ELSEVIER

Archives of Clinical Neuropsychology
20 (2005) 485–503

Archives of CLINICAL NEUROPSYCHOLOGY

Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS

Tom N. Tombaugh*

Psychology Department, Carleton University, 1125 Colonel By Drive, Ottawa, Ont., Canada K1S 5B6

Accepted 12 November 2004

Abstract

The present study explored several different procedures for determining the amount of change that occurred on the Mini-Mental State Exam [MMSE; Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198] and Modified Mini-Mental State Exam [3MS; Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry, 48,* 314–318] over short and extended test-retest intervals. The test-retest scores were drawn from a selected sample of elderly individuals who participated in the Canadian Study of Health and Aging [Canadian Study of Health and Aging. (1994). The Canadian study of health and aging: Study methods and prevalence of dementia. *Canadian Medical Association Journal, 150,* 899–913] and were tested on two occasions (CSHA-1 and CSHA-2) separated by 5 years. On each occasion the MMSE and 3MS were administered twice at approximately 3-month intervals. Thus, the mental status tests were administered four times: times 1 and 2 at CSHA-1 and times 3 and 4 at CSHA-2. Mean difference scores and percent of baseline scores showed relatively small group changes over both short and long test-retest intervals for the MMSE and the 3MS. A reliable change index based on a linear regression model controlled for practice effects, psychometric errors due to low reliability, regression to the mean, and accounted for the effects of various demographic variables. Consequently, this reliable change index provided a better estimate of the amount of change that occurred for individual participants than did the mean Retest-Test 1 difference, percent of baseline change, or a reliable change index based on a Retest-Test 1 difference score. Normative data for the change scores are provided.

Keywords: Test-retest interval; Mini-Mental State Examination; Cognitive status

* Tel.: +1 613 520 2659; fax: +1 613 520 3667.

0887-6177/$ – see front matter
doi:10.1016/j.acn.2004.11.004

The Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) is the most widely used measure to screen for cognitive status in medical and neuropsychological research (Tombaugh & McIntyre, 1992). One of the primary functions of the MMSE is to serially measure change in cognitive status. Repeated administrations of the MMSE have been used to differentiate between normal age-associated cognitive decline and the pathological cognitive decline that occurs in dementia, to track the progression of dementia, and to assess the effects of rehabilitation. Multiple administrations of the MMSE have also been employed in epidemiological research to investigate the incidence and prevalence of various neuropathological impairments. The central question in all of these applications concerns the degree to which an individual's mental (i.e., cognitive) status has declined.

At the most basic level, interpreting a change in a person's cognitive status cannot proceed without knowing how much change occurs normally in cognitively intact individuals. Attempts to determine changes over time traditionally have relied on some type of test-retest reliability coefficient. Most attempts to provide this type of information with cognitively intact individuals have used relatively short test-retest intervals of less than 6 months. These reliability estimates generally fall between .80 and .95 (Tombaugh & McIntyre, 1992). The clinical relevancy of these estimates, however, may be called into question because the duration of the test-retest interval is generally shorter than that used when a person's change in mental status is being examined clinically. Longer intervals produce less stable test scores. For example, when 1- or 2-year test-retest intervals are used, correlation coefficients less than .50 have been obtained (Escobar et al., 1986; Mitrushina & Satz, 1991). One factor that may account for this decline in reliability is the possibility that some individuals who had been classified as cognitively intact at time 1 were, in fact, mildly impaired or demented. The cognitive decline that occurred for these individuals during the test-retest interval may have been responsible, at least in part, for the lower reliability estimates that occurred with longer test-retest intervals.

Attempts to determine the stability of the MMSE over time periods greater than 6 months or 1 year have relied on the amount of change that occurs when group performance is averaged across two or more time points. Although this approach may provide information about the performance of the group, it does not provide any information about the stability of individual scores over the time period. Information of this type requires calculating average change scores, a statistic that is distinctly different from group means.

Regardless of whether the goal of the investigation is determining group trends or establishing the stability of individual scores, interpreting changes in mental status scores ideally requires some type of independent evidence showing that the cognitive status of the person has, in fact, remained unchanged over this testing interval. Otherwise, stability estimates for the MMSE are hopelessly confounded with the effects of various neuropsychological processes. Interpretation of changes in cognitive status also needs to consider the effects of measurement errors such as regression to the mean. Other potentially critical variables are demographic factors and the effects of prior experience. Since it has been well established that MMSE scores are affected by age and education (Bravo & Hebert, 1997; Crum, Anthony, Bassett, & Folstein, 1993; Tombaugh & McIntyre, 1992; Tombaugh, McDowell, Kristjannsson, & Hubley, 1996), any attempt to interpret a change score must take these factors into consideration. Prior experience may result in practice effects merely because a person has had some previous experience with the testing materials, because participants may remember some of the questions and rehearse answers given previously (Keeting, 1987), or because prior to the retest the person knows that a cognitive test will occur and is likely to be more alert (Unger, van Belle, & Heyman, 1999). The effects of practice, however, become progressively less important as the retest-interval increases.

Thus, to determine if a MMSE retest score reflects a true change in cognitive status, it needs to be compared against some type of normative standard that takes into account (1) test-retest reliability, (2) measurement error, (3) demographic variables such as age and education, and (4) practice effects. Moreover, some type of evidence should be presented to insure that the normative sample does not contain individuals who are demented or suffer from other neurological or psychiatric dysfunction.

The present study compares several different procedures for assessing the amount of change that occurred in both the MMSE and Modified MMSE (3MS; Teng & Chui, 1987) over both short (<3 months) and long (5 years) test-retest intervals. The 3MS represents an extension of the MMSE where four additional subtests (date and place of birth, word fluency, similarities, and delayed recall of words) were added. The maximum score was increased from 30 to 100 points, and a modified scoring procedure permitted assignment of partial credit on some items. One advantage of the 3MS is that both the 3MS and MMSE scores can be derived from a single administration (see Tombaugh et al., 1996 for more details). Test-retest data from the Canadian Study of Health and Aging (CSHA, 1994) provide a unique opportunity to compare mental status scores in a sample of elderly participants who were tested on two occasions separated by 5 years (i.e., 1991 and 1996) and who received a consensus diagnosis of "no cognitive impairment" (NCI) by physicians and neuropsychologists on both occasions. On each occasion, the MMSE and 3MS were administered twice yielding two short test-retest intervals of approximately 3 months duration. The consensus diagnosis of NCI reduces the likelihood that any changes in mental status scores can be attributed to dementia or other types of neurological/psychiatric impairments that might have occurred during the 5-year test-retest interval. In addition to traditional group analyses, the data were subjected to a series of analyses designed to establish a reliable change index (RCI) that could be used to make clinical decisions as to whether an individual's retest score was significantly lower than expected.

