Published in final edited form as:

Clin Neuropsychol. 2014 July ; 28(5): 714–725. doi:10.1080/13854046.2014.920923.

One week practice effects in older adults: Tools for assessing cognitive change

Kevin Duff

Center for Alzheimer's Care, Imaging and Research, Department of Neurology, University of Utah

Abstract

Although neuropsychologists are frequently asked to evaluate cognitive change in a patient, this can be a complex determination. Using data from 167 non-demented older adults tested twice across one week, the current study sought to provide a variety of reliable change indices for a brief battery of commonly used neuropsychological measures. Statistically significant improvements were observed on 7 of 9 scores examined over this short retest interval, with the largest changes occurring on memory measures. Information is provided on simple discrepancy scores, standard deviation index, reliable change index (with and without correcting for practice effects), and standardized regression based change formulae for each cognitive score. Even though a one week retesting interval is less typical clinical scenario, these results may give clinicians and researchers more options for assessing short term change in a variety of settings.

Assessment of cognitive change over time is a common task for neuropsychologists, especially those working in geriatric settings. Declines in cognition can occur in older adults due to progressive neurological conditions (e.g., Alzheimer's disease, (Wilson et al., 2010), complications of surgical procedures (e.g., delirium following joint replacement, coronary artery bypass, (Lombard & Mathew, 2010), or exacerbations of chronic medical conditions (e.g., hypothyroidism, (Hogervorst, Huppert, Matthews, & Brayne, 2008). Additionally, stability or improvements can be due to intervention (e.g., medication, (Rosenberg et al., 2013); cognitive rehabilitation, (Smith et al., 2009). However, the actual determination of "real" and "meaningful" change in cognition is complex.

Due to the extant literature on practice effects on cognitive tests in intact individuals (Calamia, Markon, & Tranel, 2012; McCaffrey, Duff, & Westervelt, 2000), clinicians might employ one of several statistical formulas that utilize a test's reliability, documented practice effects, and other variables to evaluate if a change is "normal" or not. For example, the Reliable Change Index (RCI) of Jacobson and Truax (1991), practice-adjusted RCI (Chelune, Naugle, Luders, Sedlak, & Awad, 1993; Iverson, 2001), or the standardized regression based (SRB) formulas (McSweeney, Naugle, Chelune, & Luders, 1993) are all options for assessing if a "real" and "meaningful" change has occurred. Some studies have found that SRBs have greater sensitivity (Barr, 2002; Temkin, Heaton, Grant, & Dikmen,

1999), whereas others have noted that all of these change formulae yield comparable results (Heaton et al., 2001).

In a review of the assessment of neuropsychological change in the individual patient, Duff (2012) noted a number of obstacles facing the field with regard to these formulae. Specifically, it was stated that there is a need for more work with geriatric samples, more published RCIs and SRBs are needed for under-represented groups like dementia, and these formulae need to be more widely disseminated into the hands of clinicians. The current study sought to address some of these obstacles by providing multiple change formulae on a battery of commonly-used neuropsychological measures in a cohort of older adults. Based on existing studies (Calamia, et al., 2012; McCaffrey, et al., 2000), it was hypothesized that notable practice effects would occur across a short retest period. In the SRB models, it was also anticipated that baseline test scores would best predict follow-up test scores in these regression models (Duff et al., 2004; McSweeney, et al., 1993).

Method

Participants

One hundred sixty-seven community-dwelling older adults participated in the current study, and these participants have been previously described (Duff, Beglinger, et al., 2008). Briefly, these individuals were recruited from senior centers and independent living facilities to prospectively study practice effects in older adults. Their mean age was 78.6 (7.8) years and their mean education was 15.4 (2.5) years. Most were female (81.1%) and all were Caucasian. Premorbid intellect at baseline was average (Wide Range Achievement Test – 3 Reading: $M=107.8$ [6.2]), and they reported minimal depressive symptoms (30-item Geriatric Depression Scale: $M=4.2$ [3.4]). To be classified as amnesic MCI, all participants had to complain of memory problems. MCI participants had to have objective memory deficits (i.e., age-corrected scores at or below the 7th percentile on two delayed recall measures [described below] relative to their premorbid intellectual estimate). The 7th percentile is 1.5 standard deviations below the mean, which is a typical demarcation point for cognitive deficits in MCI (Petersen et al., 2001). Cognition was, otherwise, generally intact (i.e., non-memory age-corrected scores above the 7th percentile) and no functional impairments (e.g., assistance needed with managing money, taking medications, driving) could be reported. To be classified as “cognitively intact,” all objective memory and non-memory performances were at least above the 7th percentile. All data was reviewed by two neuropsychologists. Seventy-four individuals (44% of sample) were classified with amnesic MCI according to these criteria and the remainder was classified as cognitively intact. No one was classified as demented (i.e., both impaired memory and other cognitive domains). All classifications were made following the one-week visit, so examiners were “blinded” to baseline and one-week data. However, only baseline cognitive performances were used in these classifications.

