Psychological Assessment
1996, Vol. 8, No. 1, 48–59

1040-3590/96/$3.00

Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS): A Psychometric Comparison and Normative Data

T. N. Tombaugh
Carleton University

I. McDowell and B. Kristjansson
University of Ottawa

A. M. Hubley
Carleton University

The Mini-Mental State Examination (MMSE; M. F. Folstein, S. E. Folstein, & P. R. McHugh, 1975) and the Modified MMSE (E. L. Teng & H. C. Chui, 1987) were compared psychometrically. In this study, 525 community-dwelling participants, aged 65–89, were divided into 2 groups: no cognitive impairment (NCI; n = 406) and Alzheimer's disease (n = 119). Both tests yielded comparable reliability estimates. Fewer years of education decreased specificity and increased sensitivity, whereas increasing age primarily decreased specificity. It is concluded that although the 2 tests produce comparable effects, the inclusion of a verbal fluency test would increase the sensitivity of the MMSE. Normative data for the NCI group, stratified for 2 age levels (65–79 and 80–89) and 2 educational levels (0–8 and 9+ years), are presented.

Folstein, Folstein, and McHugh (1975) introduced the Mini-Mental State Examination (MMSE) as a brief, objective assessment of cognitive functioning and as a measure of changes in cognitive status. The MMSE usually can be administered in 5–10 min and has been employed extensively in clinical settings, community surveys, and epidemiological studies.

In a recent review of the literature, Tombaugh and McIntyre (1992) concluded that the MMSE possessed moderate to high reliability coefficients, demonstrated high levels of sensitivity for cognitive deficits in patients suffering from moderate to severe Alzheimer's disease, and reflected the cognitive decline typical of dementia patients. Criticisms of the MMSE included (a) its failure to discriminate between people with mild dementia and those who are not demented, (b) a limited ability to detect impairment caused by focal lesions, particularly those in the right hemisphere, (c) overly simple language items that reduce sensitivity to mild linguistic deficits, and (d) a large number of false-positive errors because of its bias against individuals with low education.

In response to these problems, several attempts have been made to improve the MMSE. Of these, the Modified Mini-Mental State Examination (3MS; Teng & Chui, 1987) represents the most extensive revision. Teng and Chui (1987) added four additional subtests (date and place of birth, word fluency, similarities, and delayed recall of words). The maximum score was increased from 30 to 100 points, and a modified scoring procedure permitted assignment of partial credit on some items. One of the advantages of the 3MS modifications is that both a 3MS and an MMSE score can be derived from a single administration (see the Appendix).

At the present time, however, minimal information is available on whether the 3MS modifications actually increase the clinical utility of the MMSE. In the original article, the results from a small number of patients indicated that the 3MS had increased sensitivity and specificity. In a subsequent article, Teng, Chui, and Gong (1990, p. 192) reported that the 3MS "was more reliable and more sensitive in detecting dementia than the MMSE." However, because the article did not contain any data on the individual items from either the 3MS or MMSE, it is not possible to determine whether the additional items or the revised scoring system contributed to the improvements. Finally, the limited normative data that are available were based on a relatively small number of community-dwelling individuals who had not been clinically assessed to ensure that they were, in fact, cognitively intact.

In order to provide a more complete evaluation of the psychometric properties of the MMSE and 3MS, the present study analyzed data from a subset (n = 525) of the community-dwelling individuals drawn from the Canadian Study on Health and Aging (CSHA; CSHA, 1994), a recently completed epidemiological survey that employed the 3MS. All 525 participants received a complete clinical examination, including an extensive battery of neuropsychological tests. The present psychometric evaluation focused on the effects of age, education, and gender on both individual test items and total scores. Not only do these analyses permit a comparison between the MMSE and 3MS, but they also provide valuable norms from a sample of community-dwelling adults who, on the basis of a series of medical, laboratory, and neuropsychological tests, were diagnosed to be cognitively intact.

T. N. Tombaugh and A. M. Hubley, Department of Psychology, Carleton University, Ottawa, Ontario, Canada; I. McDowell and B. Kristjansson, Faculty of Medicine, Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.

We thank Penny Faulkner for all of her help. Special thanks go to Peter Aubin and Bill Petrusic for their statistical advice and assistance.

Correspondence concerning this article should be addressed to T. N. Tombaugh, Department of Psychology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada.

Method

Procedure

As part of a large epidemiological study of dementia (CSHA, 1994), the 3MS was administered to a representative sample of 8,949 community-dwelling individuals who were drawn from five geographical regions in Canada. All individuals were over the age of 64 and were given the option of taking either the English (n = 7,209) or the French (n = 1,740) language version. All participants with a 3MS score of less than 78 (n = 1,106) were referred for a clinical assessment that was designed to confirm the presence of cognitive impairment and to provide a differential diagnosis. The cutpoint of 77 was selected following the results of a pilot test and was designed to ensure high sensitivity for individuals suffering from Alzheimer's disease (AD). An additional sample of 494 randomly selected individuals having a score greater than 77 was also assessed clinically.

The data presented in the present study are based on a subset of 525 individuals who fell into two diagnostic categories: no cognitive impairment (NCI; n = 406) and AD (n = 119). Consensus diagnoses were reached by physicians and clinical neuropsychologists on the basis of history, clinical and neurological examination, and an extensive battery of neuropsychological tests that assessed memory, language, praxis, attention, visual perception, problem-solving, judgment, and social functioning.

The NCI group contained individuals who had been diagnosed as not suffering from any type of cognitive impairment. The AD group consisted of individuals with possible and probable AD as defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA; McKhann et al., 1984).

The 3MS was administered and scored according to the guidelines set forth in the administrative manual (Teng, 1990). The 3MS and MMSE scores were both derived from the 3MS (see the Appendix). However, four of the subscales employed in the MMSE were slightly modified in the 3MS (Concentration: spell *WORLD* backward replaced serial 7s; Recall: *shirt, brown,* and *honesty* replaced *penny, apple,* and *table*; Orientation to Time: *province* was substituted for *state* and awarded 2 points rather than 1, and *type of building* replaced *floor of building*; Three-Stage Command: nonpreferred hand was used rather than right hand).