1. Method

1.1. Research participants

The participants represent a subsample of community-dwelling individuals drawn from the Canadian Study of Health and Aging (Canadian Study of Health and Aging, 1994; McDowell, Hill, & Lindsay, 2001). The participants were first tested in 1991 (CSHA-1) and then again in 1996 (CSHA-2). All interviewers received intensive 4- to 5-day training using training manuals, videotapes, feedback sessions and self-testing (for more details see McDowell, Helliwell, Sykes, Hill, & Lindsay, 2001). In 1991, a home interview which contained the screening MMSE/3MS (Time 1) was conducted with 9008 individuals, aged 65 years to 99 years. Of these, the 1614 individuals who scored negative (<78) on the screening 3MS plus a random sample of 503 participants who scored positive on the screening exam were subsequently

Table 1
Mean (S.D.) retest interval (days) using Time 1 as the baseline score and the mean (S.D.) MMSE/3MS score at each test interval

Test items	Time 1	Time 2	Time 3	Time 4
Retest interval				
Mean (S.D.)	NA	87.3 (67.8)	1758.1 (190.0)	1823.2 (202.1)
Range	NA	10–319	1202–3747	1213–3794
MMSE				
Mean (S.D.)	26.91 (2.86)	27.63 (2.16)	26.98 (2.75)	27.31 (2.60)
Range	17–30	19–30	15–30	18–30
3MS				
Mean (S.D.)	87.63 (10.01)	90.54 (7.43)	87.95 (8.61)	89.25 (8.65)
Range	58–100	65–100	58–100	51–100

administered a clinical assessment which included a complete clinical and neurological examination by a research physician, an extensive battery of neuropsychological tests administered by a neuropsychologist, and a second MMSE/3MS (Time 2) given by a registered nurse. The cutting score of <78 was selected following the results of a pilot study and was used to ensure high sensitivity for individuals suffering from Alzheimer's disease. The average time interval between Time 1 (T1) and Time 2 (T2) was approximately 3 months (see Table 1). Based on all of the clinical information obtained at Time 2 (including MMSE/3MS scores from T2 but not MMSE/3MS scores from T1), neuropsychologists and physicians independently classified participants as "no cognitive impairment" (NCI), "demented", or "cognitive impairment but not demented" (CIND). These initial diagnoses were reviewed at a consensus conference before a final diagnosis was reached. The dementia diagnosis and the subclasses of Alzheimer's disease and vascular dementia were based on the DSM-III-R criteria (American Psychiatric Association, 1987), NINCDS-ARDRA criteria (McKhann et al., 1984), and the ICD-10 criteria (World Health Organization, 1992). The CIND category was comprised of individuals whose level of cognitive impairment was judged to be greater than the NCI group but less than the Dementia group and included the following subclasses: delirium, chronic drug abuse, depression, psychiatric, age-associated memory impairment, cerebral vascular stroke, general vascular, Parkinson's disease, socio-cultural, and other.

The current study focuses on the 756 participants who received a consensus diagnosis of NCI on CSHA-1. Approximately five years after the Time 2 administration of the mental status tests, 463 NCI individuals were retested (CSHA-2) using the same screening (Time 3) and clinical examination procedures (Time 4) described previously. Of the original 756 participants, 201 individuals had died, 13 people were not contacted, 12 participants were geographically inaccessible, and 67 individuals refused. The data from 82 individuals could not be used because the MMSE/3MS had been administered only once or the participants were illiterate, blind, or deaf. Of the remaining 381 individuals, the status of 232 remained unchanged, 102 were reclassified as CIND, and 47 were reclassified as "dementia". The present study was primarily interested in determining the stability of mental status scores from participants who had not developed any cognitive impairments during the test-retest period. Consequently, no scores from the Dementia and CIND groups were included. One hundred and sixty of the participants who received a diagnosis of NCI on both assessments were administered the English version, 68 received the French version (Hébert, Bravo, & Girouard, 1992), and the language of the screening exam was not available for four participants. Only the scores from the English version were used in the current study.

1.2. Test measures

The 3MS was administered and scored according to the guidelines set forth in the administration manual (Teng, 1990). The 3MS and MMSE scores were both derived from the 3MS. Some subscales in the 3MS and MMSE employed different items than used in the original MMSE (Folstein, Folstein, & McHugh, 1975). The major changes occurred in Orientation to Place: 3MS: "province" was substituted for "state" and awarded 2 points rather than 1 point, "country" was substituted for "county", "name of hospital" and "floor of hospital" were deleted and replaced with a multiple choice question "Where are you? Are you in a hospital, store or home?"; MMSE: "province" was substituted for "state" and awarded 1 point, "country" was substituted for "county", "name of hospital" and "floor of hospital" were deleted and replaced with "number (1 point) and name of street (1 point) where you live". Other modifications are as follows: Concentration: spell WORLD backward replaced serial 7 s; Registration and Recall; "shirt", "brown" and "honesty" (T1 and T2) and "socks", "black" and "charity" (T3 and T4) replaced "penny", "apple" and "table"; Three-stage command: name of the non-preferred hand (right or left) was used rather than "right" hand.

2. Results

2.1. Demographics

The mean age of the participants was 76.32 (S.D. = 5.42, range = 65–89) years with an average of 10.94 (S.D. = 3.90, range = 0–25) years of education. There were 79 male and 81 female participants. The average duration of time between each administration of the mental status tests is shown in Table 1. It should be noted that the even though the mean time for the two short test-retest intervals (i.e., T1–T2 and T3–T4) were very similar, the range of time intervals within each comparison was large. In order to determine if the mental status scores for each comparison were representative for the group, the test-retest duration was divided into five intervals, each consisting of approximately 30 participants: 1–30 days, 31–60 days, 61–90 days, 91–150 days and 151–250 days. ANOVAs performed over the mean MMSE and 3MS scores from each of the five intervals did not yield any significant differences for either T2 or T4. These results indicated that despite a large range in these two short test-retest intervals, mean scores for T2 and T4 were representative of group performance. Consequently, the group means were used in all subsequent analyses.

