

Table 1: Summary of measures of cognitive fatigability operationalised in existing research.

Candidate Measures	Studies	Procedure	Self-reported fatigue measure	Key Findings
The auditory As and auditory trails A tests (As and trails A tests)	1. Kujala et al. (1995)*	<p>Cognitive fatigability measures: Possible effects of cognitive fatigue were measured by recording the error rates for both first and second half of the test below.</p> <p>Continuous performance task: As and trails A tests</p> <p>Participants: PwMS ($n = 45$) were classified into either cognitively preserved, or cognitively mildly deteriorated and compared to ($n = 35$) healthy controls.</p>	None reported	Both MS groups showed signs of possible fatigue in the tests of sustained attention, doing significantly worse than controls. In addition, reaction times were shorter in the last part of the test in the controls compared with the first period in the MS groups.
Computerized Assessment of Response Bias (CARB).	2. Bruce, Bruce, & Arnett (2010)*	<p>Cognitive fatigability measure: Total Response time variability (RTV) reflects the total standard deviation of correct response times, across three blocks of the CARB, measured in milliseconds.</p> <p>Continuous performance task: CARB.</p> <p>Participants: People with MS (pwMS) experiencing fatigue ($n = 87$) and ($n = 24$) healthy controls were asked to complete the CARB. Results controlled for measures of secondary fatigue (depression).</p>	FIS: Physical, social and cognitive fatigue.	PwMS showed increased RTV when compared with controls, after controlling for information processing speed (Oral Symbol Digit Modalities Test (SDMT). Total RTV significantly correlated with the FIS total score ($r = .48$), and physical ($r = .28$), social, and cognitive fatigue ($r = .45$) subscales, but correlations varied across MS subtypes.
Computerised Delayed Item Recognition (DIR) task.	3. Holtzer & Foley (2009)*	<p>Cognitive fatigability measure: Experimentally manipulated executive demands. The DIR computerized test manipulates executive demands in three stepped conditions: Alone, Partial Interference (PI), and Complete Interference (CI).</p> <p>Continuous performance task: DIR</p> <p>Participants: People with relapse-remitting MS ($n = 20$) and matched controls ($n = 20$) completed the DIR. Results controlled for measures of secondary fatigue (depression).</p>	FSS: General measure of fatigue severity and impact	DIR performance was significantly slower and less accurate as executive demands increased across the three task conditions for pwMS compared to controls. Regression analyses showed self-reported fatigue (FSS) was related to DIR reaction time and accuracy only in the complete interference condition and only in the MS group.
Digit Symbol Coding (DSC1) and Relative (DSC2 as part of the broader Wechsler Adult Intelligence Scale (WAIS-III) and Wechsler Memory Scale	4. Andreasen et al. (2010)	<p>Cognitive fatigability measure: Processing speed using the scaled score of the DSC1, and Relative DSC2, described as a more conservative parameter (DSC1 divided by the average of Matrix Reasoning and Vocabulary to account for the influence of other cognitive parameters, at the start and end of the broader test battery. Cognitive fatigability brought about during the neuropsychological test procedure was defined as [DSC1/2 I] minus [DSC1/2 II].</p> <p>Demanding cognitive task: Logical Memory I, Digit Symbol Copy, Matrix Reasoning, Vocabulary and Logical Memory II of the WAIS-III and Wechsler Memory Scale (WMS-III).</p> <p>Participants: People ($n=60$) with mild to moderate levels of disability with relapse-remitting MS were stratified into two groups depending on the presence (≥ 5 FSS) ($n = 39$) or absence of fatigue ($n = 21$). The fatigue group was divided into primary ($n = 19$) and secondary fatigue ($n = 20$) based on assessments related to sleep, well-being, depression, pain, frequency, infection,</p>	FSS	DSC performance improved with repetition, and DSC1/2 I – II change scores were not significantly different between primary and secondary fatigued pwMS, or fatigued and non-fatigued pwMS. Greater self-reported fatigue (FSS) was significantly associated with slower processing speed (DSC1/2 I) at baseline ($r = -.35$).

