Repeated Assessment and Practice Effects of the Written Symbol Digit Modalities Test Using a Short Inter-Test Interval

Diana R. Pereira; Patricio Costa; Joao J. Cerqueira

Archives of Clinical Neuropsychology 2015;

doi: 10.1093/arclin/acv028

Abstract

The Symbol Digit Modalities Test (SDMT) is a widely used instrument to assess information processing speed, attention, visual scanning and tracking. Considering that repeated evaluations are a common need in neuropsychological assessment routines, the present work is aimed to explore the test-retest reliability and practice effects associated with two alternate SDMT forms in the context of a short inter-assessment interval. A total of 123 university students completed the written SDMT version in two different time points separated by a 150 minutes interval. Half of the participants accomplished the same form in both occasions, whilst the other half filled different forms. Overall, reasonable test-retest reliabilities were found ($r = .70$), and the subjects that completed the same form revealed significant practice effects ($p < .001$, $d_z = 1.61$), which were almost non-existent in those that filled different forms. Thus, these forms were found to be moderately reliable and to elicit a similar performance across participants, suggesting their utility in repeated cognitive assessments when brief inter-assessment intervals are required.

Keywords: SDMT; practice effects; repeated assessment; test-retest reliability; alternate forms.

1. Introduction

Repeated neuropsychological assessments are necessary procedures in various clinical and research contexts. The evaluation of different cognitive functions over time can generate insightful data regarding, for example, the state and progression of a given clinical condition, possible improvements and/or declines, and also the impact of adopted interventions, including surgery, pharmacological treatments and cognitive rehabilitation. Even so a careful reading of the results between assessments is crucial since different factors can impact the subject’s performance. In this realm, practice effects constitute one of the most studied variables that can contribute to bias in repeated cognitive assessment. In simple terms, practice effects refer to an improvement in the task score from the first to the second or following applications of the test attributed solely to the task repetition (McCaffrey & Westervelt, 1995). Various factors such as comfort and familiarity with the test procedures, the development of learning strategies and the memorization of specific test stimuli can contribute to this performance enhancement.

In a general manner, when the same test version is applied in different occasions, improvements in the performance are probable (Benedict & Zgaljardic, 1998; Woods, Delis, Scott, Kramer, & Holdnack, 2006; Zgaljardic & Benedict, 2001) and can even endure one year after the baseline assessment (e.g., Basso, Bornstein, & Lang, 1999; Basso, Lowery, Ghormley, & Bornstein, 2001). Furthermore, practice effects tend to be larger between the first and the second evaluation (Baird, Tombaugh, & Francis, 2007; Beglinger et al., 2005; Collie, Maruff, Darby, & McStephen, 2003; Falleti, Maruff, Collie, & Darby, 2006; Monte, Geffen, & Kwapil, 2005; Register-Mihalik et al., 2012), and the performance can continue to improve in the subsequent time points although in a minor magnitude, or even reach a plateau (Bartels, Wegrzyn, Wiedl, Ackermann, &

Ehrenreich, 2010; Beglinger et al., 2005). Additionally, some tests can reveal larger practice effects than others. For instance, research shows that instruments which involve learning a specific rule or strategy, and those related to psychomotor processing speed and with non-verbal items show more gains in retesting evaluations in comparison with verbal oriented tests (Baird et al., 2007; Calamia, Markon, & Tranel, 2012; Watson, Pasteur, Healy, & Hughes, 1994). Even when using alternate forms, verbal tests seem more resilient to practice effects comparing to nonverbal tests (Benedict & Zgaljardic, 1998). Learning, memory and executive functions tasks are also prone to show practice effects (e.g., Basso et al., 1999; Basso et al., 2001; Lemay, Bédard, Rouleau, & Trembley, 2004; Mitrushina & Satz, 1991). The same happens in demanding and complex cognitive tasks, including tests where the development of a strategy is a key element (e.g., Basso et al., 1999; Basso et al., 2001). In contrast, instruments dedicated to explore functions like visual perception/recognition, naming and attention seem to have less influence from previous testing occasions (e.g., Mitrushina & Satz, 1991; Wilson, Watson, Baddeley, Emslie, & Evans, 2000).

Another important feature that can be explored throughout repeated assessments is the test-retest reliability (or temporal stability). It provides information about the degree of measurement error and the test score consistency, taking into account the stability of the subject’s ranking positions in the scores distribution across different assessment points (Duff, 2012). Usually the test-retest reliability is based on the correlation of the scores obtained in the same test, by the same subject, in two distinct occasions (Anastasi & Urbina, 1997). If the correlation is strong and significant, the test is considered to have little change over time and good test-retest reliability. The amount of time between assessments has an impact in this psychometric quality and, as a result, longer test-retest intervals seem to be linked with decreases in the magnitude of the correlation between test and retest evaluations (Calamia, Markon, & Tranel, 2013; McCaffrey & Westervelt, 1995). It is important to note that possible practice effects are not considered in the test-retest reliability measurements (McCaffrey & Westervelt, 1995), therefore we can have a test with a good correlation coefficient that reveals concomitantly a significant overall score change between assessments.

The present study aims to explore the practice effects and test-retest reliability of two recently developed alternate forms of the Symbol Digit Modalities Test (SDMT; Smith, 1982) in the context of a short inter-assessment interval. In a general manner, the SDMT is a commonly used measure of information processing speed, entailing other components such as attention, working memory, visual scanning and tracking (Strauss, Sherman, & Spreen, 2006). This task has been applied broadly in different clinical conditions, including Traumatic Brain Injury (TBI) and Multiple Sclerosis (MS), since it is a sensitive measure of information processing speed alterations (e.g., Draper & Ponsford, 2008; Forn, Belenguer, Parcet-Ibars, & Ávila, 2008). Specifically in MS, where deficits in information processing speed appear to be a hallmark (Batista et al., 2012; Forn et al., 2008; Huijbregts, Kalkers, Sonneville, Groot, & Polman, 2006), the SDMT is incorporated in different neuropsychological batteries of reference (e.g., Rao's Brief Repeatable Neuropsychological Battery - BRNB; Minimal Assessment of Cognitive Function in MS - MACFIMS; Brief International Assessment of Cognition for MS - BICAMS; Benedict et al., 2002; Langdon et al., 2012). In this context, the SDMT shows high sensitivity to detect cognitive alterations in MS (Dusankova, Kalincik, Havrdova, & Benedict, 2012; Glanz, Healy, Hviid, Chitnis, & Weiner, 2012; Portaccio et al., 2009; Van Schependom et al., 2014).

