Taking the tester out of the SDMT: A proof of concept fully automated approach to assessing processing speed in people with MS

Viral Prakash Patel, Lingkai Shen, Jonathan Rose and Anthony Feinstein

Abstract

Background: One factor hindering the widespread use of cognitive testing for people with multiple sclerosis (pwMS) is the need for a tester to administer tests.

Objective: To undertake a proof of concept study assessing the feasibility of a fully automated speech recognition version of the Symbol Digit Modalities Test (auto-SDMT) in detecting abnormalities in processing speed in pwMS.

Methods: A sample of 50 pwMS and 32 matched healthy control (HC) subjects was tested with the auto-SDMT and the Brief International Cognitive Assessment for MS (BICAMS).

Results: The percentages of MS participants impaired on the auto-SDMT and the traditional oral SDMT were 34% and 32%, respectively. Excellent convergent validity was found between the two tests (MS: \( r = -0.806, p < 0.001 \) and HC: \( r = -0.629, p < 0.001 \)). The auto-SDMT had a similar sensitivity and specificity to the traditional oral SDMT in predicting overall impairment on the BICAMS.

Conclusion: The auto-SDMT is a sensitive measure for detecting processing speed deficits in pwMS. The test, the first entirely computer administrated oral response version of the SDMT, uses speech recognition technology, thereby eliminating the need for a human tester. Replication of the results is required in a larger representative sample of pwMS.

Keywords: Multiple sclerosis, cognitive assessment, information processing speed, computerized testing, automated testing

Introduction

The deleterious effects of cognitive dysfunction in people with multiple sclerosis (pwMS) can be considerable.¹ As such detection and monitoring change over time is important. Considerable attention has been devoted to determining the best way to do this. Approaches here range from using a more comprehensive cognitive battery to brief monitoring tools. Notwithstanding the method chosen, there are barriers that limit their use. These include a shortage of neuropsychologists,² the costs associated with the assessments and the time taken to administer the tests.³ Should the resources be available for cognitive testing, another hurdle presents itself: it is often not feasible to serially monitor a patient’s cognition in a busy neurological clinic, even with a shorter monitoring battery. As a result, attention has shifted to determining which cognitive test might be the method of choice to do so.

A consensus has emerged that the Symbol Digit Modalities Test (SDMT),⁴ a measure of information processing speed, is the best, single psychometric option for use in pwMS.⁵ Slowness in processing speed affects nearly 50% of the MS population⁶ and is associated with deficits in other cognitive domains such as memory⁷ and executive function.⁸ Consequently, the oral SDMT is a central component in many comprehensive cognitive batteries such as the Minimal Assessment of Cognitive Functioning in MS (MACFIMS)⁹ and shorter monitoring tools such as the Brief International Cognitive Assessment for MS (BICAMS).³ Furthermore, the oral SDMT has several psychometric advantages making it an attractive...
option for clinical and research settings. The test is straightforward to administer, takes less than 5 minutes and is relatively resistant to practice effects over time. The results correlate well with magnetic resonance imaging (MRI) markers of disease pathology. Importantly, a 10% change in test performance over time is now considered clinically meaningful, thereby conferring ecological validity as well.

Despite these numerous attributes, the SDMT is not always administered routinely in clinical settings as part of MS patient care. One major reason may be the need for a tester to administer and score the responses. A fully automated, computerized version of the SDMT (auto-SDMT) that dispenses entirely with a human tester and which has a minimal motor component may offer a way around this problem. The present feasibility study compares such a test with the traditional oral tester-administered version in a sample of pwMS. No serial or test–retest data will be presented here. Rather, the primary aim is to show that voice recognition software is sophisticated enough to allow for the complete removal of a human tester from the administration process.

Methods

Sample

For this proof of concept study, a consecutive sample of 50 pwMS between the ages of 21 and 60 were recruited from two outpatient clinics. Exclusion criteria included a history of another disease of the central nervous system, traumatic brain injury, major psychiatric illness (dementia or psychosis), learning disability, substance abuse and/or neuropsychological testing done within the past year.