(WMS-III) 3rd versions.		spasticity, and tiredness due to medication side-effects. Both subgroups and healthy controls ($n = 18$) completed all tests.		
Symbol Digit Modalities Test (SDMT)	5. De Giglio et al. (2015) Conference Abstract*	<p>Cognitive fatigability measure: SDMT1 followed by the SDMT2, recording the number of correct answers (NCA) for each test in 3-time intervals at 0-30s; 30-60s; 60-90s. The Information Processing Speed Deceleration Index (IPSDI) was estimated using the following equation: $(NCA\ time-3 - NCA\ time-1 / NCA\ time-1) * 100$.</p> <p>Continuous performance task: Not clear</p> <p>Participants: PwMS ($n = 55$) and healthy controls ($n = 44$) completed the SDMT twice in a row (SDMT1 and SDMT2).</p>	MFIS: Assesses the effects of fatigue in terms of physical, cognitive, and psychosocial functioning	PwMS performed worse than controls for mean NCA at both SDMT1 and SDMT2, but there was only a significant time by group interaction at SDMT2, indicating that NCA decreased over time in the MS group only. The two groups also showed a significant difference in mean IPSDI. In the MS group IPSDI was correlated with the MFIS (r not reported).
Modified Symbol-Digit Modalities Test (mSDMT) and n-back task Computerised version	6. Sandry et al. (2014)*	<p>Cognitive fatigability measure: Processing speed (mSDMT) and working memory domains (The 2-back and 0-back version of the n-back task), with different levels of cognitive load, were assessed. Accuracy rate and reaction time data of both tasks were analysed.</p> <p>Continuous performance task: mSDMT and n-back task</p> <p>Participants: PwMS ($n = 32$) and healthy controls ($n = 24$) completed processing speed and working memory tasks over two separate testing sessions within a two-week time period. Each session involved different cognitive domains; either a processing speed (i.e. mSDMT) or working memory task (i.e. The 2-back and 0-back version of the n-back task). Results partially controlled for measures of secondary fatigue (depression).</p>	MFIS and FSS VAS: State / momentary fatigue was measured at baseline and once after each block (scale of 0–100).	There were no differences between the groups for accuracy rate across both tasks. However, there was a significant group effect for reaction time data, with slower times for pwMS compared to controls. Reaction times were significantly slower in the high, rather than the low cognitive load condition, and pwMS showed a significantly larger difference between cognitive domains compared to controls. A larger difference in reaction times between pwMS and controls in the high cognitive load condition of the processing speed (mSDMT) task was also identified. The MS group reported higher depression and fatigue (FSS and MFIS), but correlations between VAS fatigue scores and reaction time or accuracy data for both tasks were not significant.
N-back task, involving attention (0-back) (see also Sandry et al 2014),	7. Bailey, Channon, & Beaumont (2007)	<p>Cognitive fatigability measure: A continuous n-back computerized task, involving attention (0-back and 1-back), at the beginning and end of one testing session. Percentage of correct responses and median reaction time was recorded. Performance was compared across the first, second and third pairs of blocks in the test. The first and second presentations of each test were also compared.</p> <p>Demanding cognitive task: The Ravens Coloured Progressive Matrices or Spot the Word plus Rule finding.</p> <p>Participants: PwMS with fatigue ($n = 14$) and matched healthy controls ($n = 17$).</p>	FSS FRS was measured four times within the testing session.	There were no differences in change in n-back performance during the sessions between pwMS and controls. PwMS did report a greater increase, than the control group, in the level of subjective fatigue during the 1-back testing session. However, change in subjective fatigue did not correlate significantly with change in cognitive performance (r were not reported).
Paced Auditory Serial Addition Test (PASAT)	8. Bryant et al. (2004)*	<p>Cognitive fatigability measures: Comparing performance on the first versus the second half of each of the four trials of the PASAT during a single administration of the PASAT, using two scoring methods: (1) Sum of correct responses for each PASAT trial, and the first and second half of each trial. Cognitive fatigue was defined as a decrease in the number of correct responses generated in the second half (“later responses”) compared with the first half (“earlier responses”)</p>	FAI	(1) Cognitively impaired PwMS produced significantly fewer correct responses compared to either non-impaired pwMS or controls, who performed at a similar level. Performance decreased reliably across trials, with a reduction in accuracy from earlier to later responses. However, pwMS

of a trial, and (2) proportion of correct responses immediately following another correct response (a “dyad”) while performing a mathematical operation.