Procedures

All participants provided informed consent prior to participation, and all procedures were approved by the local Institutional Review Board. During a baseline visit, all participants

completed the following measures: Brief Visuospatial Memory Test – Revised (BVM-T-R), Hopkins Verbal Learning Test – Revised (HVLT-R), Controlled Oral Word Association Test (COWAT), animal fluency, Trail Making Test Parts A and B (TMTA and TMTB), Symbol Digit Modalities Test (SDMT), Wide Range Achievement Test – 3 (WRAT-3) Reading subtest, and the 30-item Geriatric Depression Scale (GDS). After one week, the battery was repeated, with the exception of the WRAT-3 Reading subtest. Alternate forms were purposefully not utilized on re-evaluation, as the study sought to maximize practice effects. WRAT-3 Reading scores were age-corrected standard scores using normative data from the test manual. All other values are raw scores.

Data Analyses

To assess for practice effects across one week, dependent t-tests were calculated for each of the nine cognitive variables from the repeated battery: BVM-T-R Total Recall (i.e., learning across three trials), BVM-T-R Delayed Recall, HVLT-R Total Recall, HVLT-R Delayed Recall, COWAT total words across three 60" trials, animal fluency total words across one 60" trial, TMTA seconds to completion, TMTB seconds to completion, and SDMT total correct in 90". These t-tests compared baseline and one week scores.

The following reliable change scores were calculated based on the formulae in Table 3 of Duff (2012): simple discrepancy score, standard deviation index, RCI, RCI correcting for practice effects, RCI correcting for practice effects and using the Iverson's standard error of the difference, and both simple and complex SRB change scores. For the simple SRBs, only the baseline score was used to predict the one week score. For the complex SRBs, baseline scores and demographic variables (age, education, gender) were used to predict one week scores. It was decided to include both cognitively normal and MCI participants into the same regression models to increase the range of test scores. However, interested readers can obtain these same results based on only the cognitively intact participants by contacting the author. The data were screened for univariate and multivariate outliers with boxplots, standardized scores (not to exceed ± 3.0), and Mahalanobis distances ($p < 0.001$). Linearity and multicollinearity were assessed with scatterplots and Variance Inflation Factors (not to exceed 2.5). Normal probability plots (P-P) were also examined for distribution of error.

Results

As can be seen in Table 1, 7 of the 9 participant's one week scores were significantly larger than their scores at baseline: BVM-T-R Total Recall ($t[166]=19.5$, $p < 0.001$), BVM-T-R Delayed Recall ($t[166]=13.9$, $p < 0.001$), HVLT-R Total Recall ($t[166]=14.6$, $p < 0.001$), HVLT-R Delayed Recall ($t[166]=11.7$, $p < 0.001$), TMTA ($t[166]=5.1$, $p < 0.001$), TMTB ($t[165]=4.1$, $p < 0.001$), and SDMT ($t[164]=5.7$, $p < 0.001$). COWAT and animal fluency scores did not significantly increase across one week (p 's = 0.06 and 0.13, respectively). Test-retest correlations between baseline and one week scores for all scores are also presented in Table 1, as are the mean practice effects (i.e., one week – baseline) and standard deviation of this difference score.

Table 2 presents simple discrepancy scores at various frequencies for all nine cognitive scores. For example, on the BVM-T-R Total Recall score, an "average" amount of change

across one week (i.e., 50%) is approximately 8 raw points. However, very few individuals lose points across one week (e.g., approximately 5% of the sample scored 1 point lower after one week) and few individuals gain a lot across this interval (e.g., approximately 5% of the sample scored 16 points higher after one week).

The standard deviation index is the simple discrepancy score divided by the baseline standard deviation of the test's score in Table 1. For example, on SDMT, if a participant scored 35 at baseline and 42 at one week, then his/her simple discrepancy score would be 7 (i.e., $42 - 35 = 7$), and his/her standard deviation index would be 0.74 (i.e., $7 / 9.5 = 0.74$). This roughly reflects how much change across one week he/she made compared to the variability of the baseline scores. Standard deviation indexes for all cognitive scores are presented in Table 3.

Table 3 also presents three RCIs: RCI, RCI correcting for practice effects, and RCI correcting for practice effects using Iverson's standard error of the difference. All three of these start with a simple discrepancy score and some form of the standard error of the difference. The standard error of the difference is essentially the standard deviation of the difference scores, and it includes information about the standard deviation of the score at baseline and the test-retest correlation. In the original RCI, Jacobson and Truax (1991) divide the simple discrepancy score by the standard error of the difference. In the RCI correcting for practice, Chelune et al. (1993) subtract practice effects from the simple discrepancy score before dividing by the standard error of the difference. Finally, Iverson (2001) modified the standard error of the difference to include information about the standard deviation of the scores at baseline and follow-up.

Tables 4 and 5 present the simple and complex SRBs, respectively, for all nine scores.