As aforementioned the SDMT assumes a relevant role in diverse research and clinical fields. Nonetheless, the development of alternate forms for repeated testing and the study of its psychometric characteristics remain scarce (Benedict et al., 2012). Additionally, the exploration of these SDMT properties is also limited when considering short inter-assessment intervals. Indeed, shorter inter-test intervals are required to evaluate possible cognitive changes occurring within hours or minutes. Specific examples include fatigue studies (e.g., Johnson, Lange, Deluca, Korn, & Natelson, 1997), clinical investigations with pharmacological agents, in which few hours are needed for the medication to reach the peak effect (e.g., Pietrzak, Snyder, & Maruff, 2010), or studies exploring the impact of surgical interventions in cognition (e.g., cardiac surgery; Bruggemans, Van de Vijver, & Huysmans, 1997; Lewis, Maruff, Silbert, Evered, & Scott 2006). Considering on one hand the lack of psychometric characterization of SDMT alternate forms and, on other hand, the dearth of SDMT data for short inter-assessment intervals, we planned a simple experimental design in which two recently developed SDMT alternate forms (see Benedict et al., 2012) were tested in the context of a brief inter-assessment interval in a group of healthy subjects. It is noteworthy that preliminary data obtained under healthy good-performance conditions can produce relevant psychometric information and clarify the role of specific variables in the performance of neuropsychological tests. Thus, this approach can also contribute to a more attentive design, administration and interpretation of results in future clinical studies (Kendall & Sheldrick, 2000). In table 1, we selected and summarized some data from studies using the SDMT with repeated assessment designs and healthy groups. In these cases, it is possible to observe that different versions and inter-assessment intervals have been implemented, ranging from one week to one year. The reported test-rest reliabilities tend to vary between 0.72 and 0.98, supporting a reasonably high temporal stability. Moreover, previous research already showed that the SDMT is susceptible to practice effects (e.g., Levine, Miller, Becker, Selnes, & Cohen, 2004;

Register-Mihalik et al., 2012) and that the use of available alternate forms seems to have a positive impact in controlling for this factor (e.g., Register-Mihalik et al., 2012). In this sense, similar outcomes were anticipated for this study, and we expected to extend these findings considering short inter-assessment intervals and also the use of Benedict's alternate forms.

Insert table 1 about here

2. Methods

2.1. Participants

The study conduction was approved by the local ethical committee for research in the health sciences (Ethics Subcommittee for Life and Health Sciences of University of Minho). A total of 123 healthy university students collaborated in the present study, 77 (62.6%) females and 46 (37.4%) males, aged between 19 and 37 years old (*M* = 22.4, *SD* = 3.54, 16% above 25 years old), 117 right-handed and 6 left-handed, and with 14.9 average years of formal education (*SD* = 1.96). The subjects were recruited during a class at the School of Health Sciences, University of Minho, Portugal. There were no reports of neurologic and/or psychiatric conditions, abusive consumption of substances, such as alcohol and drugs with known impact in the cognitive functioning, and no presence of sensorial/motor variations with significant interference in the test performance.

2.2. Materials

Symbol Digit Modalities Test (SDMT)

The SDMT (Smith, 1982) was created as a measure of cognitive screening for children and adults. It is a substitution task that covers diverse neurocognitive functions, including information processing speed, psychomotor functioning, attention, working memory and visual scanning (Strauss et al., 2006). This test has two possible ways of administration, one written that can be used for individual and group settings, and one oral for individual administrations and for subjects with motor complications. The test requires the substitutions of random geometric figures for a specific number, according to a key that contains 9 different geometric designs paired with a single 1 to 9 arabic number. In the written version, the one used in this study, a sheet of paper with the key on the top, 120 blank boxes paired with one specific design, and 10 blank boxes for initial practice is presented to the subjects, followed by the instructions. Individuals have 90 seconds to complete the blank boxes with the expected number of the key, and they are instructed to work as fast and accurate as possible. The SDMT takes, approximately, 5 minutes to administer, and the score corresponds to the total correct substitutions accomplished within the 90 seconds. The score ranges from 0 to 120, with higher scores pointing to a better performance. This is a simple test, not time consuming, easily scored, and it is also a well-accepted measure for different subject groups, including clinical groups (Berrigan et al., 2014; Possa, 2010; Rogers & Panegyres, 2007; Walker et al., 2012). As abovementioned, two particular alternate forms developed by Benedict and colleagues (2012) and shown to be equivalent to the Smith's original version (1982) were used (WPS Publishing, Torrance, CA, USA).

2.3. Procedure

After giving written informed consent, participants completed a brief questionnaire with relevant personal and medical information, such as years of formal education, occupation, handedness, relevant diseases and chronic medication use. This initial self-report facilitated the screening of identifiable neurological and psychiatric conditions, sensory-motor alterations and pharmacological treatments with possible interference in the subjects' performance. The experimental design consisted of two evaluations separated by, approximately, 150 minutes. In each occasion, the participants were asked to perform one of the written alternate forms: form 1 or form 2. The forms presentation was counterbalanced across participants and a 2 x 2 design was used implying two testing conditions: same condition, in which half of the participants filled the same version - 1 or 2 - in both occasions, and different condition where the remaining half undertook distinct forms, form 1 in the first assessment and form 2 in the second or vice-versa. In both time points, the SDMT instructions were presented according to the SDMT manual (Smith, 1982) simultaneously to all the participants, which were also asked to complete the first 10 training items. Following this initial stage, participants completed one of the two forms during 90 seconds. Between the two testing phases, participants were engaged in their regular classes.

2.4. Statistical analysis

For the analysis, subjects were aggregated according to the two previously described conditions, same and different, since the alternate forms had been shown to be equivalent (Benedict et al., 2012). All the SDMT values reported are based in the raw scores obtained from the total number of correct substitutions.

A mixed design ANOVA was performed with condition as the between-subjects factor and the testing session as the within-subjects factor, in order to explore potential main and interaction effects between the two described factors on practice effects. Statistical significant interactions were further explored by using the appropriate t tests.

Forms equivalence was tested by comparing mean scores at baseline with an independent samples t test. Test-retest reliabilities of both forms and between forms were obtained by computing Pearson’s product-moment correlations coefficients across the two time points.

As convention, the results were considered statistically significant for $p < .05$, and when this significant condition was reached, partial eta squared (η^2_p) for ANOVA, and Cohen’s *d* for t-tests were reported as measures of effect size. The statistical procedures were performed with the IBM Statistical Package for the Social Sciences Statistics software for windows, version 22 (IBM SPSS Statistics for Windows, Armonk, NY, USA).