A demographically matched sample of 32 healthy individuals was also recruited from the community using online advertisement. Exclusion criteria for the healthy control (HC) group were the same as the MS group. Interested participants were screened via a telephone interview to ensure they met the exclusion criteria.

All participants were reimbursed costs for parking and transportation to the hospital.

Demographic and neurological data

Demographic (age, sex and years of education) and neurological (disease course and physical disability based on the Expanded Disability Status Scale (EDSS)) data were collected prior to the cognitive assessment.

Cognitive assessment

The auto-SDMT. In a previous study, we validated a semi-automated computerized version of the traditional paper version of SDMT in the MS population. In the test, a row of nine boxes each filled with a symbol is presented on a computer screen. Above this row of boxes, a symbol–digit key is displayed that pairs each symbol with a number from 1 to 9. Using the symbol–digit key, participants are required to verbally match each symbol in the bottom row with a number as quickly as they can from left to right. After the participant matches the last symbol–digit pairing for a given trial, the tester clicks the mouse to proceed to the next trial. This is repeated for a total of eight trials. For each trial, the symbol–digit key remains constant, but the order of the test symbols changes (see Figure 1). Unlike the traditional pencil and paper SDMT which measures the number of correct responses in 90 seconds, the computerized SDMT captures the mean time for the eight trials.

The auto-SDMT is a fully automated version of the computerized SDMT. This program runs within a specific browser, Google’s Chrome browser, on a Windows or MacOS-based computer that must be equipped with a reasonable quality microphone and audio speakers. In the current study, a Windows 10 laptop and a Snowball iCE USB microphone were used to administer the auto-SDMT. The program consists of a user interface paired with a speech recognition module that accepts audio input to the test. The code is written in JavaScript language. The speech recognition module makes use of Google’s Speech Recognition service, provided through the Internet. The program could be easily changed to use alternative speech recognition methods, including local-to-the-computer methods. Verbal instructions to the participants and data collection are all generated by the program. The auto-SDMT does not require a tester to administer the test. Once a participant is comfortably seated in front of a computer, the auto-SDMT is initiated with the click of a mouse. In the current study, the research assistant played no role in the administration of the auto-SDMT other than to instruct the participant to click the mouse to start the test. While participants completed the test, the research assistant remained in the room as an observer to ensure that the program did not crash.

The auto-SDMT begins with the computer asking the participant for general demographic information (age, sex and total years of education) before it proceeds to administer the Snellen eye test to ensure participants have a minimal visual acuity of 20/70 needed to complete the next part. The computerized version of the
Snellen eye test is identical to the paper version, and no changes were made to the format (i.e. the layout and font sizes were kept in accordance to the original Snellen eye test). If the participant fails the eye test, the program does not proceed and tells the participant to inform the study personnel. If the participant passes the Snellen eye test, the program proceeds to give instructions on how to complete the SDMT. In the instructions phase, the participant is required to successfully complete a practice trial before starting the actual test. The practice trial served as an informal measure of participants’ oral-motor ability. After the practice run, eight trials of the auto-SDMT are administered. The speech recognition system monitors the verbal responses, and the test automatically proceeds to the next trial once the participant completes the last symbol–digit pairing. If an error is made, the auto-SDMT program automatically makes note of the error. Participants are not redirected to the place on the computer screen but encouraged to continue with the next symbol if an error is made. Since only nine symbols are presented at a time on the auto-SDMT, participants were not permitted to use their finger to keep track of where they are. Once the test is complete, results are downloaded onto the computer. The auto-SDMT provides raw times (in seconds) for each of the eight trials, a mean time for the eight trials, a total time for the eight trials and the total number of errors committed. The primary measure is the mean time (in seconds) for the eight trials. Failure on the auto-SDMT was defined as a mean time greater than 1.5 standard deviations (SDs) above the normative mean obtained from the sample of 32 HCs. A video of the test can be viewed in the online supplementary section, available at http://journals.sagepub.com/doi/suppl/10.1177/1352458518792772. For serial testing, there are 10 alternate versions of the auto-SDMT available. In each version, the symbols are the same, but the symbol–number pairing is different. Test–retest data were not collected for the current feasibility study.