Participants

All participants completed the English version of the 3MS, stated that English was their preferred language, and were judged by the interviewer to be proficient in English. Participants who were blind or illiterate were excluded. The age range was restricted to participants aged 65 to 89 (M = 79.4, SD = 6.2). Number of years of education varied from 0 to 25 (M = 10.9, SD = 3.9). The overall male:female ratio was 307:437. The AD group was significantly ($p < .01$) older (M = 81.9, SD = 5.1) than the NCI group (M = 78.5, SD = 6.0) and contained a greater proportion of women (AD = 69%; NCI = 57%). The NCI group had significantly more years of education (M = 10.6, SD = 3.8) than the AD group (M = 8.9, SD = 9.1).

Data Analyses

For descriptive purposes, ages were divided into five age ranges (65–69, 70–74, 75–79, 80–84, and 85–89), and years of education were divided into three groups (0–8, 9–12, and >12 years). Although grouping the participants into five age groups and three education groups was useful at the descriptive level, the number of AD participants in many cells was prohibitively small to use the same classification scheme for calculation of other statistics such as sensitivity, positive predictive power, and negative predictive power. For these analyses, as well as for the normative data, two age groups (65–79 and 80–89) and two education groups (0–8 years and 9+ years) were used. The selection of these groups was guided by visual inspection of the data and by a review of the MMSE literature (Tombaugh & McIntyre, 1992).

Results

Effects of Age, Education, and Gender on Total Scores

Table 1 shows the mean mental status scores for five levels of age ranges, three levels of education, and gender for the AD and NCI groups. Scores for the MMSE are shown in the upper half, and scores for the 3MS are in the lower half.

In order to determine the effects of age, education, and gender on mental status scores, a separate multiple regression analysis was performed for each diagnostic category. The results from the regression analyses were similar for both mental status tests.

Table 1
Means and Standard Deviations for Education, Gender, and Age for Different Diagnostic Categories on the MMSE and 3MS

	Diagnostic category			
	No cognitive impairment		Alzheimer's disease	
Category	*M*	*SD*	*M*	*SD*
	MMSE			
Age (years)				
65–69[a]	25.5	4.4	20.3	1.5
70–74[a]	26.7	3.2	19.0	3.9
75–79	26.8	2.6	19.8	5.3
80–84	26.1	2.9	19.6	4.9
85–89	25.0	2.9	19.1	4.9
Education level (years)				
0–8	24.3	3.1	19.1	4.2
9–12	26.7	2.8	20.6	5.0
>12	27.5	2.3	20.1	6.3
Gender				
Men	25.5	3.2	20.5	3.6
Women	26.6	2.9	18.9	5.3
Total	26.1	3.1	19.4	4.9
	3MS			
Age				
65–69[a]	83.1	12.4	63.7	2.9
70–74[a]	87.8	9.7	56.3	12.6
75–79	87.5	8.7	62.2	16.4
80–84	84.3	10.1	60.8	11.8
85–89	80.7	9.8	60.6	13.4
Education level (years)				
0–8	79.1	9.2	60.7	11.4
9–12	86.2	9.6	59.7	14.2
>12	90.1	8.3	64.0	14.2
Gender				
Men	83.7	10.3	65.9	11.5
Women	85.9	9.9	58.9	13.9
Total	84.9	10.1	60.9	13.5

Note. For the no cognitive impairment category, n = 406; for the Alzheimer's disease category, n = 119. MMSE = Mini-Mental State Examination; 3MS = Modified MMSE.
[a] Less than 10 participants per group for the Alzheimer's disease category.

For the NCI category, increased age and decreased education were correlated with significantly lower scores on the MMSE and 3MS. Gender played a role only with the MMSE, with women obtaining higher scores. When R^2 difference scores were computed ($\alpha < .01$) to determine the pure effects of each variable with the others statistically controlled, neither age nor gender contributed a significant proportion of the variance over and above that accounted for by education. For the AD category, the only significant finding was a gender effect for the 3MS test, with men obtaining higher scores.

Reliability (Internal Consistency)

Moderate to high Cronbach alpha coefficients occurred for both tests (NCI: 3MS = .82, MMSE = .62; AD: 3MS = .88, MMSE = .81), indicating that they possessed a large number of subtests measuring the same construct. The consistently higher alphas for the 3MS total score reflect, at least in part, its larger number of subtests.[1]

Differential Sensitivity of the 3MS and MMSE

The differential sensitivity of the two mental status tests was directly compared using receiver operating characteristic (ROC) curves, where the probability of a hit (sensitivity) was plotted as a function of the probability of a false alarm (1-specificity). ROC curves were computed for the 2 × 2 factorial combination of age (65–79 and 80–89) and education (0–8 years and 9+ years) for each mental status test using the signal detection analysis developed by Metz and Kronman (1980). These curves are shown in Figure 1. Approximate bivariate X^2 statistics were calculated from maximum likelihood estimates of the two ROC curve parameters, intercept and slope, using the Dorfman program (Dorfman & Alf, 1969). The basic sensitivity statistic used for descriptive analysis was the area under the ROC curve (i.e., A(Z)). An A(Z) of 0.5 indicates a complete lack of sensitivity, and an A(Z) of 1.00 indicates perfect sensitivity.

A descriptive examination of Figure 1 shows that the 3MS was better at discriminating between individuals with and without AD in three out of the four individual graphs. Differences between the 3MS and MMSE were especially pronounced for the age groups with lower education (80–89 years old: .859 vs. .803, respectively; 65–79 years old: .946 vs. .907, respectively). However, none of the statistical tests were significant. Lack of significance might have been due to low power as a result of large discrepancies in sample sizes as well as very small sample sizes. Collapsing across age groups, the 3MS exhibited greater, although not statistically significant, sensitivity than the MMSE for individuals with a lower level of education (0–8 years: .901 vs. .848, respectively; 9+ years: .941 vs. .933, respectively). The two measures were identical for individuals with higher education. Collapsing across education levels indicated virtually no differential sensitivity between the 3MS and the MMSE with respect to age groups (65–79 years old: .939 vs. .919, respectively; 80–89 years old: .880 vs. .897, respectively). Finally, an overall analysis after collapsing across education and age (Figure 2) suggested a nonsignificant advantage of the 3MS over the MMSE (.926 vs. .905, respectively).