3. Results

Subjects assessed with the same and different forms in the two occasions were similar regarding age, gender and years of formal education. Of note, the groups that completed a same or a different form had a similar distribution in terms of age [$t_{(117)} < 1$, $p = .36$, $d = 0.17$], years of education [$t_{(121)} = 1.66$, $p = .099$, $d = 0.30$] and sex [$X^2_{(1, N = 123)} < 1$, $p = .415$, $\varphi = 0.07$]. A summary of the main sociodemographic characteristics of the participants in the two conditions can be consulted in Table 2.

Insert Table 2 about here

In the context of practice effects examination, the mixed design ANOVA revealed a significant main effect of test session, $F_{(1,121)} = 100.04$, $p < .001$, $\eta^2_p = .45$, and a main effect of condition, $F_{(1,121)} = 10.76$, $p = .001$, $\eta^2_p = .08$. More importantly, there was a significant condition x test session interaction, $F_{(1,121)} = 82.81$, $p < .001$, $\eta^2_p = .41$, revealing that performance across evaluations varied according to the condition.

Insert Table 3 about here

Indeed, while there were statistically significant differences between the two occasions in the same condition, $t_{(60)} = -12.54$, $p < .001$, $d_z = 1.61$, performance in the different condition group was not significantly different, $t_{(61)} = -0.69$, $p = .49$. This implies that practice effects are important when similar forms are used, but mitigated by the use of alternate forms between assessments (see Table 3 and Figure 1). It is noteworthy that, in the first assessment session, no significant differences were found between the participants in the same group ($M = 60.67$, $SD = 9.57$) and those in the different group ($M = 61.92$, $SD = 10.83$; $t_{(121)} = -0.676$, $p = .50$). Whereas in the second evaluation session, the same group revealed a better performance ($M = 75.71$, $SD = 13.16$) than the different group ($M = 62.63$, $SD = 9.71$; $t_{(121)} = 6.28$, $p < .001$, $d = 1.13$). Additionally and supporting previous data that these forms are equivalent, performance in the first assessment for both forms was similar (form 1: $M = 62.25$, $SD = 9.80$; form 2: $M = 60.37$, $SD = 10.58$; $t_{(121)} = 1.02$, $p = .31$).

Insert Figure 1 about here

Test-retest reliability was analyzed separately according to experimental condition, different or same (see Table 3). Importantly, reliability was at similar level for the group of participants which completed different forms ($r = .70$, $p < .001$) and the group exposed to the same form in both occasions ($r = .70$, $p < .001$; see Figure 1).

Overall, these data indicate that both forms are moderately reliable when using the same or alternate forms throughout distinct assessment time points.

4. Discussion

Brief test-retest intervals of hours up to few days have been implemented in diverse contexts to explore possible cognitive changes occurring within a short span of time or to attenuate the impact of some day-by-day variable factors with known influence in the test performance (Falleti et al., 2006). In spite of this, test properties under such short repeated administrations are still poorly characterized. In the present work, we assessed the practice effects and test-retest reliabilities of two the SDMT alternate forms in a group of university students using 2.5 hours inter-test interval.

Our results showed significant practice effects when the same form was administered, but not when participants undertook different, although equivalent forms. These results are in line with previous investigations that report practice effects in the SDMT when similar forms are applied, even with longer test-retest intervals (e.g., Erlanger et al., 2014; Hinton-Bayre et al., 1997; Levine et al., 2004; Register-Mihalik et al., 2012). More importantly, they also extend to a much shorter test-retest interval previous observations of attenuated practice effects with the application of SDMT alternate forms (e.g., Register-Mihalik et al., 2012). In the specific case of brief test-retest intervals, it would be expected that the use of alternate forms would be less effective in controlling practice effects, especially because the participants can recall similar test features, including instructions and test materials. Concerning this point, it is important to note that test itself comprise an initial training period, giving a first opportunity to familiarize with the test procedures. Therefore, the possible contribution of this factor to the practice effects, even when alternate forms are used, is probably stabilized from the beginning. Notably the mitigation of the expected performance improvement when alternate forms are used supports the notion that item-specific practice has an important role in the SDMT associated practice effects. As a result, when the items are slightly modified in alternate forms, it is possible to attenuate a significant performance enhancement due solely to item-specific learning. Thus, our data support the notion that alternate forms are relevant to diminish practice effects (Benedict & Zgaljardic, 1998), especially the ones associated with item-specific training (Calamia et al., 2012; Woods et al., 2006; Zgaljardic & Benedict, 2001). Even so, there are other strategies to control for practice effects worthy to mention, including: (a) the edification of dual base lines, in which the subject do enough practice trials to establish a stabilized, pre-baseline performance (e.g., Duff, Westervelt, McCaffrey, & Haase, 2001; Watson et al., 1994); (b) inclusion of a paired control group (Watson et al., 1994); (c) implementation of statistical procedures designed to have in consideration changes related to practice. In this last case, the Reliable Change Index (RCI) has been widely used (see Lewis et al., 2006), since it provides information about how big a difference between two evaluations must be in order to consider a change as clinically relevant (RCI values were also calculated for this study and can be consulted in the Supplementary Material section). The adaptation and combination of these proposals according to the nature of each situation seems the most careful approach for dealing with possible practice effects when interpreting the results obtained from repeated assessments.

Regarding the test-retest reliability, it would be expected to be >.70 as recommended (Burlingame, Lambert, Reisinger, Neff, & Mosier, 1995), especially because we used a short inter-assessment interval that is theoretically associated with higher reliability coefficients (Slick, 2006). Although considering a .70 correlation as reasonable, the values reported by other studies with healthy participants tend to be above .70 (see table 1), and in the specific case of the Benedict's and colleagues work (2012) that used the same alternate forms also in healthy subjects, they report a coefficient of .86. One of the possible explanations for this finding resides in the demographic specificities of our sample, since some groups can show a more variable pattern across time than others (Slick, 2006) and this has a reflection in terms of reliability. In table 1, if we look for the studies with samples of highly educated adults with mean age below 40 (e.g., Goretti et al., 2014; Hinton-Bayre et al., 1997; Levine et al., 2004; Register-Mihalik et al., 2012; Smith, 1982), even using distinct inter-assessment intervals and different SDMT forms and versions, it is possible to verify Pearson's Correlation Coefficients between .62 and .82, so there is some variability for this task and the value we found here can be viewed as satisfactory.