2.2. Test and retest scores

Mean (S.D.) scores from each of the four test administrations are shown in Table 1. Perhaps the most frequent way to determine whether repeated experience with a test produces a change in performance is by using a repeated-measures ANOVA. The results of this analysis produced a statistically significant effect, MMSE: $F(3,149) = 5.13$, $P < .002$; 3MS: $F(3,149) = 11.72$, $P < .001$. When relatively short test-retest intervals were used, paired comparisons (LS.D.) showed that the retest scores significantly increased—T2 scores were significantly greater than T1 scores for both the MMSE and 3MS ($P < .001$) and T4 scores were significantly higher than T3 scores for the 3MS ($P < .03$). The increased retest performances were most likely due to a practice effect attributable to the relatively short test-retest interval. A different picture emerges when longer test-retest intervals were used thereby reducing, if not completely eliminating, practice effects. This is most obvious in the comparison between T1 and T3 scores where no significant effects were obtained on either mental status test. The stability of the mental status scores is further illustrated by the fact that mean difference scores (T3 – T1) varied by less than 1/2 of a point, and percent change scores ([retest/baseline test] × 100) increased by only a single percent (see Table 2). Although the other long-term comparisons (T2 vs. T3 and T2 vs. T4) are significant ($P < .01$), they are of less interest clinically and are more difficult to interpret because of the potentially confounding effects of practice that probably inflated the T2 and T4 scores.

Table 2
Difference scores and percent of baseline scores on the MMSE and 3MS for different test-retest intervals

Mental status test	Difference score[a]	Percent of baseline score[b]
MMSE		
Short intervals		
T1–T2	0.69 (2.32, −4 to +10)	103.34 (10.50, 84–159)
T3–T4	0.35 (2.41, −9 to +8)	101.92 (10.51, 69–153)
Long intervals		
T1–T3	0.09 (2.65, −13 to +8)	100.91 (10.99, 54–141)
T1–T4	0.31 (2.37, −5 to +7)	101.74 (8.97, 80–141)
T2–T3	−0.66 (2.56, −15 to +5)	97.85 (9.48, 50–126)
T2–T4	−0.32 (2.03, −9 to +4)	98.97 (7.60, 67–118)
3MS		
Short intervals		
T1–T2	2.76 (6.30, −15 to +31)	103.82 (8.73, 83–154)
T3–T4	1.17 (6.27, −25 to +19)	101.63 (8.00, 67–133)
Long intervals		
T1–T3	0.38 (7.39, −23 to +25)	101.01 (9.39, 77–140)
T1–T4	1.27 (6.86, −19 to +22)	101.96 (8.87, 76–132)
T2–T3	−2.47 (6.55, −30 to +12)	97.42 (7.55, 66–117)
T2–T4	−1.36 (6.09, −22 to +17)	98.60 (7.21, 75–124)

[a] Difference scores were calculated by subtracting the baseline (earlier) score from the retest (later) score (e.g., T2–T1). A positive score means that performance increased at retest and a negative score means that performance decreased at retest. The numbers in parenthesis represent S.D. and lowest to highest score.

[b] Percentages of baseline scores were calculated by dividing the retest (later) score by the baseline (earlier) score (e.g., T3/T2 × 100). A percentage score greater than 100 means that performance increased at retest and a percentage score less than 100 indicates that performance decreased at retest. The numbers in parenthesis represent S.D. and lowest to highest score.

The data shown in Table 1 are primarily useful for group comparisons, but they are not particularly helpful for estimating the amount of change that occurs for individual participants. The following example illustrates the possible misinterpretations that may occur when the average change for a group is used to predict individual performance. The mean MMSE change of +.08 of a point that occurred between T1 and T3 suggests that the performance of these cognitively-intact individuals should remain relatively unchanged during the 5-year test-retest interval. This suggestion is in marked contrast to the finding that only 14% of these participants had the same MMSE score at T1 and T3. For the 3MS, the percentage decreased to 6%.

One reason for this lack of long-term stability in individual test scores is that retest scores typically show a regression to the mean whereby higher T1 scores become lower at T3, and lower T1 scores become higher at T3. The existence of the regression effects is amply illustrated by computing the percentage of participants who showed an increase, decrease, or no change in T3 scores as a function of their T1 performance (i.e., high, middle, and low). Inspection of the data suggested that regression effects for the MMSE and 3MS were best illustrated by using the following values for high, middle and low T1 performance (MMSE = 30–29, 28–26, and 25–0; 3MS = 100–93, 92–85, and 84–0). Inspection of Table 3 clearly demonstrates the expected regression to the mean for both the MMSE and 3MS. For example, 47% more of the MMSE low scores *increased* on T3 than decreased (i.e., 71–24%), while 46% more MMSE high scores *decreased* on T3 than increased (i.e., 64–18%). Test-retest reliabilities for the different retest intervals were all reliable and varied between .48 and .65 for the MMSE and between .68 and .77 for the 3MS. The regression to the mean may have reduced the magnitude of some of the retest reliability quotients. The higher reliability for 3MS scores is due, at least in part, to the positive effects that a longer test has on reliability coefficients.

Table 3
Percent of participants that increased, decreased, or remained the same at Time 3 as a function of the level of performance at Time 1

Scores	*n*	Time 3		
		Percent increase	Percent no change	Percent decrease
Time 1				
MMSE				
0–25	42	71	5	24
26–28	61	54	16	32
29–30	57	18	18	64
Total	160	45	14	41
3MS				
0–84	50	82	0	18
85–92	41	56	2	42
93–100	69	15	13	72
Total	160	46	6	48

2.3. Normative change scores

As illustrated above, analyses showing the average amount a group changes are not particularly helpful in determining how much an individual's score must change in order for it to represent a significant change. A review of the literature shows that two procedures commonly have been used to establish reliable change scores with other cognitive tests, but have not yet been applied to mental status testing. Although the two procedures differ in the precise statistic used, they share the common logic of expressing the RCI in terms of a confidence interval that is established on the basis of the distribution of change scores.