Subscale Discriminability

A logistic regression analysis was employed to identify those subscales that best differentiated between cognitively intact individuals and those suffering from AD. A logistic regression analysis was employed rather than a discriminant function analysis because of the dichotomous dependent variable. Separate analyses were performed with the language items combined into a single category and as separate items. Because the results of both procedures were identical, only the results with all language items grouped as one subscale are reported.

Because previous research has shown that the discriminability of subscales is affected by the severity of cognitive impairment (Tombaugh & McIntyre, 1992), the MMSE cutoff scores typically used to classify mild and severe AD were employed to divide AD participants into two subgroups representing mild (MMSE = ≥18) and severe (MMSE = <18) cognitive impairment. Previous research, as well as correlational analyses performed in the present study, also has shown that scores on subscales are differentially affected by age and education. Consequently, the effects of age and education were controlled by matching each AD participant with an NCI participant on these two variables. No NCI participant was used in both control groups. This matching procedure resulted in less than a half-year difference between the mean age and education of each Alzheimer's group and its corresponding control group.

Table 2 shows the sensitivity and specificity achieved for each model using stepwise entry of the variables. Sensitivity refers to the ability of the 3MS and MMSE to correctly identify participants who had been classified as suffering from AD according to the NINCDS-ADRDA criteria (e.g., true positives divided by total number of AD participants). Specificity refers to the ability of the two mental status tests to correctly identify individuals who had been classified as cognitively intact by consensus diagnosis (e.g., true negatives divided by total number of intact participants).

The pattern of subtests that discriminated AD participants from their matched NCI controls was remarkably similar for the MMSE and the 3MS. Both tests showed that Orientation to Time, First Recall, and Pentagons entered the regression equation for the mild AD group, whereas Orientation to Time entered the regression equation for the severe AD group. The logistic regressions also showed that of the four subscales added to the 3MS, only Animal Naming significantly increased sensitivity and specificity. Moreover, the increased discriminability of Animal Naming is primarily attributable to its ability to differentiate between the NCI and the mild AD participants.

Normative Data

The fact that the NCI group contained a relatively large number of participants ($n = 406$) who originally were part of a randomly drawn sample and who subsequently had been judged to be cognitively intact on the basis of an extensive clinical examination makes it an ideal sample for providing normative data for the two mental status tests. On the basis of the previous analyses, it was decided to stratify the norms on the basis of age (65–

[1] Coefficient alphas for each subscale are available on request from T. N. Tombaugh.

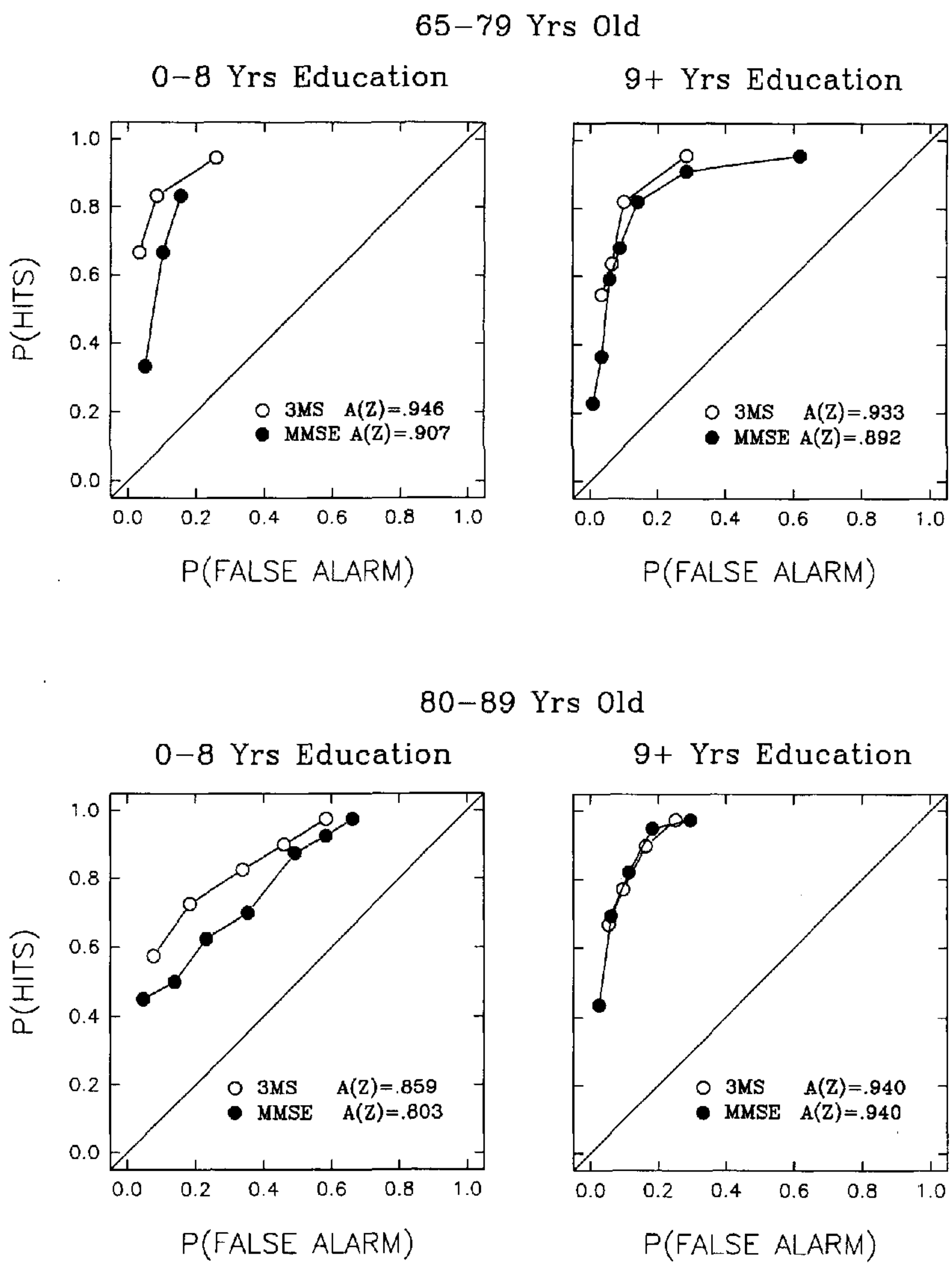

Figure 1. Receiving operator characteristic curves showing the relationship between probability (P) of a hit (true positive) and a false alarm (false positive) for the Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS) when two levels of age (65–79 years and 80–89 years) were factorially combined with two levels of education (0–8 years and 9+ years). A(Z) = area under the curve.