**BICAMS.** All participants were also administered the BICAMS battery. To briefly summarize, the battery consists of the following:

1. Information processing speed: the SDMT
2. Verbal and visual memory – learning trials: California Verbal Memory Test–II (CVLT-II) and the Brief Visuospatial Memory Test–Revised (BVMT-R)

Failure on each cognitive index is defined as a score of 1.5 SDs below the normative mean of the 32 healthy individuals. Based on previously published research, we used a criteria for global impairment as failure on one or more cognitive indices.

**Order of testing**

In order to control for practice effects between the auto-SDMT and the traditional oral SDMT, that is, part of the BICAMS, a counterbalanced study design was used. Half of the participants in the MS and HC groups were administered the auto-SDMT first and the other half the traditional SDMT first. Order of testing was switched for each consecutive participant, and the two versions of the SDMT were administered approximately 60 minutes apart. In between administrations, participants completed the demographic questionnaire and were administered the learning trails of the BVMT-R, CVLT-II and the
Multiple Sclerosis Journal 00(0)

Wechsler Test of Adult Reading (WTAR). This process took approximately 20 to 30 minutes. For the remainder of the time, participants were allowed to take a break and asked to return to the study room at a specified time.

**Premorbid IQ**
Premorbid IQ was calculated using the WTAR.\textsuperscript{18} Raw scores from the WTAR were then converted to a full-scale IQ score based on the Wechsler Adult Intelligence Scale–III (WAIS-III).\textsuperscript{19}

**Statistical analysis**
Normality of data was assessed using the Kolmogorov–Smirnov test. Cognitive and demographic comparisons between MS and HC participants were conducted using the \( t \)-test for normally distributed data and chi-square test for ordinal data. Sensitivity and specificity of the auto-SDMT were determined against the BICAMS. Pearson’s correlations were calculated between the auto-SDMT and the traditional oral SDMT. Demographic (age, sex and years of education) and neurologic (EDSS) predictors of performance on the auto-SDMT and the traditional oral SDMT were determined using a linear regression analysis. Statistical significance was set at \( p < 0.05 \).

**Informed consent**
A research ethics board approval was obtained to conduct this study, and informed consent was obtained from all participants.

**Results**

### Demographic and disease-related data
Demographic and disease-related data are presented in Table 1. There were no demographic differences between the MS and HC groups. The MS group comprised primarily participants with relapsing–remitting disease (86%), and the mean EDSS was 2.46 (SD = 1.67).

### Cognitive data
Cognitive comparisons on the BICAMS between the MS group and HC group are presented in Table 2. The mean total time for the eight trials on the auto-SDMT was 105.11 seconds (SD = 29.83). MS participants were significantly slower on the auto-SDMT compared to the HC group (mean time = 13.14 seconds (SD = 3.73) vs 11.09 seconds (SD = 2.03), respectively, \( t = -3.215, p = 0.002, \text{Cohen’s } d = 0.68 \)). The MS group was also slower on the traditional oral SDMT compared to the HC group (47.22 (SD = 11.96) vs 53.25 (SD = 8.68), respectively, \( t = 2.464, p = 0.016, \text{Cohen’s } d = 0.58 \)). There were no differences on the auto-SDMT between participants administered the test first versus second within the MS (12.46 seconds (SD = 3.55) vs 13.81 seconds, respectively, \( t = -1.288, p = 0.204 \)) or the HC (11.49 seconds (SD = 2.01) vs 10.69 seconds (SD = 2.02), respectively, \( t = 1.127, p = 0.269 \)).