Continuous performance task: PASAT

Participants: PwMS ($n = 56$) were grouped as being either cognitively impaired ($n = 27$) or cognitively non-impaired ($n = 29$) based on other neuropsychological tests and compared to matched healthy controls ($n = 39$). All subjects were then given a single administration of the PASAT.

showed the same pattern of cognitive fatigue within trials as controls, regardless of impairment level.

(2) Controls and non-impaired pwMS had more correct responses compared to the cognitively impaired pwMS. Performance was no different between controls and the non-impaired pwMS. Whilst controls only showed a significant reduction in percent dyad scores in Trial 4, cognitively impaired and non-impaired pwMS showed a significant reduction in dyad scores in Trial 3, reaching the limit of their ability to sustain central executive load at an earlier time point.

Subjective fatigue (FAI) did not correlate with number of correct responses, or percent dyad score, on the PASAT for controls, or for cognitively impaired and non-impaired pwMS.

9. Johnson et al. (1997)

Cognitive fatigability measure: The PASAT was administered four times over a 3 hour testing period with 30-min intervals between sessions. The dependent variable was the total number of correct responses summed across the four trials.

Demanding cognitive task: During the 30 min intersession period between tests, participants completed neuropsychological tests, assessing attention concentration, and memory from the WAIS (not specified) to further increase the level of participant’s fatigue.

Participants: PwMS ($n = 15$), those with depression ($n = 14$), chronic fatigue syndrome (CFS) ($n = 15$) and healthy controls ($n = 15$). Results partially controlled for measures of secondary fatigue (depression).

5-point NRS: 0 (none) to 4 (severe) measured before each administration

Findings showed no effect for a “blunting” of practice effect on the PASAT, and there were no differences in PASAT performance between pwMS, those with depression, CFS and controls. In addition, subjective fatigue and depression were not significantly related to PASAT performance (ANOVA only - r not reported).

10. Morrow, Rosehart, & Johnson (2015) (not clear)

Cognitive fatigability measure: The number of correct responses given during the first third of the test to the number given during the last third.

Continuous performance task: PASAT

Participants: ($n = 100$) and pwMS and ($n = 130$) healthy controls.

FSS

On average pwMS had 2 to 3 fewer correct responses in the last third than the first third of the test compared to controls. However, authors do not report whether these differences were statistically significant. Self-reported fatigue scores (FSS) correlated significantly, but only very weakly, with total correct responses in the last third of the test PASAT ($r = 0.11$).