2.3.1. Reliable change index-difference scores (RCI-Diff)

Jacobson and Truax (1991) are generally given credit for pioneering the procedure that uses the numeric difference between retest scores and baseline test scores to create a reliable change index (RCI-Diff). Procedures that adjust for practice effects and test-retest reliability are also commonly employed (Chelune, Naugle, Luders, Sedlak, & Awad, 1993; Jacobson & Truax, 1991). The basic procedure for establishing a RCI-Diff is to determine how much retest scores differ from baseline scores (retest score–baseline score) for each participant with standard deviation units used to create a confidence or prediction interval in a way similar to how standard deviations are used to create a standard error of measurement. That is, confidence intervals represent the distribution of change scores that would be predicted if no real change had occurred (Jacobson & Truax, 1991). Consequently, 68% of the change scores would be expected to fall between ±1.00S.D., 90% of the change scores should fall between ±1.64S.D., and 98% of change scores should fall between ±1.98S.D. Thus, if a 90% confidence interval is established with change scores that have a 3.10S.D., 90% of participants should have change scores varying between ±5.08 points (i.e., 3.10 × 1.64), 5% of individuals should have scores that have increased by more than 5.08 points, and 5% of individuals should have scores that have decreased by more than 5.08 points. Stated slightly differently, any change in a mental status score that is greater than ±5.08 would be considered to be statistically significant since it would be highly unlikely (.05 level) that the score would have occurred by chance in the normative group. In order to control for practice effects, the practice score for the normative group (mean retest score–mean baseline score) is subtracted from an individual's difference score (individual retest score–individual baseline score). This corrected difference score is considered to represent a reliable change if it exceeds the RCI-Diff.

A RCI-Diff can be computed in two ways. The first method (Jacobson & Truax, 1991) used the standard error of measurement [$\text{S.E.}_{\text{meas}} = (\text{S.D. of baseline scores}) \times \sqrt{(1 - \text{test-retest reliability coefficient})}$] to correct for possible measurement error. Using the $\text{S.E.}_{\text{meas}}$, the $\text{S.E.}_{\text{diff}}$ was computed: $\text{S.E.}_{\text{diff}} = \sqrt{2(\text{S.E.}_{\text{meas}})^2}$. The S.E.diff was then used to create a confidence interval: $\text{RCI-Diff} = \text{S.E.}_{\text{diff}} \times 1.64$. Thus, the $\text{RCI-Diff} = \sqrt{2(\text{S.E.}_{\text{meas}})^2} \times 1.64$. The second method (Dikmen, Heaton, Grant, & Temkin, 1999) simplified this procedure and merely used the S.D. of the difference scores ($\text{S.D.}_{\text{diff}}$). The RCI-Diff was calculated by multiplying the $\text{S.D.}_{\text{diff}}$ by 1.64. Both procedures were used in the current study. Since they produced similar effects, only the results from the $\text{S.E.}_{\text{meas}}$ procedure are presented.

Table 4 shows the RCI-Diff, practice effects, and percentage of change scores that exceeded the RCI-Diff in the positive and negative direction. Test-retest comparisons are divided into

Table 4
Reliable change index using $S.E._{diff}$ for MMSE and 3MS, and the percentage of scores that exceeded the RCI-Diff value

Test	RCI-Diff value	Practice effect	Percent scores > RCI-Diff	
			− (5th)	+ (95th)
MMSE				
Short intervals				
T1–T2	±4.16	0.69	1.9	6.4
T3–T4	±4.01	0.35	5.2	4.6
Long intervals				
T1–T3	±4.42	0.08	3.1	6.3
T1–T4	±4.10	0.32	5.9	5.9
T2–T3	±3.59	−0.66	3.9	2.6
T2–T4	±2.94	−0.32	4.6	7.2
3MS				
Short intervals				
T1–T2	±11.13	2.76	2.6	5.1
T3–T4	±10.19	1.17	6.5	5.2
Long intervals				
T1–T3	±13.22	0.37	3.8	4.4
T1–T4	±11.82	1.27	3.9	5.2
T2–T3	±9.81	−2.47	4.5	5.1
T2-T4	±8.76	−1.36	9.2	4.6

those with a relatively short test-retest interval (e.g., approximately 3 months) and those which involved a long test-retest interval (e.g., approximately 5 years). For each comparison the percentage of scores that exceeded the RCI-Diff value was calculated in the following manner. The difference score for each person (Retest-Baseline) was obtained for each comparison, and the practice effect was subtracted from the difference score. These corrected change scores were rank ordered to determine what percentage of people obtained a change score greater than the RCI-Diff score. Corrected changes scores were computed and used to determine what percentage of people in the NCI group actually exceeded the RCI-Diff score. As indicated previously, 5% of the sample should have scores that are larger and smaller than the RCI-Diff value.