Discussion

In an attempt to address some of the obstacles in assessing cognitive change in older adults (Duff, 2012), the current study examined one week practice effects in a large cohort of elderly participants. Statistically significant practice effects were observed across this brief retest period on 7 of the 9 cognitive scores. These findings are consistent with those observed in older patients across varying retest intervals (Benedict & Zgaljardic, 1998; Blasi et al., 2009; Calamia, et al., 2012; Cooper, Lacritz, Weiner, Rosenberg, & Cullum, 2004; Ivnik et al., 1999; Knight, McMahon, Green, & Skeaff, 2006). For example, across one week, participants significantly improved on measures tapping verbal and visual learning and memory, processing speed, and set shifting. However, significant improvements did not occur on the two tests of verbal fluency, neither phonemic nor semantic. These results provide clinicians with information about how much change should occur in older patients when they are repeatedly tested across short intervals.

When comparisons were made between practice effects on the different cognitive tests in the current battery, some interesting findings were observed. First, the largest effect sizes were seen on the two memory tests. Additionally, the effect sizes were slightly larger on BVMT-R (Total and Delayed Recall averaged = 0.89) than the HVLTR (0.78). Effect sizes were

also somewhat larger on initial learning (BVMT-R and HVLT-R averaged = 0.92) than delayed recall (0.75). This latter finding may be due to the greater range of scores on learning trials compared to delayed recall trials. Second, smaller effect sizes were consistently seen on the three speeded psychomotor tests (0.27 – 0.28). Third, negligible effects were seen on the two verbal fluency tests. This latter finding is conflicts with Zgaljardic and Benedict (2001), who found that verbal fluency tests did show significant practice effects across four testing sessions, regardless of whether the same form or an alternate form was used. These authors described how rule-based learning can be test-specific and lead to comparable improvements despite the use of alternate forms. One reason for the difference in findings may be sample characteristics. The current sample was nearly 15 years older than those of Zgaljardic and Benedict's participants, and the current sample contained almost half who were classified as amnesic MCI. Fourth, the short-term test-retest correlations were relatively high (0.70 – 0.86) in the current study. These compare quite closely to those reported in the HVLT-R/BVMT-R Professional Manual Supplement (Benedict & Brandt, 2007) in adults across longer retest intervals (e.g., BVMT-R Total Recall $r = 0.80$ and Delayed Recall $r = 0.79$ across 56 days with the same test form). However, it is expected that with much longer retest intervals (e.g., 1+ years) that these correlations would drop (Duff, 2012).

In addition to quantifying the amount of practice observed in these participants, data was used to generate change formulas that can be used by clinicians and researchers when examining change in an individual patient or subject. For example, Table 1 provides the baseline standard deviation of the test's score, which is needed to calculate the standard deviation index. Although this tends to be an imprecise index of change, it is occasionally used when no other data is available. Similarly, Table 2 presents simple discrepancy scores, another gross index of change, but one that is relatively easy to calculate and understand.

More precise indexes of change are seen in Table 3, which presents three RCIs. All of these formulae yield z-scores, which reflect the amount of change of the individual compared to normative data. Z-scores that exceed some cutoff (e.g., ± 1.96) are typically viewed as being a "reliable change." The original RCI (Jacobson & Truax, 1991) is rarely used with neuropsychological scores, as it does not correct for the known practice effects on many cognitive measures. The two RCIs that do correct for practice effects (Chelune, et al., 1993; Iverson, 2001) have become the most popular indicators of cognitive change in the field. Of these, the Iverson formula may be preferred, as it corrects the standard error of the difference for the variability in scores at both time points (i.e., T_1 and T_2). However, studies to demonstrate the clear superiority of one RCI over the others have not yet been done.

Finally, Tables 4 and 5 present the simple and complex SRBs, respectively. Unlike the aforementioned change scores that compare the difference between baseline and follow-up, SRBs compare the difference between an observed follow-up score and a predicted follow-up score (which is predicted from the baseline score [simple SRB] and demographic variables [complex SRB]). Consistent with the existing literature (Duff et al., 2005; Duff, et al., 2004; Duff, Schoenberg, et al., 2008; Hermann et al., 1996; McSweeny, et al., 1993; Sawrie, Chelune, Naugle, & Luders, 1996; Temkin, et al., 1999), the best predictor of follow-up performance (i.e., one week scores) in our sample was initial performance (i.e.,

baseline scores) on that same measure. Across cognitive measures, baseline scores shared between 49 – 74% of the variance with one week scores in the simple SRBs. Complex SRBs added variance (50 – 76% in total). Although neither the simple nor complex SRBs captured the entirety of the one week scores, they are similar to those reported by others using patient and control samples.

