modified Fatigue Impact Scale in four different European countries. *Multiple Sclerosis Journal*, *11*(1), 76–80. https://doi.org/10.1191/1352458505ms1117oa

Lamb, S. E., Jørstad-Stein, E. C., Hauer, K., & Becker, C. (2005). Development of a Common Outcome Data Set for Fall Injury Prevention Trials: The Prevention of Falls Network Europe Consensus. *Journal of the American Geriatrics Society*, *53*(9), 1618–1622. https://doi.org/10.1111/j.1532-5415.2005.53455.x

Larson, R. D. (2013). Psychometric Properties of the Modified Fatigue Impact Scale. *International journal of MS care*, *15*(1), 15–20. https://doi.org/10.7224/1537-2073.2012-019

Learmonth, Y. C., Ensari, I., & Motl, R. W. (2017). Cognitive Motor Interference in Multiple Sclerosis: Insights From a Systematic Quantitative Review. *Archives of Physical Medicine and Rehabilitation*, *98*(6), 1229–1240. https://doi.org/10.1016/j.apmr.2016.07.018

MacPherson, S. E. (2018). Definition: Dual-tasking and multitasking. *Cortex*, *106*, 313–314. https://doi.org/10.1016/j.cortex.2018.06.009

McIsaac, T. L., Fritz, N., Quinn, L., & Muratori, L. M. (2018). Cognitive-Motor Interference in Neurodegenerative Disease: A Narrative Review and Implications for Clinical Management. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02061

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J. A., Stratford, P. W., Knol, D. L., Bouter, L. M., & De Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Mokkink, L. B., Galindo-Garre, F., & Uitdehaag, B. M. J. (2016). Evaluation of the Multiple Sclerosis Walking Scale-12 (MSWS-12) in a Dutch sample: Application of item response theory. *Multiple Sclerosis Journal*. https://doi.org/10.1177/1352458516630821

Mokkink, B., AC Prinsen, C., L Patrick, D., Alonso, J., M Bouter, L., CW de Vet, H., & B Terwee, C. (2019). COSMIN Study Design checklist for Patient-reported outcome measurement instruments : Version July 2019. *COSMIN*.

Orcan, F. (2018). Exploratory and Confirmatory Factor Analysis: Which One to Use First? *Journal of Measurement and Evaluation in Education and Psychology*, 414–421. https://doi.org/10.21031/epod.394323

Pedullà, L., Tacchino, A., Podda, J., Bragadin, M. M., Bonzano, L., Battaglia, M., Bove, M., Brichetto, G., & Ponzio, M. (2020). The patients' perspective on the perceived difficulties of dual-tasking: development and validation of the Dual-task Impact on Daily-living Activities Questionnaire (DIDA-Q). *Multiple sclerosis and related disorders*, *46*, 102601. https://doi.org/10.1016/j.msard.2020.102601

Pollard, B., Dixon, D., Dieppe, P., & Johnston, M. (2009). Measuring the ICF components of impairment, activity limitation and participation restriction: an item analysis using classical test theory and item response theory. *Health and Quality of Life Outcomes*, *7*(1). https://doi.org/10.1186/1477-7525-7-41

Rodrigues, I., Adachi, J. D., Beattie, K. A., Lau, A. N., & MacDermid, J. C. (2019). Determining known-group validity and test-retest reliability in the PEQ (personalized exercise questionnaire). *BMC Musculoskeletal Disorders*, *20*(1). https://doi.org/10.1186/s12891-019-2761-3

Rooney, S., Ozkul, C., & Paul, L. (2020a). Correlates of dual-task performance in people with multiple sclerosis: A systematic review. *Gait & Posture*, *81*, 172–182. https://doi.org/10.1016/j.gaitpost.2020.07.069

Shema-Shiratzky, S., Hillel, I., Mirelman, A., Regev, K., Hsieh, K. L., Karni, A., Devos, H., Sosnoff, J. J., & Hausdorff, J. M. (2020). A wearable sensor identifies alterations in community ambulation in multiple sclerosis: contributors to real-world gait quality and physical activity. *Journal of Neurology*, *267*(7), 1912–1921. https://doi.org/10.1007/s00415-020-09759-7

Strouwen, C., Molenaar, E., Keus, S., Münks, L., Munneke, M., Vandenberghe, W., Bloem, B. R., & Nieuwboer, A. (2014). Protocol for a randomized comparison of integrated versus consecutive dual task practice in Parkinson's disease: the DUALITY trial. *BMC Neurology*, *14*(1). https://doi.org/10.1186/1471-2377-14-61