In this line of thought, it is important to recognize that variables such as cultural, ethnical, educational and age-related factors can play an essential role in the test performance. Nevertheless, investigations aimed at a better clarification of the conjugated or independent contributions of these variables in the field of repeated testing performance tend to show different results according to the neuropsychological tasks and the population cohorts. Concerning the SDMT, the results tend to be controversial. While some studies emphasize no major impact of variables like age and education in its results (e.g., Sheridan et al., 2006), others point to strong correlations between SDMT performance and the mentioned variables (e.g., Harris, Wagner, & Cullum, 2007; Vogel, Stokholm, & Jørgensen, 2013). In the case of practice effects, a study of Duff and colleagues (2012) revealed no significant correlation between the performance in SDMT and different demographic, clinical variables (e.g., age; formal education; depression; global cognition). These issues were not explored in the present study, which is an imperative limitation. In this sense, the results obtained here, derived from a group of healthy young adults highly educated, may not be generalized for other populations with different characteristics and also for subjects with a given clinical condition. Additionally, we used the written version and a group administration context, so it is not possible to perceive in what extent our results are applicable to the SDMT oral version, nor to individual testing settings. Even so, it is important to note that the raw scores obtained at baseline are close to the ones reported by other studies in different cultural settings especially with younger highly educated subjects (e.g., Bate, Mathias, & Crawford, 2001; Goretti et al., 2014; Jorm, Anstey, Christensen, & Rodgers, 2004; Nissley & Schmitter-Edgecombe, 2002; see also Tables 1 and 3). Accordingly, the results found here can be a possible addition to the normative data specifically for the population cohort of European Portuguese university students.

Another important limitation regards the number of conducted assessments and the number of alternate forms used. On one hand, the results found here support that the two alternate forms created by Benedict and colleagues (2012) are moderately reliable considering a short inter-assessment interval but, on other hand, other forms and related practice effects could be tested, including the original version by Smith (1982), the Hinton-Bayre and colleagues alternate forms (1997) and the BRNB versions. Moreover, in the research and in clinical practice, it is common to have various repeated evaluations across time. Thus, it would be important to test these SDMT alternate forms in distinct brief and long inter-assessment intervals, with more evaluation time points, in order to get some approximations with the diversity and emergent needs present in different clinical/research context. Another point that needs additional exploration is the possible association between test-retest interval length and the magnitude of the practice effects. In this context, some studies already show no differences between the practice effects obtained in various inter-assessment times (e.g., Baird et al., 2007; Hinton-Bayre & Geffen, 2005); for instance, in the study of Baird and colleagues (2007), the practice effects were similarly noticeable for test-retest intervals of 3 months, 1 week and 20 minutes.

As final remarks, the findings presented here support that the SDMT alternate forms used (Benedict et al., 2012) are moderately reliable and equivalent, suggesting its usefulness for serial neuropsychological evaluations. Even considering that the results were extracted from a specific healthy cohort of participants, this study gathers some data regarding the scores stability and practice effects of two SDMT alternate forms in the context of a short inter-assessment interval. This information can be useful for specific normative comparisons (Kendall & Sheldrick, 2000) and for the design of future investigations with other population cohorts, including with clinical conditions. The development and psychometric study of SDMT alternate forms is crucial, since this test can be applied successfully in diverse ethnical and cultural populations (Harris et al., 2007; O'Bryant, Humphreys, Bauer, McCaffrey, & Hilsabeck, 2007). Similarly, it is a promising cognitive screening tool in different clinical conditions, including MS (Morrow, Jurgensen, Forrestal, Munchauer, & Benedict, 2011; Strober, Rao, Lee, Fischer, & Rudick, 2014) and TBI (Draper & Ponsford, 2008), in which repeated cognitive evaluations over time are essential. Our results also support the SDMT as a reliable instrument to administrate across brief test-retest periods, and this fact can be advantageous for surgical and pharmacological interventions that require such short intervals (e.g., Bruggemans et al., 1997; Lewis et al., 2006; Pietrzak et al., 2010). Nevertheless, more investigations are warranted to clarify how different properties of cognitive tests may change in the context of repeated assessment, including possible variations associated with practice effects. So it is pertinent to test brief or long inter-assessment time points, and different cultural, sociodemographic and clinical characteristics (McCaffrey & Westervelt, 1995; Putnam, Adams, & Schneider, 1992; Slick, 2006). Moreover, the practice effects can be studied, on one hand, to elucidate their impact in serial testing so that their influence is accounted for when significant cognitive changes are expected and, on other hand, as a measure of cognitive performance. More specifically, the absence or diminished development of expected practice effects has been suggested as an important marker of neuropsychological dysfunction (Duff et al., 2010). Overall, new and old neuropsychological instruments require consistent investigations regarding several psychometric characteristics, practice effects associated and even variations linked to different population cohorts and clinical groups.

5. References

Akbar, N., Honarmand, K., Kou, N., & Feinstein, A. (2011). Validity of a computerized version of the Symbol Digit Modalities Test in multiple sclerosis. *Journal of Neurology*, *258*(3), 373-379. doi:10.1007/s00415-010-5760-8

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, N.J.: Prentice Hall.

Baird, B. J., Tombaugh, T. N., & Francis, M. (2007). The effects of practice on speed of information processing using the Adjusting-Paced Serial Addition Test (Adjusting-PSAT) and the Computerized Tests of Information Processing (CTIP). *Applied Neuropsychology*, *14*(2), 88-100. doi:10.1080/09084280701319912

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience*, *11*, 118. doi:10.1186/1471-2202-11-118

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *13*(3), 283-292. doi:10.1076/clin.13.3.283.1743

Basso, M. R., Lowery, N., Ghormley, C., & Bornstein, R. A. (2001). Practice effects on the Wisconsin Card Sorting Test-64 Card version across 12 months. *The Clinical Neuropsychologist*, *15*(4), 471-478. doi:10.1076/clin.15.4.471.1883

Bate, A. J., Mathias, J. L., & Crawford, J. R. (2001). Performance on the Test of Everyday Attention and standard tests of attention following severe traumatic brain injury. *The Clinical Neuropsychologist, 15*(3), 405-422. doi:10.1076/clin.15.3.405.10279

Batista, S., Zivadinov, R., Hoogs, M., Bergsland, N., Heininen-Brown, M., Dwyer, M. G., …Benedict, R. H. B. (2012). Basal ganglia, thalamus and neocortical atrophy predicting slowed cognitive processing in multiple sclerosis. *Journal of Neurology, 259*(1), 139-146. doi:10.1007/s00415-011-6147-1

Beglinger, L. J., Gaydos, B., Tangphao-daniels, O., Duff, K., Kareken, D. A., Crawford, J., …Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology, 20*(4), 517-529. doi:10.1016/j.acn.2004.12.003