79 and 80–89) and education (0–8, 9+ years), but not gender. The percentile equivalents[2] for the MMSE and 3MS normative data are shown in Tables 3 and 4.

The normative data provide a clear picture of how the distribution of scores for cognitively intact individuals vary as a function of age and education. The use of normative tables is greatly enhanced when they can be converted to specificity scores and compared to other scores showing sensitivity, positive predictive power (true positives divided by true positives plus false positives), and negative predictive power (true negatives divided by true negatives plus false negatives) for different patient groups. These data are presented in Tables 5 and 6.

[2] The lower class interval of each score was used to compute the percentiles.

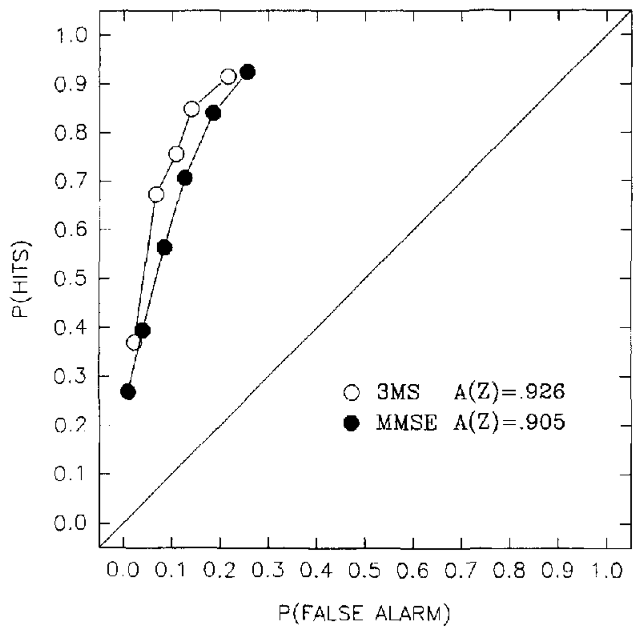

Figure 2. Receiving operator characteristics curve showing the relationship between probability (P) of a hit (true positive) and a false alarm (false positive) for the Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS). A(Z) = area under the curve.

It should be noted that the values reported here are based solely on a restricted sample of elderly individuals who have been clinically diagnosed as either cognitively intact or cognitively impaired.[3] These tables permit establishing multiple cutoffs corresponding to different ages and educational levels for both the MMSE and 3MS.

Discussion

Teng and Chui (1987) published the 3MS in an attempt to enhance the usefulness of the MMSE. In a subsequent article, Teng et al. (1990) presented data showing that the sensitivity and specificity of the 3MS was higher than that obtained for the MMSE. The present study was undertaken to provide further evidence on the psychometric properties of these two tests, thereby permitting a more comprehensive comparison than is available currently. One of the unique aspects of the present study is that all of the 525 participants (including those without cognitive impairment) received a complete medical examination, including a detailed medical history and an extensive series of neuropsychological and laboratory tests. Thus, the control participants, as well as those with clinical diagnoses, received a comparable series of medical and neuropsychological tests.

Reliability

Cronbach's coefficient alphas, a measure of internal consistency, indicate that both the MMSE and the 3MS contain relatively homogeneous subscales that measure similar cognitive domains. Equivalent alpha coefficients were obtained on both mental status tests for groups suffering from AD, whereas higher reliability coefficients were observed for the 3MS than for the MMSE in the NCI group.

Subscale Discriminability

The ability of individual subscales to differentiate between individuals with AD and those without cognitive impairment was assessed using participants matched for age and education. This represents an important methodological advance over previous studies that did not control for these potentially confounding variables. Overall, Recall 1, Orientation to Time, and Pentagons possessed a high degree of discriminability for both the MMSE and the 3MS. However, the ability of the subtests to discriminate between NCI and AD participants was determined, to a large degree, by the severity of cognitive impairment. These same three subscales showed the same degree of relative discriminability for AD participants suffering mild cognitive loss. A different picture emerged with severely impaired AD participants. Here, only Orientation to Time possessed discriminability for both mental status tests. Of the new 3MS subscales, only one (Animal Naming) possessed a high degree of discriminability overall and added to the ability of the 3MS to discriminate between NCI and mild AD participants.

ROC curves showed that the MMSE and 3MS were not significantly different in their ability to correctly identify AD patients. Thus, the present study concludes that the two tests are not differentially sensitive to AD. Moreover, any advantage that the 3MS might have over the MMSE is probably attributable to the effectiveness of the Animal Naming subscale in identifying mildly impaired AD individuals. The higher degree of discriminability associated with the Animal Naming subscale is consistent with previous experimental findings that have shown that verbal fluency tests are very sensitive measures of the type of cognitive impairment that is associated with AD (Butters, Granhold, Salmon, Grant, & Wolfe, 1987; Martin & Fedio, 1983; Ober, Dronkers, Koss, Delis, & Friedland, 1986; Rosen, 1980; Weingartner et al., 1981). In fact, verbal fluency tests frequently are used with the MMSE to increase its sensitivity. The results from the present study support this practice.

Several modifications incorporated in the 3MS did not substantially increase its utility. Some of the new questions simply were too easy. For example, participants in the NCI and mild AD groups rarely failed to provide the correct answer for Date and Place of Birth. Even severe AD participants scored higher on this subscale than on any other subscale. This is probably because knowing one's date and place of birth is a highly practiced answer and is very resistant to disruption by the disease process. Similarly, virtually all of the NCI and mild AD participants, and 85% of the severe AD individuals, were able to count backwards from 5 to 1. Inspection of the scores for the changes made to the Language subscale showed that these modifications did not increase the low sensitivity of the Language subscale that has been reported consistently in other studies (Tombaugh & McIntyre, 1992).