11. Walker et al. (2012)*
- Cognitive fatigability measures:** (a) Two PASAT assessments (2" vs. 3" inter stimulus intervals versions) and three reaction time measures of the Test of Information Processing (CTIP): Simple (SRT), Choice (CRT), and Semantic Search reaction (SSRT), and (b) second half of the PASAT compared to the first and third block of the CTIP compared to the first. All three tests were scored using three methods (similar to Bryant 2004): (1) Total number of correct responses, (2) Total dyad score and (3) Percent dyad score, defined as the proportion of time pwMS met task demands: $(1 - (\text{total correct score} - \text{dyad score}) / \text{total correct score}) \times 100$.
- Continuous performance task:** PASAT and CTIP
- Participants:** PwMS ($n = 70$) with relapsing–remitting MS and matched healthy controls ($n = 72$) completed the PASAT three times (each time with the 3" and 2" versions) and CTIP as part of larger battery of tests, which were interspersed between administrations. To reduce fatigue, tests were administered over two test sessions one week apart. The PASAT was administered two times during the first test session, and a third time during the second test session. Results controlled for measures of secondary fatigue (depression).
- FIS
- (a) There were no group differences in total number of correct responses for both PASAT 2" and 3" and CTIP, using the three scoring methods.
- (b) There were no differences in total number of correct responses between groups for the second half of PASAT 2" and 3", and first and third block of the CTIP. However, the percent dyad scoring method was significantly different on the second half of the task for both the PASAT 2" and 3" when compared to controls. Differences between groups on the three separate reaction time measures of the CTIP using the total dyad and percent dyad scoring methods were unclear. There was a significant difference between groups on the PASAT 3", where pwMS performed worse than controls, but not version 2".
- Correlations between subjective fatigue (total FIS score and the cognitive subscale) were consistently small, but significant, across the two PASAT tests and three scoring methods.
12. Schwid et al (2000) Conference Abstract only*
- Cognitive fatigability measures:** The percent decrement in correct responses during the first 10 items of the PASAT 3" compared to the last 10 items. Motor fatigue was measured during 30-second sustained contractions of four lower extremity muscle groups.
- Continuous performance task:** PASAT
- Participants:** All pwMS ($n = 30$) performed the PASAT 3" at least three times during the past year.
- FSS
RFD: Hourly VAS rating of fatigue for 48 hours.
- PwMS experienced an average decline in performance of 17.8% during the PASAT task. Individual declines in cognitive function were unrelated to cognitive impairment (total PASAT score), physical impairment, subjective fatigue, or motor fatigue (both $r = <0.3$, ns respectively).
13. Schwid et al. (2003)
- Cognitive fatigability measures:** A decline in performance from the beginning to the end of the test on two tasks: PASAT 3" and the Digit Ordering Test (DOT). Two methods of scoring for were used: (1) Percent decline in performance using the ratio of the number of correct responses for the first 20 items of the PASAT (60 items total) to the last 20 items, or the first five trials for the DOT (15 trials total) to the last five trials, (2) the slope of the linear regression of the number of correct responses per each 10 items of the PASAT versus the number of the decile, or the number correct per trial for the DOT versus the number of the trial.
- Continuous performance task:** PASAT 3" and DOT.
- Participants:** PwMS ($n = 20$), who were ambulatory and had no significant cognitive impairment or depression, and matched controls ($n = 21$), completed the PASAT 3" and DOT twice at a screening visit in an effort to stabilise performance. Within one month participants
- FSS, MFIS and RFD.
- (1) There were no significant differences in either the DOT or the PASAT performance between groups. However, the PASAT showed a 5.3% decline in performance from the start to the end of the test.
- (2) There were no significant differences between groups in the DOT or the PASAT.
- PASAT (2) scores were associated with subjective fatigue (FSS) in pwMS ($r = 0.58$), but not controls, but were not associated with the MFIS and RFD, or the cognitive subscale of the MFIS. Correlations for PASAT (1) and DOT (1 and 2) were not reported.

returned for two identical visits, separated by an average of 7 days, at which they performed the two tests with 10 minute intervals between tests.