A brief example of the use of these change formulas might be useful. A 70 year old patient completed TMTB in 195" at baseline and in 142" at one week. As noted in Table 6, the simple discrepancy score would be –53, which falls between the 84% and 95% of peers (based on Table 2). This patient's standard deviation index would be –0.93 (using the standard deviation of the baseline score in Table 1). Since TMTB is a timed test with lower scores indicating better performance, we would reverse the sign on this z-score, and see that it falls at the 82nd percentile on a normal distribution chart. The original RCI yields a z-score of –1.26 (reversed to +1.26), which falls at the 90th percentile. However, this original RCI does not correct for practice. Using Chelune et al.'s (1993) RCI, the patient's practice-corrected z-score is –0.97 (about 83rd percentile of reversed z-score). Including retest variability into the equation with Iverson's (2001) formula, the z-score is now –1.02 (84th percentile). The simple SRB, using only the baseline score to predict the one week score, leads to a z-score of –0.42 (66th percentile). Lastly, using baseline score and demographics to predict the one week score, the patient's z-score now changes to +0.22 (42nd percentile of reversed z-score). This example was not meant to identify the "best" change formula, but to demonstrate how these various formulae are calculated on a single individual. This does highlight the complexity of assessing cognitive change, even with the compliment of all available change scores. For example, the reader might notice that as the formulae got more sensitive (e.g., including reliability of test, baseline scores, and demographics), the z-score became less unusual (i.e., drifted closer to the 50th percentile).

Despite the obvious benefits of having these tools for assessing change in geriatric patients, there may be some concern about the retest interval in the current study, which was approximately one week (M=7.6 days [2.2]). Obviously, this may be a less typical clinical scenario. Individuals using the change information from this study need to use caution if their retest intervals significantly differ from this one week period. However, putting this caveat aside, it is not usual to have test-retest data across brief intervals. First, many test manuals report reliability data on intervals that range from days to weeks. So, these results can be used to compare psychometric properties to those from standardization samples. Second, there may be some clinical situations that do call for repeated assessment across brief intervals. For example, clinicians may wish to track short-term improvements within a rehabilitation setting. Benefits of fast acting medications may also be studied across a week or less (Rosenberg, et al., 2013). Rapidly evolving neurological conditions (e.g., delirium) may worsen or improve across a week or less. Third, other studies (Benedict & Zgaljardic, 1998; Darby, Maruff, Collie, & McStephen, 2002; Falsetti, Maruff, Collie, & Darby, 2006; Krenk, Rasmussen, Siersma, & Kehlet, 2012; Salthouse & Tucker-Drob, 2008) have examined short-term practice effects (e.g., across 10 minutes, a week, or a month) with the hopes of better isolating these improvements in test scores. When longer periods of time are utilized, extraneous, non-practice effects variables affect the observed changes. For example, when looking at practice effects over one year in elderly subjects, true cognitive

decline (both normal and abnormal) are influencing the magnitude and direction of "practice effects." Third, our own work has indicated that short-term practice effects hold unique information about cognition. For example, using practice effects across one week, we have found relationships with diagnosis (Duff, Beglinger, et al., 2008), prognosis (Duff et al., 2007; Duff et al., 2011), treatment response (Duff, Beglinger, Moser, Schultz, & Paulsen, 2010), and the amount of amyloid deposition in the brain (Duff, Foster, & Hoffman, in press). So, although we understand the limited immediate clinical utility of short-term practice effects, we also realize the potential that these change scores might be able to provide.

Another aspect of the current study that warrants comment is the composition of the sample used. Prior studies have tended to use relatively homogeneous samples to generate their change formulae. For example, in their original study on SRBs, McSweeney et al. (1993) used only seizure patients to develop change formulas. Conversely, Temkin et al. (1999) used only neurological healthy individuals to predict follow-up cognitive scores. The current study used both healthy elders and those classified with MCI. In some ways, these two sub-samples do reflect a single group: non-demented, community-dwelling elders. However, almost by definition, one group suffers from at least "mild" memory problems, whereas the other group does not. It was our intent to combine both sub-samples to increase the variability of cognitive scores, which increases the potential of developing change scores that would be applicable across a broad segment of the older adult population. In a related vein, Heaton et al (2001) has observed that SRBs and other change formulae developed on healthy samples might be less applicable in clinical samples, and these authors suggested that more diverse samples (e.g., neurological stable but not necessarily cognitively normal) be used to develop these change scores, so that a wide range of baseline and follow-up scores are represented. Our current sample seems to achieve this directive. Nonetheless, readers that want these change formulae on only the intact subjects can contact the author for this information.