The lack of discriminability for the Similarities subscale, one

[3] Those interested in sensitivity and specificity data based on a random sample using a correction for verification bias that allows the results to be generalized to an older population are referred to Kristjansson, McDowell, Hebert, and Bravo (1993).

Table 2
Summary of Logistic Regression Analysis for Participants With Alzheimer's Disease (AD) and No Cognitive Impairment (NCI) Matched on Age and Education

	All AD and NCI		Mild AD[a] and NCI		Severe AD[b] and NCI	
Subscale	MMSE	3MS	MMSE	3MS	MMSE	3MS
Orientation Time	X	X	X	X	X	X
Orientation Place						
Registration					X	
Concentration	X	X				
First Recall	X	X	X	X		
Language						
Pentagons	X	X	X	X		
Date and Place of Birth[c]						
Animals[c]		X		X		
Similarities[c]						
Second Recall[c]						
Sensitivity (%; AD)	81	87	81	89	96	86
Specificity (%; matched NCI)	83	85	77	82	100	96

Note. Xs indicate those subscales that entered into the regression formula for each group. Sensitivity and specificity scores for the subscales that entered into the equation were calculated using the frequency distribution tables generated by the logistic regression. MMSE = Mini-Mental State Examination; 3MS = Modified MMSE.
[a] Mild AD corresponds to MMSE scores ≥18. [b] Severe AD corresponds to MMSE scores <18.
[c] Subscales present only on 3MS.

of the new subscales, is probably because so few participants without any cognitive impairment were able to answer it correctly. The distribution of the scores for each word pair revealed that approximately 75% of the NCI participants failed to answer the question correctly. This low base rate, in turn, decreased the likelihood that scores from the AD groups would possess a high degree of discriminability. It is uncertain if this lack of sensitivity is attributable to the scoring system used, the type of word pairs (e.g., *eat–sleep*) employed, or the level of skill required for this type of task.

The failure of Recall 2 to enter the regression equation is surprising in view of the evidence showing that delayed recall is one of the most sensitive measures of memory impairment. Comparison of the scores for Registration, First Recall, and Second Recall showed that for all three groups, a substantial decrease in scores occurred from initially learning the three words (i.e., Registration) to the first retention test (i.e., First Recall). However, virtually no additional change in scores occurred on the second retention test (i.e., Second Recall), showing that the amount forgotten did not change from the first to second recall test. Thus, the information contained in Second Recall was redundant with that in First Recall.

Among changes to the scoring procedures, three are particularly noteworthy. First, the scoring of the Pentagons was changed from a maximum of 1 point to a maximum of 10 points, with each pentagon and the intersection of the pentagons scored separately. A comparison of the MMSE and 3MS percentage change scores for the Pentagons showed that little difference existed between them, with the MMSE being slightly superior for the mild AD group (19% vs. 16%) and severe AD group (49% vs. 43%). This suggests that the modified scoring procedure did not substantially alter the discriminability of the subscale from that observed with the original MMSE scoring procedure. This suggestion is supported by the moderate to high point-biserial correlations that occurred between the MMSE and 3MS Pentagon scores for the different groups (mild AD = .78, NCI controls = .78; severe AD = .52, NCI controls = .66).

Second, the weighting of the answers for three of the Orientation to Time subscale items was modified so that credit could

Table 3
MMSE Norms (Percentile Scores) Stratified for Age and Years of Education for Participants Diagnosed as No Cognitive Impairment

	Age 65–79		Age 80–89	
MMSE score	0–8 years (*n* = 58)	9+ years (*n* = 168)	0–8 years (*n* = 65)	9+ years (*n* = 115)
30	98	86	100	93
29	88	62	97	77
28	76	41	89	57
27	62	29	83	37
26	48	21	66	30
25	36	14	58	18
24	26	9	49	11
23	19	6	35	6
22	16	5	23	4
21	10	4	14	3
20	5	4	8	1
19	5	3	5	<1
18	4	1	<5	
17	4	<1		
16	3			
<16	<3			

Note. MMSE = Mini-Mental State Examination.

Table 4
3MS Norms (Percentile Scores) Stratified for Age and Years of Education for Participants Diagnosed as No Cognitive Impairment

	Age 65–79		Age 80–89	
3 MS Score	0–8 years (*n* = 58)	9+ years (*n* = 168)	0–8 years (*n* = 65)	9+ years (*n* = 115)
100	100	98	100	98
99	100	95	100	97
98	100	86	100	95
97	100	79	100	90
96	95	73	100	88
95	93	67	99	82
94	90	61	99	77
93	86	56	95	70
92	83	52	91	61
91	79	48	90	58
90	78	43	84	54
89	76	38	82	51
88	72	36	81	49
87	69	32	80	42
86	65	29	80	39
85	60	26	80	38
84	57	24	77	37
83	53	22	72	37
82	48	22	71	34
81	47	20	69	31
80	45	20	68	30
79	44	19	66	29
78	43	19	63	28
77	38	16	59	25
76	26	10	51	20
75	17	8	46	17
74	12	7	42	12
73	10	7	38	10
72	9	5	34	7
71	8	5	32	5
70	6	4	25	2
69	5	4	22	<2
68	4	3	19	
67	3	2	12	
66	<3	<2	9	
65			8	
64			3	
63			<2	
62				
61				
60				
<60				

Note. 3MS = Modified Mini-Mental Examination.

be given for answers that differed in their degree of "correctness." For example, answers for "date" were scored on a 4-point continuum (3 = correct, 2 = incorrect 1–2 days, 1 = incorrect 3–5 days, 0 = missed >5 days). Similar procedures were used for "month" (2–0) and "year" (8–0). The overall effect of these changes was an increased discriminability between the NCI and the severe AD group for Orientation to Time. This was largely due to the increased discriminability attributed to "year." However, because Orientation to Time already constituted the most discriminating subscale for severely impaired AD participants, the relative merits of the additional discriminability are diminished.

Third, the administration and scoring of First Recall and Second Recall were modified so as to assess word recall by three different procedures (i.e., free recall, cued recall, and recognition). That is, if a word is not recalled, the participant is provided with a cue to facilitate recall (e.g., "It is a name of a color"). If the word still is not recalled, the participant is asked to select the word from one of three alternatives (e.g., "Blue, Black, Brown"). Although procedures similar to these have been shown to reveal important information about an individual's memory abilities (Tombaugh & Schmidt, 1992; Delis, 1989), they did not increase scale discriminability in the present case.