Tower of London (TOL), Paired Associates Learning Test (PALT), Stroop, and The Wisconsin Card Sorting Test (WCST)	14. Parmenter, Denney, & Lynch (2003)	<p>Cognitive fatigability measure: In an initial session, pwMS completed the TOL, PALT, Stroop and WCST of planning, selective attention, and paired associate learning. During the second session (not clear how long after), the same tests were re-administered, but The WCST was not. Groups were counterbalanced in terms of the order of receiving the tests. Each test had their own unique scoring method.</p> <p>Demanding cognitive / continuous performance task: Not clear.</p> <p>Participants PwMS ($n = 30$) who had substantial fatigue, and who reported significant daily variation in fatigue severity, were tested on two occasions during a self-reported period of high fatigue and relatively low fatigue.</p>	<p>FSS and FIS completed at the start of the first and second session.</p> <p>Abbreviated version of the POMS: A state measure of fatigue.</p>	<p>PwMS experienced greater self-reported fatigue during the period of high fatigue, feeling they had performed worse during this period. However, there were no differences in cognitive performance that could be attributed to fatigue. Rather all subjects showed improvement from the first to the second session regardless of whether the latter was a period of high or low fatigue.</p>
Six-trial version Selective Reminding Test (SRT), Spatial Recall Test (SRT) and Tower of Hanoi Test (TOH Test).	15. Krupp & Elkins (2000)*	<p>Cognitive fatigability measure: Performance on the SRT, SRT and TOH (visual memory, verbal memory, and verbal fluency) before and after: a continuous effortful task.</p> <p>Continuous performance task: Alpha-Arithmetic Test (A-A Test).</p> <p>Participants: The SRT, SRT and TOH were administered with pwMS experiencing fatigue ($n = 45$) and healthy controls ($n = 14$), followed by the A-A Test (completing mental arithmetic problems administered on a computer), and then the first three tests were repeated. Results partially controlled for measures of secondary fatigue (depression).</p>	<p>FSS</p> <p>PANAS: Momentary assessments of subjective physical and mental fatigue, and mood, recorded three times during the evaluation.</p>	<p>Following the A-A Test, performance on the SRT, SRT and TOH tests declined for pwMS and improved for controls.</p> <p>There were differences in mood across the two groups over the three time points, but MS and control participants reported an increase in perceived mental and physical fatigue (PANAS) across the testing session compared to baseline. However, baseline self-reported fatigue (FSS) did not correlate with changes in cognitive fatigability assessments (r not reported). PwMS with baseline cognitive impairment were also compared to pwMS without, showing no significant differences in SRT, SRT and TOH scores between the two subgroups.</p>
Test battery for Attention Performance (TAP-M/version mobility): The alertness subtest computerised version	16. Claros-Salinas et al. (2013)*	<p>Cognitive fatigability measure: Performed alertness, selective, and divided attention subtests from the TAP twice: during rest (baseline) and before and after treadmill training and cognitive load. Attention tests were performed on three different days on (a) a weekend morning before and after a rest period, (b) a weekday before and after cognitive load, and (c) a week day before and after treadmill training. Performance on the alertness task was median reaction time; selective attention median reaction time and errors; divided attention median reaction times and errors.</p> <p>Demanding cognitive task: Standardised battery of neuropsychological tests lasting 2.5 hours (unclear), including domains of attention, word recognition, verbal fluency, memory, calculation, and visuo-spatial and reasoning.</p> <p>Participants: PwMS ($n = 32$) with fatigue and healthy controls ($n = 20$).</p>	<p>FSMC: Cognitive fatigue and 10-point NRS completed before each testing session</p>	<p>PwMS showed significantly increased reaction times on the alertness test after treadmill training and after cognitive load, whilst control subjects had no change in performance. No significant increases in reaction times we shown in the divided and selective attention tasks.</p> <p>Self-reported cognitive, motor and overall fatigue (FSMC) were only significantly related to the reaction differences of the alertness test in the cognitive load condition ($r = .48, .36, \text{ and } .44$ respectively), but the 10-point NRS was not.</p>