Some limitations with the present study are acknowledged. First, as with most regression-based prediction formulas (Tabachnick & Fidell, 1996), they are less accurate estimates for individuals whose cognitive functioning falls at the extremes (e.g., <2nd percentile or >98th percentile) at baseline. In these cases, the prediction equations are more susceptible to regression to the mean and other fluctuations. However, the current study utilized a sample with both intact and impaired participants, which might lessen the chance of these statistical fluctuations. Caution should also be exercised when using these formulas outside the demographic and situational parameters of the sample (e.g., <65 or >96 years old; relatively brief or extended retest intervals; non-Caucasians; largely male samples). There is limited information in the field about the effects of gender, culture, or linguistics on repeated testing with various neuropsychological measures, and future investigations should examine these other demographic variables. Second, the current study focused on statistical methods for accounting for score changes in a repeated assessment situation. Alternate test forms could also be used to reduce artificial inflation of scores. However, some studies have found that alternate forms do not entirely control for practice effects (Beglinger et al., 2005; Zgaljardic & Benedict, 2001), and multiple methods may be needed to most accurately interpret change values. Third, although no stand-alone performance validity measures were administered as

part of this battery, an embedded indicator of effort (Silverberg, Wertheimer, & Fichtenberg, 2007) found the >97% of the sample seemed to provide adequate effort. Removing the few participants with questionable effort was considered, but it was ultimately decided to leave them in, as older individuals with memory impairments may perform poorly on this effort indicator (Hook, Marquine, & Hoelzle, 2009). Finally, the accuracy of these change scores need to be validated in an independent sample. Despite these limitations, the current results have the potential to provide more accurate assessments of short-term cognitive change in older adults by considering the influence of initial performance, practice effects, and other demographic variables.

Acknowledgments

The project described was supported by research grants from the National Institutes on Aging: K23 AG028417. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

References

- Barr WB. Neuropsychological testing for assessment of treatment effects: methodologic issues. *CNS Spectr.* 2002; 7(4):300–302. 304–306. [PubMed: 15122116]
- Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, Crawford J, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. *Arch Clin Neuropsychol.* 2005; 20(4):517–529. [PubMed: 15896564]
- Benedict, RHB.; Brandt, J. *HVLT-R/BVMT-R Professional Manual Supplement*. Lutz, FL: Psychological Assessment Resources; 2007.
- Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. *J Clin Exp Neuropsychol.* 1998; 20(3):339–352. [PubMed: 9845161]
- Blasi S, Zehnder AE, Berres M, Taylor KI, Spiegel R, Monsch AU. Norms for change in episodic memory as a prerequisite for the diagnosis of mild cognitive impairment (MCI). *Neuropsychology.* 2009; 23(2):189–200. [PubMed: 19254092]
- Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *Clin Neuropsychol.* 2012; 26(4):543–570. [PubMed: 22540222]
- Chelune GJ, Naugle RI, Luders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology.* 1993; 7(1):41–52.
- Cooper DB, Lacritz LH, Weiner MF, Rosenberg RN, Cullum CM. Category fluency in mild cognitive impairment: reduced effect of practice in test-retest conditions. *Alzheimer Dis Assoc Disord.* 2004; 18(3):120–122. [PubMed: 15494616]
- Darby D, Maruff P, Collie A, McStephen M. Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology.* 2002; 59(7):1042–1046. [PubMed: 12370459]
- Duff K. Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. *Arch Clin Neuropsychol.* 2012; 27(3):248–261. [PubMed: 22382384]
- Duff K, Beglinger L, Schultz S, Moser D, McCaffrey R, Haase R, et al. Practice effects in the prediction of long-term cognitive outcome in three patient samples: A novel prognostic index. *Arch Clin Neuropsychol.* 2007; 22(1):15–24. [PubMed: 17142007]
- Duff K, Beglinger LJ, Moser DJ, Schultz SK, Paulsen JS. Practice effects and outcome of cognitive training: preliminary evidence from a memory training course. *Am J Geriatr Psychiatry.* 2010; 18(1):91. [PubMed: 20104658]
- Duff K, Beglinger LJ, Van Der Heiden S, Moser DJ, Arndt S, Schultz SK, et al. Short-term practice effects in amnesic mild cognitive impairment: implications for diagnosis and treatment. *Int Psychogeriatr.* 2008; 20(5):986–999. [PubMed: 18405398]