2.3.2. Reliable change index-regression scores (RCI-Reg)

Although the RCI-Diff controls for practice effects and psychometric errors due to low reliability coefficients, it does not correct for regression to the mean. In addition, the effects of various demographic variables such as age, education, and gender are not accounted for. In view of this, linear regression equations have been used to provide a quantitative measure of test-retest change (Dikmen et al., 1999; McSweeny, Naugle, Chelune, & Luders, 1993; Salinsky, Storzbach, Dodrill, & Binder, 2001; Temkin, Heaton, Grant, & Dikmen, 1999). The basic procedure was to regress the baseline scores and various demographic variables (e.g., age, gender, education) on retest scores for each individual in the normative sample. After the predicted retest score was obtained, it was subtracted from the observed retest score. If this change score exceeded the RCI-Reg it was considered to represent a significant change at the .05 level (one tailed). The RCI-Reg was computed by multiplying the S.E.$_{est}$ (readily available as an SPSS output) by 1.64. Some authors (Temkin et al., 1999) advocate using the residual S.D. (also readily available on SPSS) rather than the S.E.$_{est}$. Both measures were employed in the present investigation and produced virtually identical results. Consequently, only the S.E.$_{est}$ procedure is reported. A hierarchical regression procedure was used in which the variables were entered in the following order: age, education, sex and baseline score. Table 5 shows the regression formulas when a baseline was the sole regressor and when a baseline score was combined with all demographic variables (coefficients were included only if they exceed the .10 level of significance), the percent of variance explained by the baseline score alone (R^2—Baseline only) and by all variables (R^2—all variables), S.E.$_{est}$, the RCI-Reg value, and the percentage of participants who had a negative or positive score greater than the RCI-Reg. Table 6 shows the total amount of variance explained when each demographic variable was entered separately and when all variable were entered together. Inspection of this table reveals that when each demographic variable was entered separately, education generally accounted for the greatest percent of variance. However, the percent of variance accounted for by education was a distant second to that accounted for by the baseline test which often was two to three times greater than education. This is most dramatically obseved in the 3MS where the baseline test accounted for approximately 50–60% of the variance. Although the effects of the baseline test are diminished when all demographic variables are entered together, it still accounts for approximately 20–30% of the variance. This value is roughly equivalent to the amount of variance explained by the combined effect of all demographic variables. The overall findings shown in Table 6 and described above are consistent with those previously reported in Table 5 which showed that RCIs calculated using baseline scores as the only regressor were very similar to those that occurred when Baseline test scores were combined with the demographic variables.

Although RCIs provide a cut-off score which has both clinical and experimental utility, they represent an arbitrary criterion for identifying scores that represent a "true" change in performance. Since other decision rules could be used (e.g., 90th rather than 95th percentile), reference values upon which these decisions could be based were calculated by using SPSS to create a frequency distribution of scores (retest score observed-retest score predicted) and then obtaining percentiles that corresponded to the various scores (Table 7). This procedure provides information about the relative ranking of scores in the same way as does any normative table that uses percentile values. Of the two short test-retest intervals, the T1–T2 interval probably holds the greatest interest clinically since it conforms to the type of comparison typically used in clinical situations. However, the data from T3–T4 are also included to show the potential effects that the repeated administrations of the mental status tests have on performance. Normative data is also presented for two of the long test-retest interval. The T1–T3 interval has the greatest clinical utility and is relatively uncontaminated by practice effects. That is, it is unlikely that effects of the mental status tests at T2 will be sustained over the 5-year test-retest interval. The data for T2–T3 are included but are judged to be less relevant clinically since few, if any, clinical situations will parallel the administrative procedures used in this comparison.

Table 5
Regression formulas predicting retest scores, percent of variance explained by the baseline test and by all variables[a], $S.E._{est}$, RCI-Regression (RCI-Reg) value, and the percentage of scores that exceeded RCI-Reg value

					Percent scores > RCI-Reg	
Variable	Regression formula	R^2	$S.E._{est}$	RCI-Reg score	− (5th)	+ (95th)
MMSE						
Short intervals						
T1–T2						
Test 1 only	.46 (test 1) + 15.36	.36	1.73	±2.84	7.69 (−2.94)	3.85 (+2.06)
All variables	.38 (test 1) − .07 (age) + .10 (educ) + 21.65	.41	1.67	±2.73	7.69 (−3.07)	3.21 (+2.51)
T3–T4						
Test 3 only	.56 (test 3) + 12.28	.36	2.09	±3.43	6.54 (−4.16)	2.60 (+2.60)
All variables	.46 (test 3) + .12 (educ) + .83 (sex) + 15.21	.41	2.02	±3.31	21.60 (−6.58)	0.00 (0.00)
Long intervals						
T1–T3						
Test 1 only	.53 (test 1) + 12.61	.31	2.29	±3.75	6.28 (−3.86)	1.26 (+2.73)
All variables	.45 (test 1) − .09 (age) + 1.06 (sex) + 19.12	.37	2.20	±3.60	3.15 (−2.79)	7.55 (+3.76)
T1–T4						
Test 1 only	.57 (test 1) + 11.87	.39	2.04	±3.34	9.43 (−3.98)	1.31 (+2.31)
All variables	.48 (test 1) − .05 (age) + .07 (educ) + .91 (sex) + 16.24	.43	1.99	±3.26	9.80 (−4.15)	0.00 (+2.33)
T2–T3						
Test 2 only	.61 (test 2) + 10.10	.23	2.43	±3.98	5.80 (−4.18)	0.00 (+2.65)
All variables	.46 (test 2) + .09 (educ) + 1.36 (sex) + 15.53	.31	2.33	±3.82	20.32 (−7.87)	0.00 (0.00)
T2–T4						
Test 2 only	.78 (test 2) + 5.77	.42	1.98	±3.24	5.52 (−3.15)	0.00 (+2.95)
All variables	.68 (test 2) + .07 (educ) + 1.09 (sex) + 7.40	.47	1.92	±3.14	12.50 (−4.50)	0.00 (+1.34)

Table 5 (*Continued*)

Variable	Regression formula	R^2	$S.E._{est}$	RCI-Reg score	Percent scores > RCI-Reg: − (5th)	Percent scores > RCI-Reg: + (95th)
3MS						
Short intervals						
T1–T2						
Test 1 only	.60 (test 1) + 39.68	.60	4.74	±7.77	8.97 (−10.68)	1.92 (+3.92)
All variables	.53 (test 1) – .27 (age) + .20 (educ) + 62.20	.60	4.52	±7.41	5.79 (−8.28)	4.49 (+7.03)
T3–T4						
Test 3 only	.74 (test 3) + 24.26	.54	5.86	±9.61	7.19 (−10.34)	3.26 (+8.24)
All variables	.61 (test 3) – .19 (age) + .35 (educ) + 2.94 (sex) + 42.25	.60	5.56	±9.12	5.23 (−8.10)	5.22 (+8.83)
Long intervals						
T1–T3						
Test 1 only	.60 (test 1) + 35.63	.48	6.22	±10.20	8.18 (−11.63)	3.14 (+8.57)
All variables	.52 (test 1) – .23 (age) + .30 (educ) + 1.93 (sex) + 53.86	.53	5.99	±9.82	6.22 (−10.04)	3.14 (+8.88)
T1–T4						
Test 1 only	.65 (test 1) + 32.40	.53	5.96	±9.77	7.18 (−10.80)	1.31 (+7.15)
All variables	.54 (test 1) – .26 (age) + .31 (educ) + 3.40 (sex) + 53.53	.60	5.54	±9.08	7.84 (−11.12)	3.37 (+7.00)
T2–T3						
Test 2 only	.78 (test 2) + 17.35	.46	6.38	±10.46	5.80 (−10.75)	1.94 (+7.95)
All variables	.67 (test 2) + .37 (educ) + 2.49 (sex) + 25.15	.50	6.18	±10.13	16.80 (−16.06)	0.00 (+2.41)
T2–T4						
Test 2 only	.84 (test 2) + 13.14	.53	5.99	±9.82	7.89 (−13.02)	2.63 (+6.90)
All variables	.71 (test 2) + .34 (educ) + 3.78 (sex) + 25.76	.59	5.61	±9.20	33.60 (−21.00)	0.00 (−0.68)