Benedict, R. H. B., Fischer, J. S., Archibald, C. J., Arnett, P. A., Beatty, W. W., Bobholz, J., …Munschauer, F. (2002). Minimal neuropsychological assessment of MS patients: A consensus approach. *The Clinical Neuropsychologist, 16*(3), 381-397. doi:10.1076/clin.16.3.381.13859

Benedict, R. H. B., Smerbeck, A., Parikh, R., Rodgers, J., Cadavid, D., & Erlanger, D. (2012). Reliability and equivalence of alternate forms for the Symbol Digit

Modalities Test: Implications for multiple sclerosis clinical trials. *Multiple Sclerosis Journal*, *18*(9), 1320-1325. doi:10.1177/1352458511435717

Benedict, R. H. B., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 339-352. doi:10.1076/jcen.20.3.339.822

Berrigan, L. I., Fisk, J. D., Walker, L. A. S., Wojtowicz, M., Rees, L. M., Freedman, M. S., & Marrie, R. A. (2014). Reliability of regression-based normative data for the oral symbol digit modalities test: An evaluation of demographic influences, construct validity, and impairment classification rates in multiple sclerosis samples. *The Clinical Neuropsychologist*, *28*(2), 281-299. doi:10.1080/13854046.2013.871337

Bruggemans, E. F., Van de Vijver, F. J. R., & Huysmans, H. A. (1997). Assessment of cognitive deterioration in individual patients following cardiac surgery: Correcting for measurement error and practice effects. *Journal of Clinical and Experimental Neuropsychology*, *19*(4), 543-559. doi:10.1080/01688639708403743

Burlingame, G. M., Lambert, M. J., Reisinger, C. W., Neff, W. M., & Mosier, J. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *Journal of Mental Health Administration*, *22*(3), 226-236. doi:10.1007/BF02521118

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: A meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*(4), 543-570. doi:10.1080/13854046.2012.680913

Calamia, M., Markon, K., & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test-retest correlations. *The*

Clinical Neuropsychologist, *27*(7), 1077-1205. doi:10.1080/13854046.2013.809795

Collie, A., Maruff, P., Darby, D. G., & McStephen, M. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *Journal of the International Neuropsychological Society*, *9*(3), 419-428. doi:10.1017/S1355617703930074

Draper, K., & Ponsford, J. (2008). Cognitive functioning ten years following traumatic brain injury and rehabilitation. *Neuropsychology*, *22*(5), 618-625. doi:10.1037/0894-4105.22.5.618

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology, 27*(3), 248-261. doi:10.1093/arclin/acr120

Duff, K. (2014). One-week practice effects in older adults: Tools for assessing cognitive change. *The Clinical Neuropsychologist*, *28*(5), 714-725. doi:10.1080/13854046.2014.920923

Duff, K., Beglinger, L. J., Moser, D. J., Paulsen, J. S., Schultz, S. K., & Arndt, S. (2010). Predicting cognitive change in older adults: The relative contribution of practice effects. *Archives of Clinical Neuropsychology*, *25*(2), 81-88. doi:10.1093/arclin/acp105

Duff, K., Callister, C., Dennett, K., & Tometich, D. (2012). Practice effects: A unique cognitive variable. *The Clinical Neuropsychologist*, *26*(7), 1117-1127. doi:10.1080/13854046.2012.722685

Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S., ...McCaffrey, R. J. (2011). Practice effects predict cognitive outcome in amnestic mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, *19*(11), 932-939. doi:10.1097/JGP.0b013e318209dd3a

Duff, K., Westervelt, H. J., McCaffrey, R. J., & Haase, R. F. (2001). Practice effects, test-retest stability, and dual baseline assessments with the California Verbal Learning Test in an HIV sample. *Archives of Clinical Neuropsychology*, *16*(5), 461-476. doi:10.1016/S0887-6177(00)00057-3

Dusankova, J. B., Kalincik, T., Havrdova, E., & Benedict, R. H. B. (2012). Cross cultural validation of the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS) and the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). *The Clinical Neuropsychologist*, *26*(7), 1186-1200. doi:10.1080/13854046.2012.725101

Erlanger, D. M., Kaushik, T., Caruso, L. S., Benedict, R. H. B., Foley, F. W., Wilken, J., …Deluca, J. (2014). Reliability of a cognitive endpoint for use in a multiple sclerosis pharmaceutical trial. *Journal of the Neurological Sciences*, *340*(1-2), 123-129. doi:10.1016/j.jns.2014.03.009

Falleti, M. G., Maruff, P., Collie, A., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState Battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1095-1112. doi:10.1080/13803390500205718

Forn, C., Belenguer, A., Parcet-Ibars, M. A., & Ávila, C. (2008). Information-processing speed is the primary deficit underlying the poor performance of multiple sclerosis patients in the Paced Auditory Serial Addition Test (PASAT). *Journal of Clinical and Experimental Neuropsychology*, *30*(7), 789-796. doi:10.1080/13803390701779560

Glanz, B. I., Healy, B. C., Hviid, L. E., Chitnis, T., & Weiner, H. L. (2012). Cognitive deterioration in patients with early multiple sclerosis: A 5-year study. *Journal of Neurology, Neurosurgery & Psychiatry, 83*(1), 38-43. doi:10.1136/jnnp.2010.237834

Goretti, B., Niccolai, C., Hakiki, B., Sturchio, A., Falautano, M., Eleonora, M., …Amato, M. P. (2014). The brief international cognitive assessment for multiple sclerosis (BICAMS): Normative values with gender, age and education corrections in the Italian population. *BMC Neurology, 14*(1), 171. doi:10.1186/s12883-014-0171-6

Harris, J. G., Wagner, B., & Cullum, C. M. (2007). Symbol vs. digit substitution task performance in diverse cultural and linguistic groups. *The Clinical Neuropsychologist*, *21*(5), 800-810. doi:10.1080/13854040600801019

Hinton-Bayre, A., & Geffen, G. (2005). Comparability, reliability, and practice effects on alternate forms of the digit symbol substitution and symbol digit modalities tests. *Psychological Assessment*, *17*(2), 237-241. doi:10.1037/1040-3590.17.2.237

Hinton-Bayre, A. D., Geffen, G., & McFarland, K. (1997). Mild head injury and speed of information processing: A prospective study of professional rugby league players. *Journal of Clinical and Experimental Neuropsychology, 19*(2), 275-289. doi:10.1080/01688639708403857