Age and Education

Typically, mental status tests have classified individuals as cognitively intact or impaired on the basis of a fixed cutoff criterion that is not adjusted for education and age. For example, it is common practice to classify individuals as cognitively intact if their MMSE score is 24 or greater. However, the current results show that the chances of a person without cognitive impairment being correctly classified as cognitively intact are highly dependent upon age and education level. This is best illustrated in Table 5, in which the MMSE normative data were transformed into the percentage of cognitively intact individuals who were correctly classified (i.e., specificity). Using the generally accepted cutoff score of 23, approximately 90% of the younger and older individuals with 9 or more years of education were correctly identified. However, if the participant had less than 9 years of education, only 74% of the younger group and 51% of the older group were correctly identified. A similar situation occurs with the 3MS when individuals were classified as cognitively intact using a criterion score of 77 (Table 6). In that case, 81% of the younger participants and 72% of the older individuals with 9 or more years of education were appropriately classified, but only 57% of the younger participants and 37% of the older participants with 0–8 years of education were correctly identified.

Interestingly, neither education nor age affected scores for AD participants on either the MMSE or the 3MS. This suggests that these variables no longer bias test performance after the onset of dementia. In the case of education, perhaps the changes in mental status produced by dementia are so similar to those caused by lower levels of education that they either mask or duplicate their effects. Certainly, these findings should not be misinterpreted to mean that age and education are not associated with the prevalence or incidence of AD or that they do not constitute a risk factor for AD. Age is well accepted as a risk factor for AD, and there is growing evidence that education plays a similar role (see Berkman, 1986; Kittner, White, & Farmer, 1986).

Norms

There are two strengths to the normative information presented here. First, the present study contains normative data from a relatively large number of participants who were classified as cognitively intact on the basis of an extensive battery of neuropsychological and medical tests. To the best of our knowledge, no other study has presented norms in which cognitively intact participants have undergone such an extensive series of tests. Second, because both age and education were shown to

Table 5
Sensitivity, Specificity, Positive Predictive Power, and Negative Predictive Power for the MMSE

MMSE criterion score	Education							
	0–8 years				9+ years			
	SEN	SPE	PPP	NPP	SEN	SPE	PPP	NPP
	Age 65–79							
27	100	24	29	100	96	59	23	98
26	100	38	33	100	93	71	30	98
25	100	52	39	100	91	79	36	99
24	100	64	46	100	82	86	46	97
23	100	74	55	100	68	91	55	96
22	89	81	59	96	59	94	60	95
21	83	84	63	94	52	95	62	93
20	67	90	67	90	46	96	59	93
19	33	95	67	82	36	96	59	92
18	28	95	63	81	27	98	62	91
17	24	96	60	79	25	99	86	91
	Age 80–89							
27	100	10	41	100	100	43	37	100
26	100	17	43	100	100	63	48	100
25	98	34	48	96	97	70	52	99
24	93	42	49	90	95	82	63	98
23	88	51	52	87	82	89	71	94
22	70	65	55	78	69	94	79	90
21	63	77	63	77	44	96	77	83
20	50	86	69	74	39	97	83	82
19	48	92	79	74	36	98	93	82
18	45	95	86	74	28	98	100	80
17	35	96	82	71	26	100	100	80

Note. All results are given in percentages. For the age 65–79 category, n = 18 for Alzheimer's disease (AD) participants and n = 58 for no cognitive impairment (NCI) participants with 0–8 years education, and n = 22 for AD participants and n = 168 for NCI participants with 9+ years education. For the age 80–89 category, n = 40 for AD participants and n = 65 for NCI participants with 0–8 years education, and n = 39 for AD participants and n = 115 for NCI participants with 9+ years education. MMSE = Mini-Mental State Examination; SEN = sensitivity (number of AD participants correctly identified by the score to have AD divided by total number of AD participants); SPE = specificity (number of NCI participants correctly identified by the score to be "normal" divided by total number of NCI participants); PPP = positive predictive power (number of participants correctly identified by the score to have AD divided by the total number of participants correctly and incorrectly identified with AD); NPP = negative predictive power (number of participants correctly identified by the score to be "normal" divided by the total number of participants correctly and incorrectly identified as "normal").

influence scores on both mental status tests, the current set of norms were stratified across two age groups (65–79 and 80–89) and two educational levels (0–8 and 9+ years). Given that a modified version of the MMSE was used in the present study (see the Method section), it could be argued that these norms may not be directly applicable when the original version is administered. However, it is unlikely that these modifications substantially altered the content of the original MMSE. Moreover, some of these modifications have been incorporated in other versions of the MMSE (Tombaugh & McIntyre, 1992). Consequently, it is reasonable to assume that the present set of norms has a wide range of application.

The age and education results presented here, as well as those presented elsewhere (Anthony, LaResche, Niaz, von Korff, & Folstein, 1982; Escobar et al., 1986; Teng et al., 1990; Uhlman & Larson, 1991), suggest that it might be useful to employ different cutoff scores for different age and educational levels. The normative data generated in the present study permit the establishing of multiple cutoff scores for both the MMSE and the 3MS. It is important to remember, however, that cutoff scores cannot be determined solely on the basis of the percentage of cognitively intact participants who are identified. The ability of the tests to correctly identify cognitively impaired individuals must also be considered. Traditionally, this has been accomplished by computing the percentage of impaired participants that a specific score correctly identifies (i.e., sensitivity) and comparing this to the percentage of nonimpaired participants that the score correctly identifies (i.e., specificity). For example, if the MMSE cutoff (criterion) score was 23, sensitivity represents the percentage of AD patients whose scores fall at or below 23 (i.e., 23–0), and specificity refers to the percentage of NCI individuals whose scores were higher than 23 (i.e., 24–30). Although sensitivity and specificity provide valuable psychometric information about how well a test score identifies individuals within a given diagnostic category, they have limited clinical applicability because they do not predict how well a specific test score identifies people who are from different categories. That is, sensitivity and specificity do not provide any information about whether a score will identify a cognitively impaired patient from a group that contains both cognitively intact and impaired patients. This type of information is provided by the positive predictive power (PPP). Likewise, the ability of a score to identify a cognitively intact individual from a mixed group of participants is provided by negative predictive power (NPP). PPP refers to the percentage of patients that are truly impaired out of those that the test score classified as impaired [true positives/(true positives + false positives)], and NPP refers to the percentage of patients who are truly not impaired out of those that the test score classified as not impaired [true negatives/(true negatives + false negatives)]. Thus, in the present study, a PPP of 27% and NPP of 90% mean that only 27% of the people that the test classified as AD actually had been previously diagnosed as having AD according to NINCDS-ADRDA criteria, and 90% of the people classified as cognitively intact had been previously diagnosed as cognitively intact.