In calculating the percentage of scores that exceeded the RCI-Reg value, all predicted scores that were greater than the maximum possible score (MMSE = 30; 3MS = 100) were changed to equal the maximum value. Age and education are expressed in years. Sex was coded as male = 1 and female = 2.

[a] Demographic variables were added to regression equation if $P < .10$.

Table 6
Percentage of total variance explained when each demographic variable was entered separately and when all variables were entered together

Test items	Age	Education	Gender	All	Baseline test
MMSE					
Short intervals					
T1–T2					
Separately	6.0	14.3	3.0		35.8
Together				22.9	18.9
T3–T4					
Separately	2.4	10.6	9.4		35.9
Together				22.5	18.7
Long intervals					
T1–T3					
Separately	3.9	7.5	8.3		30.9
Together				19.7	17.5
T1–T4					
Separately	2.6	10.6	9.4		38.5
Together				22.5	20.4
T2–T3					
Separately	3.2	7.5	8.3		22.9
Together				20.6	9.9
T2–T4					
Separately	2.3	10.6	9.4		42.3
Together				22.7	24.5
3MS					
Short intervals					
T1–T2					
Separately	8.1	17.1	1.4		59.5
Together				26.0	38.0
T3–T4					
Separately	5.4	18.0	8.8		54.3
Together				32.2	27.5
Long intervals					
T1–T3					
Separately	4.0	16.6	3.6		48.2
Together				23.8	29.0
T1–T4					
Separately	5.7	18.0	8.8		52.9
Together				32.2	27.8
T2–T3					
Separately	3.8	16.6	3.6		45.5
Together				25.3	24.5
T2–T4					
Separately	5.7	18.0	8.8		52.6
Together				32.2	27.0

Table 7
Percentiles for the difference between obtained retest scores and retest scores predicted on the basis of regression equations using baseline scores and demographic information as regressors for the MMSE and 3MS

	Short test-retest intervals		Long test-retest intervals	
Percentiles	T1–T2	T3–T4	T1–T3	T2–T3
MMSE				
98 (+2S.D.)	3.08	0.87	4.26	0.00
95	2.53	0.01	3.79	0.00
90	2.05	0.00	3.44	0.00
84 (+1S.D.)	1.43	0.00	3.16	−1.00
75	1.06	−0.67	2.55	−1.00
50 (0S.D.)	0.13	−2.00	1.18	−2.00
25	−0.78	−3.00	0.01	−4.00
16 (−1S.D.)	−1.50	−4.00	−0.84	−5.00
10	−2.16	−4.96	−1.67	−6.14
05	−3.20	−7.30	−2.99	−8.00
02 (−2S.D.)	−4.22	−8.12	−4.67	−10.81
3MS				
98 (+2S.D.)	11.27	11.62	12.06	4.34
95	7.48	9.72	8.90	2.60
90	4.90	6.46	7.05	0.82
84 (+1S.D.)	3.99	5.08	4.62	−0.68
75	2.66	3.61	3.45	−1.52
50 (0S.D.)	0.21	1.11	0.47	−4.19
25	−1.54	−2.31	−3.12	−8.58
16 (−1S.D.)	−2.96	−4.27	−5.47	−10.52
10	−5.40	−6.35	−7.79	−12.56
05	−8.47	−10.51	−10.15	−16.78
02 (−2S.D.)	−10.28	−13.82	−17.31	−24.75

The other two long test-retest intervals are not included because of their lack of clinical utility.

3. Discussion

The primary purpose of the present study was to compare several methods for determining the degree of change that occurred in MMSE and 3MS over various test-retest intervals. In addition to using relatively short intervals, similar to those used in studies designed to establish test-retest reliability coefficients, the present investigation employed a substantially longer test-retest interval of 5 years. In addition, the status of all participants was documented to remain unchanged during the 5-year, test-retest interval.

Traditionally, stability of mental status performance has been investigated by using mean difference scores (retest–baseline score) or by determining the percent that retest scores change relative to those obtained at baseline. Results showing that the MMSE and 3MS scores increased by less the 1/2 of a point supports previous research showing that MMSE practice effects are minimal and group means generally change less than a single point over a wide range of test-retest intervals, with slightly larger changes occurring for participants over the age of 75 years (Aguero, Fratiglioni, Guo, Viitanen, & Winbald, 1998; Aevarsson & Skogg, 2000; McCaffrey, Duff, & Westervelt, 2000; Schmand et al., 1995; Starr, Deary, Inch, Cross, & MacLennan, 1997; Tombaugh & McIntyre, 1992).

Based on the findings presented above, one might conclude that the performance of the NCI participants was relatively stable over the 5-year, test-retest interval. However, test-retest correlations and evidence from regression analyses showed that considerable variability existed among the participants. A regression to the mean was clearly evident for the NCI group where only 14% of the MMSE scores and 6% of the 3MS scores remained the same. In general, these results are consistent with those reported by Olin and Zelinski (1991) and Mitrushina and Satz (1991) who reported that MMSE scores changed by at least 2 points for 50 and 42% of community-dwelling participants over a 1-year or 2-year period, respectively. In the present study MMSE scores changed by at least 2 points for 48% of the participants over a 5-year period.