Huijbregts, S. C. J., Kalkers, N. F., Sonneville, L. M. J., Groot, V., & Polman, C. H. (2006). Cognitive impairment and decline in different MS subtypes. *Journal of Neurological Sciences*, *245*(1-2), 187-194. doi:10.1016/j.jns.2005.07.018

Johnson, S. K., Lange, G., Deluca, J., Korn, L. R., & Natelson, B. (1997). The effects of fatigue on neuropsychological performance in patients with chronic fatigue syndrome, multiple sclerosis, and depression. *Applied Neuropsychology*, *4*(3), 145-153. doi:10.1207/s15324826an0403_1

Jorm, A. F., Anstey, K. J., Christensen, H., & Rodgers, B. (2004). Gender differences in cognitive abilities: The mediating role of health state and health habits. *Intelligence, 32*(1), 7-23. doi:10.1016/j.intell.2003.08.001

Kendall, P. C. & Sheldrick, R. C. (2000). Normative data for normative comparisons. *Journal of Consulting and Clinical Psychology, 68*(5), 767-773. doi:10.1037//0022-006X.68.5.767

Langdon, D. W., Amato, M. P., Boringa, J., Brochet, B., Foley, F., Fredrikson, S., …Benedict, R. H. B. (2012). Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). *Multiple Sclerosis Journal*, *18*(6), 891-898. doi:10.1177/1352458511431076

Lemay, S., Bédard, M., Rouleau, I., & Trembley, P. L. G. (2004). Practice effect and test-retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neuropsychologist*, *18*(2), 284-302. doi:10.1080/13854040490501718

Levine, A. J., Miller, E. N., Becker, J. T., Selnes, O. A., & Cohen, B. A. (2004). Normative data for determining significance of test-retest differences on eight common neuropsychological instruments. *The Clinical Neuropsychologist*, *18*(3), 373-384. doi:10.1080/1385404049052420

Lewis, M. S., Maruff, P., Silbert, B. S., Evered, L. A., & Scott, D. A. (2006). The influence of different error estimates in the detection of post-operative cognitive dysfunction using reliable change indices with correction for practice effects. *Archives of Clinical Neuropsychology*, *21*(2), 421-427. doi:10.1016/j.acn.2006.05.004

McCaffrey, R. J., & Westervelt, H. J. (1995). Issues associated with repeated neuropsychological assessments. *Neuropsychology Review*, *5*(3), 203-221. doi:10.1007/BF02214762

Mitrushina, M., & Satz, P. (1991). Effect of repeated administration a neuropsychological battery in the elderly. *Journal of Clinical Psychology*, *47*(6), 790-801.

Monte, V. E., Geffen, G. M., & Kwapil, K. (2005). Test-retest reliability and practice effects of a rapid screen of mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 624-632. doi:10.1080/13803390490918589

Morrow, S. A., Jurgensen, S., Forrestal, F., Munchauer, F. E., & Benedict, R. H. B. (2011). Effects of acute relapses on neuropsychological status in multiple sclerosis patients. *Journal of Neurology, 258*(9), 1603-1608. doi:10.1007/s00415-011-5975-3

Nissley, H. M., & Schmitter-Edgecombe, M. (2002). Perceptually based implicit learning in severe closed-head injury patients. *Neuropsychology, 16*(1), 111-122. doi:10.1037//0894-4105.16.1.111

O'Bryant, S. E., Humphreys, J. D., Bauer, L., McCaffrey, R. J., & Hilsabeck, R. C. (2007). The influence of ethnicity on Symbol Digit Modalities Test performance: An analysis of a multi-ethnic college and hepatitis C patient sample. *Applied Neuropsychology, 14*(3), 183-188. doi:10.1080/09084280701508986

Pietrzak, R. H., Snyder, P. J., & Maruff, P. (2010). Amphetamine-related improvement in executive function in patients with chronic schizophrenia is modulated by practice effects. *Schizophrenia Research, 124*(1-3), 176-182. doi:10.1016/j.schres.2010.09.012

Portaccio, E., Goretti, B., Zipoli, V., Siracusa, G., Sorbi, S., & Amato, M. P. (2009). A short version of Rao's Brief Repeatable Battery as a screening tool for cognitive impairment in multiple sclerosis. *The Clinical Neuropsychologist*, *23*(2), 268-275. doi:10.1080/13854040801992815

Possa, M. F. (2010). Neuropsychological measures in clinical practice. *Neurological Sciences*, *31*, S219-S222. doi:10.1007/s10072-010-0374-6

Putnam, S. H., Adams, K. M., & Schneider, A. M. (1992). One-day test-retest reliability of neuropsychological tests in a personal injury case. *Psychological Assessment*, *4*(3), 312-316.

Register-Mihalik, J. K., Kontos, D. L., Guskiewicz, K. M., Mihalik, J. P., Conder, R., & Shields, E. W. (2012). Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries. *Journal of Athletic Training*, *47*(3), 297-305. doi:10.4085/1062-6050-47.3.13

Rogers, J. M., & Panegyres, P. K. (2007). Cognitive impairment in multiple sclerosis: Evidence-based analysis and recommendations. *Journal of Clinical Neuroscience*, *14*(10), 919-927. doi:10.1016/j.jocn.2007.02.006

Sheridan, L. K., Fitzgerald, H. E., Adams, K. M., Nigg, J. T., Martel, M. M., Puttler, L. I., …Zucker, R. A. (2006). Normative Symbol Digit Modalities Test performance in a community-based sample. *Archives of Clinical Neuropsychology*, *21*(1), 23-28. doi:10.1016/j.acn.2005.07.003

Slick, D. J. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. S. Sherman, & O. Spreen (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 3-43). New York: Oxford University Press.

Smith, A. (1982). *Symbol Digit Modalities Test*. Los Angeles, CA: Western Psychological Services.