### Comparisons with the oral SDMT
The percentages of MS participants impaired on the auto-SDMT and the traditional oral SDMT were 34%...
and 32%, respectively. This did not differ statistically (McNemar’s test, \( p > 0.999 \)). Based on the 1.5 SD threshold, the sensitivity and specificity of the auto-SDMT in detecting overall cognitive impairment relative to the BICAMS were 67% and 96%, respectively (Table 3). Percentages for the traditional oral SDMT were 67% and 100%, respectively.

**Convergent validity.** Significant correlations between the auto-SDMT and the traditional oral SDMT were found in both the MS (\( r = -0.806, p < 0.001 \)) and HC (\( r = -0.629, p < 0.001 \)) groups. Scatterplots comparing the performance on the auto-SDMT and the traditional oral SDMT are presented in Figure 2(a) and (b).

In the MS group, the auto-SDMT also correlated significantly with CVLT-II–immediate recall (\( r = -0.479, p < 0.001 \)) and BVMT-R–immediate recall (\( r = -0.537, p < 0.001 \)). In the HC group, the auto-SDMT correlated with CVLT-II–immediate recall (\( r = -0.598, p < 0.001 \)), but only a trend emerged with BVMT-R–immediate recall (\( r = -0.314, p = 0.08 \)).

**Demographic predictors of the auto-SDMT**

Based on a linear regression analysis, age predicted performance on the auto-SDMT in both the MS (\( B = 0.102, t = 2.205, p = 0.033 \)) and HC (\( B = 0.123, t = 2.777, p = 0.008 \)) groups. On the traditional oral SDMT, age predicted performance in the MS group (\( B = -0.457, t = -3.228, p = 0.002 \)), but not in the HC group (\( B = -0.295, t = -1.999, p = 0.055 \)). Sex and years of education were not associated with test performance on either the auto-SDMT or the traditional oral SDMT in both the MS and HC groups. EDSS was not associated with performance on either the auto-SDMT (\( B = 0.452, t = 1.443, p = 0.156 \)) or the traditional oral SDMT (\( B = -1.080, t = -1.130, p = 0.265 \)).

**Test preference**

When asked if participants prefer completing the auto-SDMT or the traditional oral SDMT, more participants preferred the former in both the MS (70% vs 30%, respectively; McNemar’s test, \( p = 0.005 \)) and HC (78% vs 22%, respectively; McNemar’s test, \( p = 0.001 \)) groups.

**Discussion**

In the present study, we demonstrate the feasibility and utility of using the auto-SDMT in pwMS. The auto-SDMT was able to detect differences in processing speed between the MS and HC groups, slightly outperforming the traditional tester-administered SDMT with which it had excellent convergent validity. Performance on both versions of the SDMT in pwMS was influenced by age with participants preferring to complete the auto-SDMT over the traditional version.
Neuropsychological testing is generally conducted with a trained psychometrician administering the cognitive tests. In the last few decades, the advent of computer-based tests has reduced the neuropsychological expertise required to administer cognitive tests. This has made cognitive testing more widely available and potentially less time-consuming given the ability of computers to generate results quickly. The SDMT has lent itself well to this process. We have previously shown how one semi-automated version matched the traditional version of the SDMT when it comes to testing pwMS.13,20 Of note, however, is that our earlier computerized version differed from the oral version in that the primary outcome recorded is a mean response time across a fixed number of symbol–digit trials, whereas the latter captures the number of correct responses in 90 seconds. The auto-SDMT also captures a mean response time for eight trials with each trial encompassing nine symbol–digit pairings.

One barrier to the more widespread use of recently developed computerized tests is that they still require administration by a human tester. A superior approach here would be to develop fully automated rather than semi-automated tests that require no human supervision. Full automation, however, does present an additional challenge, namely, having cognitively impaired patients essentially test themselves. Artificial intelligence – namely, the speech recognition technology embedded within the auto-SDMT – can overcome this drawback, as our data show. The program has two checks to ensure that patients will be able to complete testing. The first is an eye test to make sure patients have sufficient visual acuity to see the symbols and numbers. The second is a practice trial that must be successfully completed before the full test can begin. Should a patient fail either of these two preliminary steps, the program will not proceed to administer the test and will inform the patient to contact a health care worker.