The preceding evidence indicates that while group means are useful in comparing performance across different types of experimental conditions, they do not provide adequate normative information about how much a person's cognitive status is expected to change over time. In order to provide this type of information some other approach is needed. The present investigation examined two such procedures that have found favour recently among neuropsychologists: RCI-Difference and RCI-Regression. Although difference scores are widely used in assessing the stability of mental status scores, their utility is limited by several factors even when corrected for potential practice effects, a factor that was relatively unimportant in the present investigation. One limitation, regression to the mean, makes it very difficult to apply the RCI-Diff values uniformly across all levels of Test 1 scores. For example, a decline greater than the suggested cut-off score of 4.23 points would actually represent a greater change if a person had a low score on Test 1 rather than a high score. Moreover, difference scores do not take into consideration the effects that demographic variables have on mental status scores. Although more precise RCI-Diff scores could be achieved by creating separate RCIs for important demographic variables and different levels of scores on Test 1, the use of regression equations is a more economical approach, since they can be used to create a single RCI that controls for both regression to the mean and the potential effects of demographic variables (Hermann et al., 1996; McSweeney, Naugle, Chelune, & Luders, 1993; Temkin et al., 1999). Using this approach, a person's MMSE score would have to decline by at least by 4 points over the 5-year period (10 points on the 3MS) to represent a significant change (<.05 level). As indicated above, this magnitude of change is in stark contrast to what might be predicted on the basis of group means which showed less than a 1/2 point change over the 5 years.

Regression analyses showed that most of the variance in retest scores was due to the score on Test 1, with age and education definitely playing a secondary role. The importance of the score on Test 1 was further demonstrated by the similarity of the RCIs when all variables were entered and when Test 1 score was the sole regressor. The ability of Test 1 scores to predict the score on subsequent administration of a test has been reported by several studies using a variety of neuropsychological tests (Salinsky et al., 2001; Sawrie, Chelune, Naugle, & Luders, 1996; Temkin et al., 1999).

The procedure used for establishing cut off scores has been employed in several previous studies. However, it has two potential drawbacks in the present study. The first limitation relates to the assumption that a 90% confidence interval is created when the $S.E._{est}$ is multiplied by ± 1.64. This will only occur when change scores are normally distributed. This assumption does not appear to be fulfilled in the present investigation. There is a consistent trend for greater than 5% of the participants to have change scores that exceeded the lower limit of the RCI and less than 5% of the participants have change scores that exceeded the upper limit of the RCI. A second limitation is that cut-off scores only permit a categorical classification of whether a mental status score has significantly changed. They do not permit determining the relative percentage of change scores that do not exceed the cut off value. Both of these limitations were addressed in the present study by using a set of percentile scores created by a using a non-linear transformation of scores. It is recommended these values should be used, rather than group means and standard deviations or difference scores, to determine the degree to which a change in a MMSE or 3MS score deviates from what would be expected over a five year period. It should be noted that these norms can be used clinically to document serial change in mental status scores, and experimentally as a surrogate control group to determine the percentage of individuals in a specific group whose scores have changed. Caution should be exercised to avoid over interpreting the clinical significance of change scores. Anyone attempting to interpret change scores should heed the advice of Hermann et al. (1996) that regression statistics do not determine the clinical significance of the change, but only provide psychometric information on the statistical reliability of the change. This is particularly sage advice in the present instance since the MMSE/3MS represents a screening examination, whose major function is to refer low scoring individuals for further medical or neuropsychological examination. That is, a significant change score on a cognitive screening test is suggestive of but not prescriptive of a bona fide cognitive impairment. This caution underscores the conclusion set forth in the review article by Tombaugh and McIntyre (1992) that the MMSE, or in this case, a change in the MMSE, “should *not* serve as the sole criterion for diagnosing dementia” (p. 931).

There are several cautions in using these values. First, a 5-year test-retest interval was used. Since several studies have reported that the duration of the test-retest interval may affect retest performance on various cognitive tests (Cohen, Swerdlilk, & Smith, 1992; Dikmen et al., 1999; Unger et al., 1999), it may not be appropriate to employ the present set of norms for test-retest intervals that are substantially shorter than those used in the present investigation. However, comparison of the mean change scores from the current study with those obtained in other studies using a test-retest interval of greater than 1 year, suggests that the current set of normative data could be applied to any retest intervals greater than 1 year. Second, particular attention should be paid to the demographic composition of the sample. Even though the regression formula included the variables age and education, it is unknown if the same weighting will apply to a younger cohort who have had different educational experiences. Third, the effects of the intervening clinical MMSE/3MS that was administered between the two screening exams are largely unknown. As explained previously, the scores from the clinical exam but not from the screening exam were made available for the consensus diagnosis. In the present context, the major impact of the intervening administration was to produce a putative practice effect on the scores obtained from the second screening exam. However, since the second screening exam occurred 4.52 years (S.D. = .17 years) later, the likelihood of the intervening "practice" exerting a significant impact is, at best, minimal. Fourth, there may be some concerns about the appropriateness of applying the results when the standard MMSE is administered alone. Such a concern is based on the assumption that embedding the MMSE within the 3MS somehow contaminates or invalidates the MMSE score. Although this is possible, it is highly unlikely that the types of additions and changes made in the 3MS would exert any significant effect on how an individual performs on the core MMSE items. Finally, modification of some MMSE/3MS items raises the question as to whether the results would also apply to administrations of the mental status tests that do not employ the same modifications. Although this question cannot be answered directly by the current results, there are several sources of information that suggest the current findings have general applicability. For example, it can be argued that requiring the person to accept the paper in the non-preferred hand on the three-stage command, rather than in the right hand, increases the difficulty of the test. However, the finding that 92.6% (Test 1) and 85.4% (Retest) of participants received the paper with the correct hand, yielding a mean .07 point change between the two administrations shows that it is unlikely that the scores would have been much different had the original instructions been used. It also should be noted that the alternative words for registration and recall were selected on the basis of category similarity and were judged to be of comparable difficulty. Most of the other changes were made to "Orientation to Place" questions. For example, since all participants were administered the MMSE in the home, it would have been inappropriate in the CSHA, as well as with other home administrations, to ask "What is the name of this place" and/or "What floor are you on?". Consequently, the address (number and street name) was used as a substitute. Also "province" was substituted for "state". Awarding two points for the correct answer for "province" may have inflated the score for the 3MS, but not for the MMSE since only one point was awarded. However, since 98.5% (Time 1) and 97.6% (Time 2) of participants answered the "province" correctly, the differential weighting had virtually no effect on the total change score. A similar resistance to change was observed when the amount of change was computed for all five "Orientation to Place" questions. A .01 point change occurred on the 3MS (92.38% of scores showed no change), while .08 point changed on the MMSE (85.3% of scores remained unchanged). Thus, overall it appears that the current results have wide applicability and can provide the basis for a large database that will permit researchers and clinicians to better evaluate serial performance on the MMSE/3MS.