17. Moyano et al. (2013) Conference Abstract only	<p>Cognitive fatigability measure: Omissions and mistakes during the flexibility and divided attention tasks of the TAP in two separate testing sessions (not clear).</p> <p>Continuous performance task / Demanding cognitive task: Not clear</p> <p>Participants: PwMS ($n = 43$) and controls ($n = 37$) with similar age and education level completed a one hour neuropsychological testing session, which was split into two parts.</p>	FSMC cognitive subscale and fatigue VAS at the beginning and at the end of testing.	There were no significant differences between groups for the cognitive flexibility domain of the TAP, but there were differences for omissions in the divided attention task, in the second testing session. There were no differences between pwMS and control group for any VAS and FSMC measures, showing a similar level of subjective fatigue. No significant correlations between the VAS were identified. In addition, cognitive fatigue and omissions in both the first or second part of the divided attention test were not significantly related.
18. Neumann et al. (2014)*	<p>Cognitive fatigability measure: Median reaction times of the alertness subtest from the TAP was measured three times: At rest, following a 2.5 hour test session inducing high cognitive load, and during a one hour recovery period.</p> <p>Demanding cognitive task: Paper and pencil-tests (not specified, but reportedly the same as Claros-Salinas et al. 2013), including the domains of attention, word recognition, verbal fluency, memory, calculation as well as visuo-spatial and reasoning abilities.</p> <p>Participants: pwMS ($n = 30$) with self-reported cognitive fatigue (FSMC) and healthy controls ($n = 15$). Secondary fatigue was accounted by excluding participants with sleep problems and depression (i.e. Epworth Sleepiness Scale and the BDI-II).</p>	FSMC	Performance was significantly worse for pwMS than controls following the test session. During the one hour recovery period pwMS reaction times returned to baseline level. In contrast, performance of controls continued to gradually improve across the three conditions. Self-reported cognitive fatigue (FSMC) and reaction time alertness were positively correlated ($r = 0.54$).
19. Weinges-Evers et al. (2010)*	<p>Cognitive fatigability measure: Cross-sectional study asking pwMS to complete three tests within a single session lasting approximately 1 hour. Tests included the TAP Alertness, Visual Scanning and Executive Control subtests.</p> <p>Continuous performance task / Demanding cognitive task: Not clear</p> <p>Participants: PwMS ($n = 110$) were classified into groups after completing several neuropsychological tests based on these findings, of which $n = 56$ were fatigued and $n = 53$ were not fatigued according to the FSS. Results controlled for measures of secondary fatigue (depression).</p>	FSS	Fatigued pwMS had significantly longer mean reaction times only on the alertness subtest compared those who were not fatigued. In contrast to other subtests, regression findings showed that self-reported fatigue was an independent predictor of performance in the alertness subtest.
Word List Learning 1 and vigilance (Distractibility Task (Gordon	20. Paul, et al. 1998) <p>Cognitive fatigability measure: Grip strength tests, Word list learning and vigilance tasks before and after 30 minutes of demanding cognitive tasks.</p> <p>Demanding cognitive task: Verbal fluency and vocabulary and comprehension from the Wechsler Adult Intelligence Scale-Revised (WAIS-R).</p>	Authors developed separate NRS for physical and cognitive fatigue: 1 (not	PwMS reported more self-reported physical and cognitive fatigue than controls at baseline, and performed more poorly on the grip strength, word list learning, and vigilance tasks. However, following cognitive tasks pwMS reported increased physical and cognitive fatigue (r not reported), but their performance on grip strength, learning, and vigilance tasks were no different from baseline.

Systems Inc., DeWitt, NY)		Participants: PwMS ($n = 39$) and matched healthy control ($n = 19$).	at all) and 5 (a great deal)	Controls showed no change in self-reported fatigue ratings or performance on any tests.
Name not specified	21. Jennekens -Schinkel et al (1988)*	Cognitive fatigability measure: A stimulus and response panel and a reaction time and error recording device, which measured simple and disjunctive reaction times on visual and auditory tasks before and after a demanding cognitive task. Demanding cognitive task: A neuropsychological assessment lasting 4 hours, assessing motor speed, intelligence, reasoning, memory span, recall, recognition and list learning, interference sensitivity, rule application, copying drawings, confrontation naming, reading, writing and calculation (tests not specified). Participants: Ambulatory pwMS ($n = 39$) and healthy controls ($n = 25$).	Not reported	Reaction times for the visual stimulus tasks before and after the demanding cognitive task were significantly longer for pwMS than controls, but not for the combined visual-auditory stimulus condition. Visual tasks reaction time was related to disease duration and neurological disability.

Abbreviations: Analysis of Variance (ANOVA); r (Pearson's r coefficient).

Fatigue self-report scales:

1. FIS: Fatigue Impact Scale
2. FRS: Fatigue Rating Scale
3. FSMC: Fatigue Scale of Motor and Cognition
4. FSS: Fatigue Severity Scale
5. MFIS: Modified Fatigue Impact Scale
6. NRS: Numerical Rating Scale
7. PANAS: Positive and Negative Affect Schedule
8. POMS: Profile of Mood States
9. RFD: Rochester Fatigue Diary
10. VAS: Visual Analogue Scale