In order to provide the greatest clinical utility for the data in the present study, the specificity values for the normative data are presented in Tables 5 and 6, along with sensitivity, PPP, and NPP values. The usefulness of these data to establish different sets of cutting scores depends on at least two factors. First, it relies on a cost–benefit analysis to determine whether detection errors in sensitivity or specificity would be more "acceptable." The selection of a cutting point depends on the relative benefits of detecting a dementia case, compared to the disadvantages of labeling a normal person as impaired. The former hinges on how treatable the dementia is; that is, does early detection imply a better prognosis? The latter considers the possible distress that may be caused by a fuller clinical assessment, as well as the financial cost involved. The present data certainly offer an opportunity to reduce the margin of error involved in this type of determination that is not afforded with the currently employed

Table 6
Sensitivity, Specificity, Positive Predictive Power, and Negative Predictive Power for the 3MS

3MS criterion score	Education							
	0–8 years				9+ years			
	SEN	SPE	PPP	NPP	SEN	SPE	PPP	NPP
				Age 65–79				
85	100	35	34	100	100	71	33	100
84	100	40	35	100	100	74	35	100
83	100	43	37	100	100	76	37	100
82	100	47	39	100	100	78	39	100
81	100	52	40	100	100	79	39	100
80	100	53	40	100	100	80	40	99
79	100	54	42	100	98	81	40	99
78	100	56	42	100	97	81	40	99
77	100	57	45	100	96	81	46	99
76	100	62	55	100	86	85	53	98
75	94	74	63	98	82	90	56	97
74	92	83	70	96	73	92	57	96
73	91	88	73	96	70	93	58	95
72	89	90	76	96	67	94	64	95
71	86	91	79	95	64	95	64	95
70	83	93	79	95	59	95	65	95
69	80	94	81	92	58	96	63	94
68	76	95	81	92	56	96	80	94
67	72	95	87	92	55	98	80	94
66	69	97	86	92	53	98	79	94
65	67	97	86	90	52	98	79	94
				Age 80–89				
85	100	20	44	100	100	61	47	100
84	100	23	46	100	100	62	48	100
83	100	28	47	100	100	63	49	100
82	100	29	47	100	100	64	50	100
81	100	30	47	100	100	66	52	100
80	100	31	48	100	100	69	53	100
79	100	32	48	100	100	70	54	100
78	100	34	49	100	100	71	55	100
77	100	37	51	100	100	72	57	100
76	98	42	54	97	97	75	63	99
75	93	49	55	92	95	80	67	98
74	91	54	57	91	93	84	72	97
73	90	58	60	91	90	88	75	96
72	88	62	61	90	77	90	79	92
71	83	66	62	87	72	93	82	91
70	80	68	67	86	67	95	92	90
69	79	75	69	85	62	98	96	88
68	78	78	72	86	59	99	100	88
67	73	82	74	83	56	100	100	87
66	63	84	81	80	54	100	100	87
65	60	91	83	79	51	100	100	86

Note. For the 65–79 age category, n = 18 for Alzheimer's disease (AD) participants and n = 58 for no cognitive impairments (NCI) participants with 0–8 years education, and n = 22 for AD participants and n = 168 for NCI participants with 9+ years of education. For the 80–89 age category, n = 40 for AD participants and n = 65 for NCI participants with 0–8 years education, and n = 39 for AD participants and n = 115 for NCI participants with 9+ years of education. 3MS = Modified Mini-Mental State Examination; SEN = sensitivity (number of AD participants correctly identified by the score to have AD divided by total number of AD participants); SPE = specificity (number of NCI participants correctly identified by the score to be "normal" divided by total number of NCI participants); PPP = positive predictive power (number of participants correctly identified by the score to have AD divided by the total number of participants correctly and incorrectly identified with AD); NPP = negative predictive power (number of participants correctly identified by the score to be "normal" divided by the total number of participants correctly and incorrectly identified as "normal").

fixed, cutoff criterion that is not adjusted for age and education. Within this context, it must be noted that the MMSE and the 3MS were designed as screening tests to detect cognitive impairment. The tests were not intended to be used for the diagnosis of dementia and should not be employed in that capacity.

The second factor concerns the interpretation of PPP and NPP values. Although these measures are clinically useful, they are very sensitive to baseline rates of the illness and should be interpreted cautiously (Baldessarini, Finklestein, & Arana, 1983; Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984). For example, a PPP of 76% is excellent if 30% of all patients in a given setting (e.g., general geriatric ward) have AD, but it is less than adequate if, in another setting (e.g., Alzheimer's and related disease ward), 85% of all patients have AD. Thus, the clinical usefulness of the PPP and NPP values varies among clinical settings and the prevalence of the specific illness. In a similar vein, the calculation of the PPP and NPP values in the present article reflects the prevalence (i.e., AD:NCI ratio) in the various age and educational groupings used in the current sample.

References

Anthony, J. C., LaResche, L., Niaz, U., von Korff, M. R., & Folstein, M. F. (1982). Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients. *Psychological Medicine, 12,* 397–408.

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40,* 569–573.

Berkman, L. F. (1986). The association between educational attainment and mental status examinations: Of etiologic significance for senile dementias or not? *Journal of Chronic Diseases, 39,* 171–174.

Butters, N., Granhold, E., Salmon, D. P., Grant, I., & Wolfe, J. (1987). Episodic and semantic memory: A comparison of amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology, 9,* 479–497.

Canadian Study of Health and Aging. (1994). The Canadian study of health and aging: Study methods and prevalence of dementia. *Canadian Medical Association Journal, 159,* 899–913.