Acknowledgements

The author would like to thank Betsy Kristjansson, Ph.D. and Margaret Stewart for their assistance with managing the CSHA data set and to Denise Wosney, and Liz Sykes for their helpful suggestions made on an earlier version of the article.

References

Aguero, T. H., Fratiglioni, L., Guo, Z., Viitanen, M., & Winblad, B. (1998). Prognostic factors in very old demented adults: A seven-year follow-up from a population-based survey in Stockholm. *Journal of the American Geriatric Society*, *46*, 444–452.

Aevarsson, O., & Skogg, I. (2000). A longitudinal population study of the Mini-Mental State Examination in the very old: Relation to dementia and education. *Dementia and Geriatric Cognitive Disorders*, *11*, 166–175.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author (Review).

Bravo, G., & Hébert, R. (1997). Age- and education-specific reference values for the mini-mental and modified mini-mental state examination derived from a non-demented elderly population. *International Journal of Geriatric Psychiatry*, *12*, 1008–1018.

Canadian Study of Health Aging. (1994). The Canadian study of health and aging: Study methods and prevalence of dementia. *Canadian Medical Association Journal*, *150*, 899–913.

Chelune, G., Naugle, R. I., Luders, H., Sedlak, J., & Award, I. A. (1993). Individual change after surgery: Practice effects and base-rate information. *Neuropsychology*, *7*, 41–52.

Cohen, R. J., Swerdlik, M. E., & Smith, D. K. (1992). (2nd ed.). *Psychological testing and assessent:* Mountain View, California: Mayfield Publishing Company.

Crum, R., Anthony, J., Bassett, S., & Holstein, M. (1993). Population-based norms for the Mini-Mental State Examination by age and education level. *Journal of the American Medical Association*, *269*, 386–391.

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of Expanded-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, *5*, 346–356.

Escobar, J., Burnam, A., Karno, M., Forsythe, A., Landsverk, J., & Golding, J. M. (1986). Use of the Mini-Mental Status Examination (MMSE) in a community population of mixed ethnicity: Cultural and linguistic artifacts. *Journal of Nervous and Mental Disorders*, *174*, 607–614.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Hébert, R., Bravo, G., & Girouard, D. (1992). Validation de l'adaptation francaise du modified mini-mental state (3MS). *Geriatric Review*, *17*, 443–450.

Hermann, B. P., Seidenberg, M., Schoenfeld, J., Peterson, J., Leveroni, C., & Wyler, A. R. (1996). Empirical techniques for determining the reliability, magnitude, and pattern of neuropsychological change after epilepsy surgery. *Epilepsia*, *37*, 942–950.

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*, 12–19.

Keeting, J. J. (1987). "Studying" for the Mini-Mental Status Exam. *Journal of the American Geriatric Society*, *35*, 594–597.

McCaffrey, R. J., Duff, K., & Westervelt, H. J. (2000). *Practitioner's guide to evaluating change with neuropsychological assessment instruments*. New York: Kluwer Publishing.

McDowell, I., Helliwell, B., Sykes, E., Hill, G., & Lindsay, J. (2001). Study organization in the Canadian Study of Health and Aging. *International Psychogeriatrics*, *13*(Suppl. 1), 41–48.

McDowell, I., Hill, G., & Lindsay, J. (2001). An overview of the Canadian Study of Health and Aging. *International Psychogeriatriacs*, *13*(Suppl. 1), 7–18.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., et al. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

McSweeney, J., Naugle, R. I., Chelune, G. J., & Luders, H. (1993). "*T*-scores for change": An illustration of a regression approach to depicting change in neuropsychology. *Clinical Neuropsychology*, *7*, 300–312.

Mitrushina, M., & Satz, P. (1991). Reliability and validity of the Mini-Mental State Exam in neurologically intact elderly. *Journal of Clinical Psychology*, *47*, 537–541.

Olin, J. T., & Zelinski, E. M. (1991). The 12-month reliability of the Mini-Mental State Examination. *Psychological Assessment*, *3*, 427–432.

Salinsky, M. C., Storzbach, D., Dodrill, C. B., & Binder, L. M. (2001). Test-retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12-16 week period. *Journal of the International Neuropsychological Society*, *7*, 597–605.

Sawrie, S. M., Chelune, G. L., Naugle, R. I., & Luders, H. O. (1996). Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. *Journal of the International Neuropsychological Society*, *2*, 556–564.

Schmand, B., Lindeboom, J., Launer, L., Dinkgreve, M., Hooijer, C., & Jonker, C. (1995). What is a significant score change on the Mini-Mental State Examination? *International Journal of Geriatric Psychiatry*, *10*, 411–414.

Starr, J. M., Deary, I. J., Inch, S., Cross, S., & MacLennan, W. J. (1997). Age-associated cognitive decline in healthy old people. *Age Ageing*, *26*, 295–300.

Temkin, N. R., Heaton, R. K., Grant, I., & Dikman, S. S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*, *5*, 257–369.

Teng, E. L. (1990). *Manual for the administration and scoring of the Modified Mini-Mental State (3MS) test*. Unpublished manual.

Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) Examination. *Journal of Clinical Psychiatry*, *48*, 314–318.

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. (1996). Meni-Mental State Examination (MMSE) and the Modified MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment*, *8*, 4–59.

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatric Society*, *40*, 935–992.

Unger, J. M., van Belle, G., & Heyman, A. (1999). Cross-sectional versus longitudinal estimates of cognitive change in nondemented older people: A CERAD study. *Journal of the American Geriatric Society*, *47*, 559–563.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.