The auto-SDMT is, to our knowledge, the second automated test of processing speed that does not require a tester, the other being the Processing Speed Test (PST).21 The iPad based PST was found to have good test–retest reliability, was slightly more sensitive in discriminating MS patients from HCs than the oral SDMT and correlated well with MRI indices. Importantly, the results were not dependent on whether a tester was present or absent during the assessment. However, the PST requires subjects to move an arm and hand across the iPad screen to find the correct number to press and as such is dependent on relatively intact upper limb function. For this reason, it is programmed to run for 30 seconds longer than the traditional oral SDMT. Notwithstanding this modification, significant weakness or coordination problems would make the completion of the test problematic given that approximately 50% of pwMS experience some type of upper limb disability.22,23 Our auto-SDMT was developed with this in mind. There is no motor component to the test apart from the need for intact oral communication, as is the case with the traditional oral SDMT.

Another potential advantage to the auto-SDMT is that one can obtain times for each of the eight symbol–digit lines. Performance can therefore be automatically tracked serially during the test providing useful data on whether responses speed up with practice or slow with wavering attention, fatigue or both. While similar data

![Figure 2. Comparison of performance on the auto-SDMT versus the traditional oral SDMT in healthy controls (a) and people with MS (b).](image-url)
may be obtained from the traditional oral version, the process is much more laborious, requiring the tester to not only record errors and correct responses but also track these in blocks of time.

Our study is not without limitations. First, the sample is modest and comprises mainly subjects with relapsing–remitting MS. Second, we have yet to collect test–retest data. Here, it is germane to note that the auto-SDMT is programmed with 10 alternate symbol–digit variations to minimize the effects of practice with serial administration. The question of inter-rater reliability is largely answered by the computer never varying in the administration of the test. Third, the auto-SDMT is currently only programmed for the English language. Finally, the automation of the current version of the auto-SDMT does not include the calculation of z-scores. However, once a larger normative database has been established, the programming could be expanded to provide these scores automatically as well.

In summary, the auto-SDMT has three potential advantages over the traditional version. (1) The test uses speech recognition technology, which eliminates the need for a human tester. This opens up its use to all professionals in an MS clinic, be they nurses, occupational therapists, clinic coordinators, psychologists or neurologists. (2) The administration is completely standardized, thereby eliminating in theory tester-induced inter- and intra-related variability. It also standardizes assessments across multiple sites; (3) automatically and instantly provides the results on test completion. Interpretation of these results is identical to that for the traditional oral SDMT, namely, pass or fail based on a threshold score obtained from normative, HC data. For these reasons, the auto-SDMT can be a potentially useful screening tool for assessing cognition both in the clinic and in the patient’s home. Results now require replication in a larger sample. To this end, the next steps in our development of the test include collecting a large normative database, establishing test–retest reliability, expanding the administration to pwMS who have primary and secondary MS and exploring associations between performance on the test and MRI indices of brain involvement.

Declaration of Conflicting Interests
The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: V.P.P. declares that there is no conflict of interest. L.S. declares that there is no conflict of interest. A.F. has received speaker honoraria from Merck Serono, Teva Pharmaceuticals Industries Ltd., Novartis and Biogen Idec; serves on the editorial boards of Multiple Sclerosis Journal; receives publishing royalties for The Clinical Neuropsychiatry of Multiple Sclerosis (Cambridge University Press, 2007); chairs the Medical Advisory Committee for the Multiple Sclerosis Society of Canada; conducts neuropsychiatric evaluation, cognitive testing and brain imaging in neuropsychiatry in his clinical practice; and receives research support from the Multiple Sclerosis Society of Canada.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Multiple Sclerosis Society of Canada.

ORCID iD
Jonathan Rose https://orcid.org/0000-0002-3551-2175

References