Strauss, S., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Strober, L. B., Rao, S. M., Lee, J. C., Fischer, E., & Rudick, R. (2014). Cognitive impairment in multiple sclerosis: An 18 year follow-up study. *Multiple Sclerosis and Related Disorders, 3*(4), 473-481. doi:10.1016/j.msard.2014.03.004

Van Schependom, J., D'hooghe, M. B., Cleynhens, K., D'hooge, M., Haelewyck, M. C., De Keyser, J., & Nagels, G. (2014). The Symbol Digit Modalities Test as sentinel test for cognitive impairment in multiple sclerosis. *European Journal of Neurology, 21*(9), 1219-1225. doi:10.1111/ene.12463

Vogel, A., Stokholm, J., & Jørgensen, K. (2013). Performances on Symbol Digit Modalities Test, Color Trails Test, and modified Stroop test in a healthy, elderly Danish sample. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development*, *20*(3), 370-382. doi:10.1080/13825585.2012.725126

Walker, L. A. S., Cheng, A., Berard, J., Berrigan, L. I., Rees, L. M., & Freedman, M. S. (2012). Test of information processing speed: What do people with multiple sclerosis think about them? *International Journal of Multiple Sclerosis Care, 14*(2), 92-99. doi:10.7224/1537-2073-14.2.92

Watson, F. L., Pasteur, M. A. L., Healy, D. T., & Hughes, E. A. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology, 9*(1), 51-61. doi:10.1002/hup.470090107

Wilson, B. A., Watson, P. C., Baddeley, A. D., Emslie, H., & Evans, J. J. (2000). Improvement or simply practice? The effects of twenty repeated assessments on people with and without brain injury. *Journal of the International Neuropsychological Society*, *6*(4), 469-479.

Woods, S. P., Delis, D. C., Scott, J. C., Kramer, J. H., & Holdnack, J. A. (2006). The California Verbal Learning Test-second edition: Test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Archives of Clinical Neuropsychology*, *21*(5), 413-420. doi:10.1016/j.acn.2006.06.002

Zgaljardic, D. J., & Benedict, R. H. B. (2001). Evaluation of practice effects in language and spatial processing test. *Applied Neuropsychology*, *8*(4), 218-223. doi:10.1207/S15324826AN0804_4

Table 1. Brief Systematization of SDMT Test-retest Reliabilities and Scores Reported across Different Intervals for Healthy Subjects

Study	Participants: N of the sample, N of the most frequent sex, mean age (*SD*), mean years of formal education (*SD*)	SDMT version	Number of assessment sessions	Test-retest interval and test-retest reliability	Baseline raw score mean (*SD*)	Final assessment raw score mean (*SD*)
Smith, 1982	N = 80 48 women; Age = 34.8 (11.32); Education = 16.2 (2.50)	Original oral and written versions	2	*M* = 29.40 days Written SDMT: *r* = .80 Oral SDMT: *r* = .76	Written SDMT: 56.79 (9.84) Oral SDMT: 64.99 (11.91)	Written SDMT: 60.46 (11.16) Oral SDMT: 69.15 (11.97)
Hinton-Bayre, Geffen, & McFarland, 1997 (study 1)	N = 54 professional rugby players Age = 19.4 (2.1); Education = 12.3 (1.1)	Original version and 3 alternate forms	2	1-2 weeks Collapsed for all forms: *r* = .72	Collapsed for all forms: 55.8 (13.1)	Collapsed for all forms: 57.4 (11.5)
Levine et al., 2004	N = 1047 healthy male participants from which 465 completed the SDMT Age = 38.1 (7.8); Education = 16.3 (2.3)	Original version	2	*M* = 192 days (*SD* = 53) *r* = .80	57.3 (8.64)	59.6 (9.47)
Hinton-Bayre & Geffen, 2005	N = 112 semiprofessional athletes from which 31 did form 1 (Age = 19.7, *SD* = 3.2), 30 form 2 (Age = 21.1, *SD* = 4.0), 26 form 3 (Age = 20.5, *SD* = 3.3) and 25 form 4 (Age = 20.7, *SD* = 3.7)	Original written version and 3 alternate forms (from Hinton-Bayre et al., 1997)	2	1-2 weeks Form 1 and 2: ICC = .97 Form 1 and 3: ICC = .87 Form 1 and 4: ICC = .98 Form 2 and 3: ICC = .96 Form 2 and 4: ICC = .95 Form 3 and 4: ICC = .96	Form 1: 53.5 (9.6) Form 2: 55.8 (10.5) Form 3: 58.4 (9.9) Form 4: 58.0 (13.1)	n/a
Duff et al., 2010	N = 127 community-dwelling older adults 103 women; Age = 78.7 (7.8); Education = 15.5 (2.5)	n/a (the same version was used throughout moments)	3	1 week; 1 year	40.6 (12.4)	40.1 (11.7)
Akbar, Honarmand, Kou, & Feinstein, 2011	N = 119 participants with Multiple Sclerosis (MS); 38 healthy subjects MS: 90 women; Age = 44.7 (8.5); Education = 15.0 (2.2) Controls: 29 women; Age = 41.8 (11.0); Education = 15.9 (1.7)	Computerized version; oral paper version (from the Rao's Brief Repeatable Neuropsychological Battery - BRNB)	1 (2 for the temporal consistency calculations)	*M* = 103 days (*SD* = 16) ICC = .94 (this value was obtained from a randomly selected sub-sample of 17 MS participants)	Only for the oral administration MS: 45.1 (11.6) Controls: 57.6 (12.5)	n/a

(Continued)

Duff et al., 2011	N = 26 participants with amnestic Mild Cognitive impairment (MCI) with minimal Practice Effects (PE); 25 MCI with large PE; 57 cognitively intact MCI minimal PE: 19 women; Age = 83.2 (6.7); Education = 15.1 (2.1) MCI large PE: 22 women; Age = 81.6 (6.4); Education = 15.8 (3.0) Cognitively intact: 46 women; Age = 77.1 (7.9); Education = 15.4 (2.7)	n/a	2	1 week	MCI minimal PE: 32.5 (9.3) MCI large PE: 41.1 (8.8) Cognitively intact: 40.8 (7.8)	MCI minimal PE: 33.6 (10.4) MCI large PE: 42.2 (8.9) Cognitively intact: 44.2 (8.9)
Benedict et al., 2012	N = 25 19 women; Age = 42.0 (15.6); Education = 14.8 (1.9)	Original version; 2 alternate forms from the BRNB; 2 alternate forms created in the context of the study	5	Collapsed for all forms Between time 1 and 2: r = .84 Time 2 and 3: r = .86 Time 3 and 4: r = .89 Time 4 and 5: r = .90 Between new form 1 and new form 2 (used in our study): r = .86	Collapsed for all forms: 59.3 (11.7)	Collapsed for all forms: 64.9 (13.5)
Duff, Callister, Dennett, & Tometich, 2012	N = 268 community-dwelling older adults 211 women; Age = 73.3 (7.6); Education = 15.3 (2.6)	n/a (the same version was used throughout moments)	2	1 week	39.6 (9.3)	42.2 (10.1)
Register-Mihalik et al., 2012	N = 40 20 women and 20 men	Three distinct alternate forms	3	Between session 1 and 2: M = 1.8 days (SD = 0.61) Between session 2 and 3: M = 1.6 days (SD = 0.59) Session 1 and 2: r = .795 Session 2 and 3: r = .743 Session 1 and 3: r = .621	College group: 50.04 (14.53) High school group: 41.00 (5.85)	College group: 33.74 (0.22) High school group: 41.95 (5.94)
Duff, 2014	N = 167 community-dwelling older adults 136 women; Age = 78.6 (7.8); Education = 15.4 (2.5)	n/a	2	1 week r = .86	39.5 (9.5)	42.1 (10.1)
Goretti et al., 2014	N = 273 from which 243 completed the baseline and the retest assessment 180 women; Age = 38.9 (13.0); Education = 14.9 (3.0)	Oral version	2	r = .815	56.2 (11.6)	60.3 (12.0)