- Duff K, Foster NL, Hoffman JM. Practice effects and amyloid deposition: Preliminary data on a method for enriching samples in clinical trials. *Alzheimer Disease & Associated Disorders*. (in press).
- Duff K, Lyketsos CG, Beglinger LJ, Chelune G, Moser DJ, Arndt S, et al. Practice effects predict cognitive outcome in amnesic mild cognitive impairment. *Am J Geriatr Psychiatry*. 2011; 19(11): 932–939. [PubMed: 22024617]
- Duff K, Schoenberg MR, Patton D, Paulsen JS, Bayless JD, Mold J, et al. Regression-based formulas for predicting change in RBANS subtests with older adults. *Arch Clin Neuropsychol*. 2005; 20(3): 281–290. [PubMed: 15797165]
- Duff K, Schoenberg MR, Patton DE, Mold J, Scott JG, Adams RA. Predicting change with the RBANS in a community dwelling elderly sample. *Journal of the International Neuropsychological Society*. 2004; 10:828–834. in press. [PubMed: 15637773]
- Duff K, Schoenberg MR, Patton DE, Mold JW, Scott JG, Adams RL. Predicting cognitive change across 3 years in community-dwelling elders. *Clin Neuropsychol*. 2008; 22(4):651–661. [PubMed: 17853149]
- Falletti MG, Maruff P, Collie A, Darby DG. 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. *J Clin Exp Neuropsychol*. 2006; 28(7):1095–1112. [PubMed: 16840238]
- Hermann BP, Seidenberg M, Schoenfeld J, Peterson J, Leveroni C, Wyler AR. Empirical techniques for determining the reliability, magnitude, and pattern of neuropsychological change after epilepsy surgery. *Epilepsia*. 1996; 37(10):942–950. [PubMed: 8822692]
- Hogervorst E, Huppert F, Matthews FE, Brayne C. Thyroid function and cognitive decline in the MRC Cognitive Function and Ageing Study. *Psychoneuroendocrinology*. 2008; 33(7):1013–1022. [PubMed: 18640783]
- Hook JN, Marquine MJ, Hoelzle JB. Repeatable battery for the assessment of neuropsychological status effort index performance in a medically ill geriatric sample. *Arch Clin Neuropsychol*. 2009; 24(3):231–235. [PubMed: 19549722]
- Iverson GL. Interpreting change on the WAIS-III/WMS-III in clinical samples. *Arch Clin Neuropsychol*. 2001; 16(2):183–191. [PubMed: 14590186]
- Ivnik RJ, Smith GE, Lucas JA, Petersen RC, Boeve BF, Kokmen E, et al. Testing normal older people three or four times at 1- to 2-year intervals: defining normal variance. *Neuropsychology*. 1999; 13(1):121–127. [PubMed: 10067783]
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991; 59(1):12–19. [PubMed: 2002127]
- Knight RG, McMahon J, Green TJ, Skeaff CM. Regression equations for predicting scores of persons over 65 on the rey auditory verbal learning test, the mini-mental state examination, the trail making test and semantic fluency measures. *Br J Clin Psychol*. 2006; 45(Pt 3):393–402. [PubMed: 17147104]
- Krenk L, Rasmussen LS, Siersma VD, Kehlet H. Short-term practice effects and variability in cognitive testing in a healthy elderly population. *Exp Gerontol*. 2012; 47(6):432–436. [PubMed: 22469814]
- Lombard FW, Mathew JP. Neurocognitive dysfunction following cardiac surgery. *Semin Cardiothorac Vasc Anesth*. 2010; 14(2):102–110. [PubMed: 20478950]
- McCaffrey, RJ.; Duff, K.; Westervelt, HJ. *Practitioner's Guide to Evaluating Change with Neuropsychological Assessment Instruments*. New York: Plenum/Kluwer; 2000.
- McSweeney AJ, Naugle RI, Chelune GJ, Luders H. T scores for change: An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist*. 1993; 7:300–312.
- Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, et al. Current concepts in mild cognitive impairment. *Arch Neurol*. 2001; 58(12):1985–1992. [PubMed: 11735772]
- Rosenberg PB, Lanctot KL, Drye LT, Herrmann N, Scherer RW, Bachman DL, et al. Safety and efficacy of methylphenidate for apathy in Alzheimer's disease: a randomized, placebo-controlled trial. *J Clin Psychiatry*. 2013; 74(8):810–816. [PubMed: 24021498]

- Salthouse TA, Tucker-Drob EM. Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*. 2008; 22(6):800–811. [PubMed: 18999354]
- Sawrie SM, Chelune GJ, Naugle RI, Luders HO. Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. *J Int Neuropsychol Soc*. 1996; 2(6):556–564. [PubMed: 9375160]
- Silverberg ND, Wertheimer JC, Fichtenberg NL. An Effort Index for the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). *Clin Neuropsychol*. 2007; 21(5):841–854. [PubMed: 17676548]
- Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. A cognitive training program based on principles of brain plasticity: results from the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) study. *J Am Geriatr Soc*. 2009; 57(4): 594–603. [PubMed: 19220558]
- Tabachnick, BG.; Fidell, LS. *Using Multivariate Statistics*. New York: HarperCollins Publishers; 1996.
- Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: a comparison of four models. *J Int Neuropsychol Soc*. 1999; 5(4):357–369. [PubMed: 10349298]
- Wilson RS, Aggarwal NT, Barnes LL, Mendes de Leon CF, Hebert LE, Evans DA. Cognitive decline in incident Alzheimer disease in a community population. *Neurology*. 2010; 74(12):951–955. [PubMed: 20308679]
- Zgaljardic DJ, Benedict RH. Evaluation of practice effects in language and spatial processing test performance. *Appl Neuropsychol*. 2001; 8(4):218–223. [PubMed: 11989725]