Delis, D. C. (1989). Neuropsychological assessment of learning and memory. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 3, pp. 3–32). Amsterdam: Elsevier Science.

Dorfman, D. D., & Alf, E. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals: Rating method data. *Journal of Mathematical Psychology, 6,* 487–496.

Escobar, J. I., Burnam, A., Karno, M., Forsythe, A., Landsvark, 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 Disease, 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.

Kittner, S. J., White, L. R., & Farmer, M. E. (1986). Methodological issues in screening for dementia: The problem of education adjustment. *Journal of Chronic Diseases, 39,* 163–170.

Kristjansson, E. A., McDowell, I., Hebert, R., & Bravo, G. (1993, August). *Screening tests for dementia: The Mini-Mental State and the Modified Mini-Mental State Examination in a community survey.* Paper presented at the 101st Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and Language, 19,* 124–141.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (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.

Metz, C. E., & Kronman, H. B. (1980). Statistical significance tests for binormal ROC curves. *Journal of Mathematical Psychology, 22,* 218–243.

Ober, B. A., Dronkers, N. F., Koss, E., Delis, D., & Friedland, R. P. (1986). Retrieval from semantic memory in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology, 8,* 75–92.

Rosen, W. G. (1980). Verbal fluency in aging and dementia. *Journal of Clinical Neuropsychology, 2,* 135–146.

Teng, E. T. (1990). *Manual for the administration and scoring of the Modified Mini-Mental State (3MS) test.* Unpublished manual, School of Medicine, University of Southern California, Los Angeles.

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

Teng, E. L., Chui, H. C., & Gong, A. (1990). Comparisons between the Mini-Mental State Exam (MMSE) and its modified version—The 3MS test. *Psychogeriatrics: Biomedical and Social Advances* (pp. 189–192). Tokyo: Excerpta Medica.

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society, 40,* 922–935.

Tombaugh, T. N., & Schmidt, J. P. (1992). The Learning and Memory Battery (LAMB): Development and standardization. *Psychological Assessment, 4,* 193–206.

Uhlman, R. F., & Larson, E. B. (1991). Effects of education on the Mini-Mental State Examination as a screening test for dementia. *Journal of the American Geriatric Society, 39,* 876–880.

Weingartner, H., Kaye, W., Smallberg, S. A., Ebert, M. H., Gillin, J. C., & Sitaram, N. (1981). Memory failures in progressive idiopathic dementia. *Journal of Abnormal Psychology, 90,* 187–196.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency and DSM III. *Archives of General Psychiatry, 41,* 1005–1012.

(*Appendix follows on next page*)

Appendix

Modified Mini-Mental State (3MS)

Reprinted from "The Modified Mini-Mental State (3MS) Examination" by E. L. Teng, 1987, *Journal of Clinical Psychiatry, 48,* pp. 314–318. The 3MS administration and scoring manual may be obtained from E. L. Teng, Department of Neurology, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033.

3MS	MMSE	
5		DATE AND PLACE OF BIRTH Date: year ____, month __, day ___, Place: town ____, state __ (1 pt each)
3	3	REGISTRATION Shirt, Brown, Honesty (1 pt each)
7	5	MENTAL REVERSAL 5 TO 1 (0–2 pts) DLROW (0–5 pts)
9	3	FIRST RECALL Spontaneous recall (3 pts) Cued: "Something to wear" (2 pts) Recog: "Shoes, Shirt, Socks" (1 pt) Spontaneous recall (3 pts) Cued: "A color" (2 pts) Recog: "Blue, Black, Brown" (1 pt) Spontaneous recall (3 pts) Cued: "A good personal quality" (2 pts) Recog: "Honesty, Charity, Modesty" (1 pt)
15	5	TEMPORAL ORIENTATION Year (accurate = 8 pts; within 1 year = 4 pts; within 2–5 years = 2 pts) Season (accurate or within 1 month = 1 pt) Month (accurate = 2 pts; missed by 1 month = 1 pt) Day of Month (accurate = 3 pts; within 2 days = 2 pts; within 2–5 days = 1 pt) Day of Week (accurate = 1 pt)
5	5	SPATIAL ORIENTATION State (2 pts) County (1 pt) City (1 pt) Hospital/Building/Home (1 pt)
5	2	NAME (MMSE: Pencil __, Watch ___) Forehead __, Chin __, Shoulder __ Elbow __, Knuckle __
10		FOUR-LEGGED ANIMALS IN 30 SEC (1 pt each)
6		SIMILARITIES Arm-Leg (body part = 2; less correct = 1, 0) Laughing-Crying (feeling, emotions = 2; less correct = 1, 0) Eating-Sleeping (essential for life = 2; less correct = 1, 0)
5	1	REPETITION "I would like to go home/out" (2 pts) "NO IFS, ANDS, OR BUTS" (0–3 pts)
3	1	READ AND OBEY "CLOSE YOUR EYES" (no prompts = 3 pts; prompt = 2 pts; read only = 1 pt; incorrect = 0)
5	1	WRITING (1 minute) "(I) WOULD LIKE TO GO HOME/OUT" (0–5)

3MS	MMSE	
___10	___1	COPYING TWO PENTAGONS
		5 equal sides: Each 4, Pentagon 4 5 unequal sides: Each 3, Pentagon 3 other enclosed figures: Each 2, Pentagon 2 2 or more liness: Each 1, Pentagon 1 Intersection 4 corners: 2 not 4-corner enclosure: 1
___3	___3	TAKE THIS PAPER WITH YOUR L/R (non-preferred) HAND, (1 pt) FOLD IT IN HALF, AND (1 pt) HAND IT BACK TO ME (1 pt)
___9		SECOND RECALL Spontaneous recall (3 pts), Cued: "Something to wear" (2 pts) Recog: "Shoes, Shirt, Socks" (1 pt) Spontaneous recall (3 pts), Cued: "A color " (2 pts) Recog: "Blue, Black, Brown" (1 pt) Spontaneous recall (3 pts), Cued: "A good personal quality" (2 pts) Recog: "Honesty, Charity, Modesty" (1 pt)
___100	___30	

Received August 22, 1994
Revision received August 16, 1995
Accepted August 21, 1995 ■