Note. BRNB = Brief Repeatable Neuropsychological Battery; ICC = Intra-class Correlation Coefficient; M = Mean; MCI = Mild Cognitive Impairment; MS = Multiple Sclerosis; n/a = Not Applicable/ Not Available; PE = Practice Effects; r = Pearson's Correlation Coefficient; SD = Standard Deviation; SDMT = Symbol Digit Modalities Test.

Table 2. Sum of the Main Sociodemographic Characteristics According to the Conditions Same and Different

	Condition same (*n* = 61)	Condition different (*n* = 62)
	M (SD)	*M (SD)*
Female/male %	62.3 / 37.7 %	62.9 / 37.1 %
Age	22.7 (3.88)	22.1 (3.17)
Years of formal education	15.2 (2.24)	14.6 (1.63)

Note. M = Mean; SD = Standard Deviation.

Table 3. Sum of the Main Results Concerning the Practice Effects and Test-retest Reliabilities Coefficients in the Same and in the Different Condition

Condition	Moments		Total	Test-retest reliability	$F_{(1,121)}$	η^2_p
	M1	M2				
	M (SD)	*M (SD)*	*M (SD)*			
Same	60.7 (9.57)	75.7 (13.16)	68.2 (1.28)	.70**	10.76*	.082
	$t_{(60)} = - 12.54^{**}$, $d_z = 1.61$					
Different	61.9 (10.83)	62.6 (9.71)	62.3 (1.27)	.70**		
	$t_{(61)} = - 0.69$ (*ns*)					
Total *M (SD)*	61.3 (0.92)	69.2 (1.04)	-	-	100.04**	.453

Note. M = Mean; M1 = Moment 1; M2 = Moment 2; ns = Non-significant; SD = Standard Deviation.
* $p < .01$ ** $p < .001$.

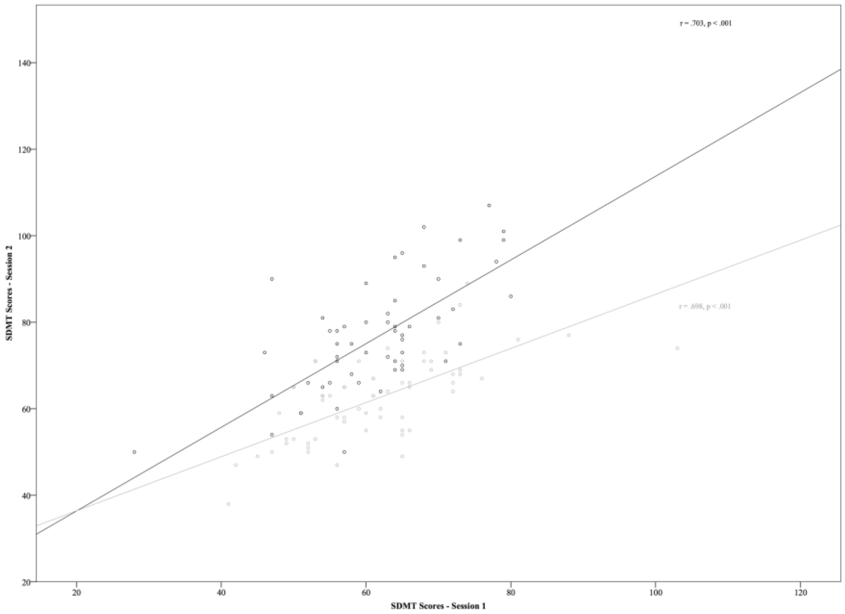

Figure 1. Scatterplot of the SDMT scores obtained in moment 1 and in moment 2 color-coded for condition same (black color) and condition different (grey color). Equations for the regression lines: y = 0.97x + 17.03 (same condition); y = 0.63x + 23.90 (different condition).

Supplementary data

Methods

Statistical analysis

Reliable change indices (RCI) were also planned and since there are several of possible calculations (e.g., Chelune, Naugle, Lüders, Sedlak, & Awad, 1993; Jacobson & Truax, 1991; Lewis et al., 2006) and according to the nature of the present study, the RCI cut-off calculations were based in the Chelune and colleagues (1993) approach. Thus, test-retest correlations, Standard Error of Measurement (SE_m) and Standard Error of Difference (SE_{diff}) were used for the RCI calculations. More specifically, the SE_m was extracted from the following formula - $S1(1 - r)^{1/2}$ - where the S1 corresponds to the standard deviation obtained at baseline and the r is the value of test-retest reliability; in the next phase, the SE_m was included in the SE_{diff} calculation according to the formula: $[2(SE_m)^2]^{1/2}$. The SE_{diff} was then multiplied by 1.96 of the standardized normal distribution to obtain the RCI considering a 95% confidence interval.

Results

The RCI results and relevant values included for the RCI calculations are summarized in Supplementary Table 1, separated by condition. Taking into account that the RCI values inform about how great a change in the subject score has to be in order to exceed possible random variation effects alone and be considered reliable, the results obtained here reveal that within this short time frame individuals have to change about ± 15 to 16 points from the baseline in their SDMT scores to reach a reliable improvement or decline. Regardless of the condition same or different, the RCI results can be observed as high, since a significant improvement means that the same subject has to complete more 15 correct substitutions than the ones accomplish at the first assessment moment.

Table 1. Sum of the Reliable Change Index Scores, Standard Error of Measurement and Standard Error of Difference for the Conditions Same and Different

	SE_m	S_{diff}	RCI (95%)
Condition same	5.24	7.41	± 14.52
Condition different	5.93	8.39	± 16.44

Note. RCI = Reliable Change Index; SE_{diff} = Standard Error of Difference; SE_m = Standard Error of Measurement.

References

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*(1), 41-52. doi:10.1037/0894-4105.7.1.41

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12-19. doi:10.1037/0022-006X.59.1.12