Van Schependom, J., D'hooghe, M. B., Cleynhens, K., D'hooge, M., Haelewyck, M., Keyser, D. J., & Nagels, G. (2014). The Symbol Digit Modalities Test as sentinel test for cognitive impairment in multiple sclerosis. *European Journal of Neurology*, *21*(9), 1219-e72. https://doi.org/10.1111/ene.12463

Veldkamp, R., Baert, I., Kalron, A., Tacchino, A., D'hooge, M., Vanzeir, E., Van Geel, F., Raats, J., Goetschalckx, M., Brichetto, G., Shalmoni, N., Hellinckx, P., De Weerdt, N., De Wilde, D., & Feys, P. (2019). Structured Cognitive-Motor Dual Task Training Compared to Single Mobility Training in Persons with Multiple

Sclerosis, a Multicenter RCT. *Journal of Clinical Medicine*, *8*(12), 2177. https://doi.org/10.3390/jcm8122177

Veldkamp, R., Romberg, A., Hämäläinen, P., Giffroy, X., Moumdjian, L., Leone, C., Feys, P., & Baert, I. (2019). Test-Retest Reliability of Cognitive-Motor Interference Assessments in Walking With Various Task Complexities in Persons With Multiple Sclerosis. *Neurorehabilitation and Neural Repair*, *33*(8), 623–634. https://doi.org/10.1177/1545968319856897

Yardley, L., Beyer, N., Hauer, K., Kempen, G. I. J. M., Piot-Ziegler, C., & Todd, C. (2005). Development and initial validation of the Falls Efficacy Scale-International (FES-I). *Age and Ageing*, *34*(6), 614–619. https://doi.org/10.1093/ageing/afi196

Zhou, X. K., Obuchowski, N. A., & McClish, D. K. (2011). Measures of Diagnostic Accuracy. In *John Wiley & Sons, Inc. eBooks* (pp. 13–55). https://doi.org/10.1002/9780470906514.ch2

Appendices

Appendix A
Dual-Tasking Questionnaire (DTQ)

Instructies: De volgende vragen gaan over problemen die iedereen van tijd tot tijd ervaart, maar waarvan sommige vaker gebeuren dan andere. We willen weten hoe vaak deze dingen bij u zijn gebeurd in de afgelopen weken. Er zijn 5 opties, gaande van 'zeer vaak' tot 'nooit', of 'niet van toepassing'. Gelieve het gepaste vakje aan te vinken.

Heeft u een van deze moeilijkheden…	Zeer vaak	Eerder vaak	Occasioneel	Zeer zelden	Nooit	NVT
	4	3	2	1	0	0
1. Aandacht geven aan meer dan één ding tegelijk?						
2. Nood hebben aan een activiteit te stoppen om te praten?						
3. Praten wanneer je een andere activiteit aan het doen bent?						
4. Volgen of deelnemen in een gesprek waar verscheidene personen tegelijk aan het spreken zijn?						
5. Verslechteren van het stappen wanneer je aan het spreken of luisteren bent naar iemand?						
6. Verdiept zijn in je eigen gedachten, dus zonder op te merken wat er rondom u gebeurd?						
7. Een drankje morsen tijdens het dragen ervan.						
8. Meer morsen indien je op hetzelfde moment spreekt.						
9. Tegen mensen aanbotsen of dingen laten vallen indien je tevens iets anders doet?						
10. Moeilijkheden om te eten en televisie te kijken of te luisteren naar de radio tegelijkertijd.						
Subtotalen						

Totaal:

Appendix B

Dual-Task Screening List (DTSL)

Instructies: De volgende vragenlijst gaat na of u problemen ondervindt bij het uitvoeren van een dubbeltaak. Hieronder vindt u enkele situaties waarin er dubbeltaken gebruikt worden. Bedoeling is dat u 'ja' aanduidt indien u moeilijkheden ondervindt bij:
- het stappen: gaat het stappen moeilijker wanneer u tegelijkertijd een taak moet uitvoeren vergeleken met enkel stappen zonder bijkomende taak? Met moeilijker bedoelen we bv. of u dan trager stapt, kleinere passen neemt, enzovoort. EN/OF
- de dubbeltaak: merkt u dat u de taak die u tegelijkertijd uitvoert moeilijker gaat? Met moeilijker bedoelen we vb plots stoppen met een gesprek omdat u moeite heeft om de woorden te vinden, fouten maken die men anders niet zou maken, dingen laten vallen, etc.