Table 1

Repeated cognitive scores, practice effects, and test-retest correlations

Cognitive scores	Baseline	One Week	Practice Effects	r	d
BVMT-R Total Recall	14.6 (7.0)	22.9 (8.9)	8.2 (5.4)	0.79	1.04
BVMT-R Delayed Recall	5.6 (3.4)	8.0 (3.2)	2.3 (2.1)	0.79	0.73
HVLT-R Total Recall	23.2 (5.6)	27.7 (5.7)	4.4 (3.9)	0.76	0.80
HVLT-R Delayed Recall	6.7 (3.4)	9.0 (2.6)	2.2 (2.4)	0.70	0.76
COWAT	38.3 (11.2)	39.4 (11.6)	1.1 (7.4)	0.79	0.10
Animal fluency	17.1 (5.3)	17.6 (5.3)	0.5 (3.9)	0.72	0.09
TMTA	44.1 (17.1)	39.5 (16.1)	-4.2 (10.6)	0.78	0.28
TMTB	117.1 (57.2)	102.7 (50.5)	-12.4 (38.5)	0.73	0.27
SDMT	39.5 (9.5)	42.1 (10.1)	2.3 (5.1)	0.86	0.27

Note. BVMT-R = Brief Visuospatial Memory Test – Revised, HVLT-R = Hopkins Verbal Learning Test – Revised, COWAT = Controlled Oral Word Association Test, TMTA = Trail Making Test Part A, TMTB = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test. Values represent means and standard deviations (in parentheses). All scores are raw scores. Practice effects scores are One Week – Baseline. r = Pearson correlation. d = Cohen’s d.

Table 2

Simple discrepancy scores

Cognitive scores	2%	5%	16%	50%	84%	95%	98%
BVMT-R Total Recall	-5	-1	2	8	13	16	18
BVMT-R Delayed Recall	-2	-1	0	2	4	6	7
HVLT-R Total Recall	-5	-2	0	4	8	10	12
HVLT-R Delayed Recall	-2	-1	0	1	4	6	9
COWAT	-19	-12	-5	1	8	13	15
Animal fluency	-9	-6	-4	0	4	7	8
TMTA	21	10	4	-5	-14	-22	-31
TMTB	55	42	19	-9	-40	-92	-125
SDMT	-9	-6	-3	2	6	10	16

Note. BVMT-R = Brief Visuospatial Memory Test – Revised, HVLT-R = Hopkins Verbal Learning Test – Revised, COWAT = Controlled Oral Word Association Test, TMTA = Trail Making Test Part A, TMTB = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test. Values represent means and standard deviations (in parentheses). All scores are raw scores, including TMTA and TMTB, which is why they appear reversed. Simple discrepancy scores are One Week - Baseline.

Table 3

Standard deviation and reliable change indexes

Cognitive scores	SDI	RCI	RCI+PE	RCI+PE _{Iverson}
BVMT-R Total Recall	$T_2 - T_1 / 7.0$	$T_2 - T_1 / 4.5$	$(T_2 - T_1) - 8.2 / 4.5$	$(T_2 - T_1) - 8.2 / 5.2$
BVMT-R Delayed Recall	$T_2 - T_1 / 3.4$	$T_2 - T_1 / 2.2$	$(T_2 - T_1) - 2.3 / 2.2$	$(T_2 - T_1) - 2.3 / 2.1$
HVLT-R Total Recall	$T_2 - T_1 / 5.6$	$T_2 - T_1 / 3.9$	$(T_2 - T_1) - 4.4 / 3.9$	$(T_2 - T_1) - 4.4 / 3.9$
HVLT-R Delayed Recall	$T_2 - T_1 / 3.4$	$T_2 - T_1 / 2.7$	$(T_2 - T_1) - 2.2 / 2.7$	$(T_2 - T_1) - 2.2 / 2.4$
COWAT	$T_2 - T_1 / 11.2$	$T_2 - T_1 / 7.2$	$(T_2 - T_1) - 1.1 / 7.2$	$(T_2 - T_1) - 1.1 / 7.4$
Animal fluency	$T_2 - T_1 / 5.3$	$T_2 - T_1 / 3.9$	$(T_2 - T_1) - 0.5 / 3.9$	$(T_2 - T_1) - 0.5 / 3.9$
TMTA	$T_2 - T_1 / 17.1$	$T_2 - T_1 / 11.3$	$(T_2 - T_1) + 4.2 / 11.3$	$(T_2 - T_1) + 4.2 / 11.0$
TMTB	$T_2 - T_1 / 57.2$	$T_2 - T_1 / 42.0$	$(T_2 - T_1) + 12.4 / 42.0$	$(T_2 - T_1) + 12.4 / 39.6$
SDMT	$T_2 - T_1 / 9.5$	$T_2 - T_1 / 5.0$	$(T_2 - T_1) - 2.3 / 5.0$	$(T_2 - T_1) - 2.3 / 5.2$

Note. BVMT-R = Brief Visuospatial Memory Test – Revised, HVLT-R = Hopkins Verbal Learning Test – Revised, COWAT = Controlled Oral Word Association Test, TMTA = Trail Making Test Part A, TMTB = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test. T_1 = baseline score, T_2 = one week score.