Als een situatie niet voor u van toepassing is, duidt u het vakje met 'niet van toepassing' aan. Indien er echter bepaalde situaties zijn die u vermijdt in het dagelijks leven omdat u weet dat deze u moeilijkheden bezorgen bij het stappen of bij het goed uitvoeren van de taak, gelieve dan het vakje 'ja' aan te duiden voor deze situaties.

Heeft u een van deze moeilijkheden…	JA	NEE	NVT
1. Heeft u moeite met de combinatie stappen en praten?			
2. Heeft u moeite met de combinatie stappen en telefoneren?			
3. Heeft u moeite met de combinatie van stappen en een boodschappentas dragen?			
4. Heeft u moeite met de combinatie van stappen en het dragen van een bord dat gevuld is met eten?			
5. Heeft u moeite met de combinatie van stappen en het dragen van een gevuld glas/kopje?			
6. Heeft u moeite met de combinatie van stappen en het ontwijken van obstakels (op de grond)?			
7. Heeft u moeite met de combinatie van stappen en het tevoorschijn halen van iets uit je zakken (vb een busabonnement, zakdoek, geld, gsm)?			
8. Heeft u moeite met de combinatie van stappen en het richten van je aandacht op iets anders (vb het verkeer)?			
9. Heeft u moeite met de combinatie van stappen en het onthouden van iets belangrijks (vb een telefoonnummer, een adres)?			
10. Heeft u moeite met de combinatie van stappen en het denken aan andere dingen?			
11. Heeft u moeite om tijdens het winkelen naar koopwaren te zoeken?			
12. Heeft u moeite met de combinatie van stappen en het sluiten van de rits van uw jas?			
13. Heeft u moeite met de combinatie van stappen en het zoeken van uw weg (vb in een station, luchthaven)?			

Totaal:

Appendix C

Dual-task Impact on Daily-living Activities Questionnaire (DIDA-Q)

Instructies: In het dagelijks leven voeren we vaak twee taken tegelijk uit, zoals wandelen en een boodschappenlijstje onthouden. Dit kan de moeilijkheid van de taak verhogen, wat kan leiden tot het verminderen van de wandelsnelheid, het stoppen met wandelen, of tot het veranderen van de focus van aandacht. Geef hieronder de moeilijkheidsgraad van elk van de volgende combinaties van taken aan.

Geef de moeilijkheidsgraad van de volgende taken aan:	Niet moeilijk	Beetje moeilijk	Enigszins moeilijk	Erg moeilijk	Extreem moeilijk
	0	1	2	3	4
1. Wandelen en een bord gevuld met eten dragen					
2. Wandelen en drinken uit een flesje of blikje					
3. Wandelen en iets onthouden (bijvoorbeeld de naam van een restaurant of de titel van een boek of film)					
4. Wandelen en luisteren naar iemand die praat					
5. Wandelen en een planning bedenken (bijvoorbeeld een maaltijd bereiden)					
6. Wandelen en de rits van je jas sluiten					
7. Met iemand een gesprek voeren en op een pad met bochten wandelen					
8. Wandelen en snel reageren op visuele prikkels (bijvoorbeeld stoppen voor het rode licht, lezen van verkeersborden)					
9. Met iemand een gesprek voeren en je evenwicht bewaren op je voeten[1]					
10. Met iemand praten en met een hoge snelheid wandelen					
11. Met iemand een gesprek voeren en snel van looprichting veranderen					
12. Met iemand een gesprek voeren en aan een spontane[2] snelheid wandelen					
13. Wandelen en aandacht hebben voor verkeersgeluiden in de straat					
14. Wandelen en iets uit je zak halen					

[1] Rechtopstaand, in stand

[2] Gebruikelijke, gewoonlijke

15. Wandelen en hoofdrekenen (bijv. wisselgeld berekenen)					
16. Wandelen en je telefoon gebruiken (bijvoorbeeld een contact opzoeken, een sms versturen)					
17. Een trap nemen en een zak dragen					
18. Wandelen en spreken					
19. Wandelen en luisteren naar muziek op de radio					

Scores:
Bovenste lidmaat vaardigheid (MMI): ___/24
Cognitie (MCI): ___/32
Balans en mobiliteit (MCI): ___/20
Totaal: ___/76

Appendix D

Decision-making tree for statistical analysis