Table 4

Simple standardized regression based change scores

Cognitive scores	F(df)	R ²	SEE	Predicted T ₂
BVMT-R Total Recall	279.7 (1,166)	0.63	5.46	$7.87 + (T_1 * 1.02)$
BVMT-R Delayed Recall	282.5 (1,166)	0.63	1.95	$3.71 + (T_1 * 0.75)$
HVLT-R Total Recall	229.3 (1,166)	0.58	3.71	$9.18 + (T_1 * 0.79)$
HVLT-R Delayed Recall	158.3 (1,166)	0.49	1.90	$5.34 + (T_1 * 0.54)$
COWAT	268.6 (1,163)	0.62	7.19	$7.91 + (T_1 * 0.82)$
Animal fluency	172.8 (1,161)	0.52	3.69	$5.07 + (T_1 * 0.73)$
TMTA	257.9 (1,166)	0.61	10.10	$5.11 + (T_1 * 0.79)$
TMTB	186.6 (1,165)	0.53	34.64	$24.00 + (T_1 * 0.68)$
SDMT	465.3 (1,164)	0.74	5.14	$4.08 + (T_1 * 0.95)$

Note. BVMT-R = Brief Visuospatial Memory Test – Revised, HVLT-R = Hopkins Verbal Learning Test – Revised, COWAT = Controlled Oral Word Association Test, TMTA = Trail Making Test Part A, TMTB = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test. All F tests are significant at $p < .001$. SEE = Standard error of the estimate, T_1 = Unstandardized beta weight for the Time 1 raw score. To calculate the Predicted T₂ score, use the formula in the final column. To calculate the reliable change score, $(\text{Observed } T_2 - \text{Predicted } T_2) / \text{SEE}$.

Table 5

Complex standardized regression based change scores

Cognitive scores	F(df)	R ²	SEE	Predicted T ₂
BVMT-R Total Recall	103.6 (3,166)	0.66	5.29	$14.13 + (T_1 * 0.92) + (ed * 0.43) - (age * 0.14)$
BVMT-R Delayed Recall	105.7 (3,166)	0.66	1.88	$0.28 + (T_1 * 0.72) + (ed * 0.18) + (sex * 0.99)$
HVLT-R Total Recall	229.3 (1,166)	0.58	3.71	$9.18 + (T_1 * 0.79)$
HVLT-R Delayed Recall	82.7 (2,166)	0.50	1.88	$8.87 + (T_1 * 0.50) - (age * 0.04)$
COWAT	139.3 (2,163)	0.63	7.12	$19.85 + (T_1 * 0.82) - (age * 0.15)$
Animal fluency	172.8 (1,161)	0.52	3.69	$5.07 + (T_1 * 0.73)$
TMTA	137.0 (2,164)	0.63	9.93	$-15.10 + (T_1 * 0.72) + (age * 0.29)$
TMTB	110.1 (2,165)	0.57	33.13	$-85.54 + (T_1 * 0.57) + (age * 1.56)$
SDMT	255.2 (2,164)	0.76	4.97	$24.55 + (T_1 * 0.86) - (age * 0.21)$

Note. BVMT-R = Brief Visuospatial Memory Test – Revised, HVLT-R = Hopkins Verbal Learning Test – Revised, COWAT = Controlled Oral Word Association Test, TMTA = Trail Making Test Part A, TMTB = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test. All F tests are significant at $p < .001$. SEE = Standard error of the estimate, T₁ = Unstandardized beta weight for the Time 1 raw score, ed = years of education, age = years old at baseline, sex is coded as male = 0 and female = 1. To calculate the Predicted T₂ score, use the formula in the final column. To calculate the reliable change score, $(\text{Observed } T_2 - \text{Predicted } T_2) / \text{SEE}$.

Table 6

Case example

Change formula	Formula	Worked Formula	Result
Simple discrepancy score	$T_2 - T_1$	142 – 195	–53 (84 – 95%)
SDI	$T_2 - T_1 / SD_1$	142 – 195 / 57.2	$z = -0.93$
RCI	$T_2 - T_1 / SED$	142 – 195 / 42	$z = -1.26$
RCI+PE	$(T_2 - T_1) - PE / SED$	$(142 - 195) - (-12.4) / 42$	$z = -0.97$
RCI+PE _{Iverson}	$(T_2 - T_1) - PE / SED_{Iverson}$	$(142 - 195) - (-12.4) / 39.6$	$z = -1.02$
Simple SRB	$(T_2 - T_2') / SEE$	142 – 156.6 / 34.64	$z = -0.42$
Complex SRB	$(T_2 - T_2') / SEE$	142 – 134.8 / 33.13	$z = 0.22$

Note. SDI = standard deviation index, RCI = reliable change index, RCI+PE = reliable change index correcting for practice effects, RCI+PE_{Iverson} = reliable change index correcting for practice effects using Iverson (2001) modification, SRB = standardized regression based formula, T_1 = baseline score, T_2 = one week score, SD_1 = standard deviation at baseline, SED = standard error of the difference, PE = practice effects, T_2' = predicted one week score, SEE = standard error of the estimate.