ANALYSIS OF STRUCTURE AND DISCRIMINATIVE POWER OF THE MATTIS DEMENTIA RATING SCALE

SCOTT M. HOFER AND ANDREA M. PICCININ

University of Southern California

DOUGLAS HERSHEY

George Mason University

The study examines the factor structure and provides test of the discriminative properties of the 38-item Mattis Dementia Rating Scale (MDRS). The MDRS was designed a priori to measure five broad domains of cognitive abilities: attention, initiation/perseveration, conceptualization, construction, and memory. Complete item level data were collected at the USC Alzheimer Disease Research Center from 19 probable Alzheimer's Disease (AD) patients, 17 cases with dementia of various etiologies (e.g., multiple infarct), and 49 contrast subjects. Factor analyses, with rotation to equamax criterion, were performed on education partialled data. Five and six factor solutions accounted for most of the reliable variance and permitted simple structure theoretical description for separate subscales. These factors, similar to Mattis' design, can be characterized as Memory (Recall)/Verbal Fluency, Construction, Memory (short-term), Initiation/Perseveration, and Simple Commands. Cross-validated discriminant analyses performed on five unit-weighted composite variables derived from factor analysis provided better classification (72% vs 62%) than the 38 Mattis items alone. © 1996 John Wiley & Sons, Inc.

To assess the cognitive status of individuals who perform at the lower ranges of intellectual functioning, such as in Alzheimer's Disease or multiple infarct dementia, measures must be sensitive to variations in very low levels of cognitive functioning. Although little is known regarding the nature of these low level abilities nor the precise processes involved in their decline in dementia, what is known and thought important in distinguishing these individuals from normal adults has been summarized in mental status exams and dementia screening devices; measures constructed to discriminate dementia from normal functioning.

In assessing any level of cognitive functioning, however, mental status exams provide only cursory, initial information—consideration of the subtest scores as well as further examination are necessary to augment basic findings and increase confidence in the diagnosis. It would be advantageous to have a measure that not only provided valid cutoff score information but subscale measures of distinct ability to allow greater specificity of diagnosis. The psychometric properties of such dementia scales have not, however, been well documented. The focus of this paper is to examine these properties in a well known and widely used assessment device.

The 38-item Mattis Dementia Rating Scale (MDRS, Mattis, 1976) is a popular screening device used in the initial diagnosis of dementia and in the assessment of low levels of cognitive functioning. There are several reasons for its popularity: the MDRS derives mainly from several other well-known devices used in clinical work or research (see Vitaliano, Breen, Russo, Albert, Vitiello, & Prinz, 1984 for a list of studies which used specific items), it

---

The authors wish to thank John L. Horn for his insightful comments and suggestions on this manuscript.
Correspondence should be addressed to: Scott M. Hofer, Age and Cognitive Performance Research Centre, University of Manchester, Oxford Road, Manchester, England M13 9PL.

Journal of Clinical Psychology, Vol. 52(4) 395–409 (1996)

CCC 0021-9762/96/040395-15

measures aspects of cognition that researchers and clinicians regard as important for understanding dementia, it is easily administered without extensive training, and it was designed with a time-saving scoring procedure which decreases overall test time when assessing non-demented individuals.

Although the MDRS was developed as a dementia screening device, designed to allow the use of cut-off scores for clinical purposes, it is useful to study its psychometric properties. As a continuous cognitive measure, individual differences on the items can be analyzed in terms of correlations, and tests can be made of the underlying factors of the MDRS. Demonstration of a multi-factor structure would provide evidence for subscale construct validity, specifically of convergent and discriminant validity.

A test that distinguishes poorly between indicators of separate processes or factors, however, can still discriminate between groups of subjects such as demented versus normal. When a test is evaluated for its clinical utility, discrimination using total or subscale scores is of interest. Discrimination between populations of demented patients and normal adults would also provide evidence for construct validity, increasing confidence that the MDRS is measuring constructs for which it was designed. This paper will attempt to synthesize both perspectives and develop a practical bridge from which to regard this scale.

According to the theory implicit in its development, the MDRS measures five distinct features of cognitive functioning: attention, memory, conceptualization, initiation and perseveration, and construction. The items and the features they define are shown in Table 1.

Following Mattis' scoring procedures, the item subscores of a particular feature (i.e., memory) are added together to obtain a score for that feature. This indicates an implicit theory that each of the five collections of items measures a separate dimension of cognition. It also implies that each such dimension is reliable—in internal consistency and in stability—and that each subscale is reliably independent of the others. Evidence in support of these hypotheses would indicate important aspects of construct validity (Cronbach, 1989).

The theory specifying functional unities (see Cattell, 1957) for the features implied in the MDRS is not clear—particularly when judged from the perspective of what is known about the organization of human cognitive capacities in normal adults (e.g., Ekstrom, French, & Harman, 1979; Hakstian & Cattell, 1974; Horn, 1972, 1985, 1988; Horn & Hofer, 1992). Independent of the MDRS, there is a substantial body of evidence showing that distinct processes of short-term, intermediate term, and long-term memory (Craik, 1977; Horn, 1988; Horn & Hofer, 1992), and several features of attention (McDowd & Birren, 1990) exist in intact adults. The evidence is also clear in demonstrating that these processes of memory and attention are functionally distinct from processes of reasoning and storage of knowledge and information (Horn & Hofer, 1992). There are, however, no studies relating the evidence on normal functioning to the qualities measured in very low range cognitive assessment scales, such as the MDRS. To establish such studies on a solid foundation it is necessary to construct information about the factors from measures with good convergent and divergent construct validity.

Vitaliano et al. (1984) provide evidence often interpreted as indicating that the MDRS subscales measure the constructs they are intended to measure. Using a sample of mildly (n = 18) and moderately (n = 16) demented individuals and a healthy control group (n = 23), they found acceptable coefficient alphas (Cronbach, 1951) for Attention (.95), Initiation and Perseveration (.87), Conceptualization (.95), and Memory (.75).

The alpha coefficient alone, however, is not a sufficient indicator of subscale construct validity. Needed, also, is evidence of reliable independence (Horn, Wanberg, & Foster, 1987) or divergent validity. Under basement or ceiling effects, where the scale is not sensitive, not only do intercorrelations for the items within the subscales increase but those across scales increase as well. This would be seen with the MDRS in groups including both severely demented and cognitively intact adults. Since alpha reliability is based on overall item correlations, regardless of how these are produced, the reliability of the measure can increase as the degree of dementia within the sample increases or decreases. Under these conditions,

Table 1
Items and Subscales of the Mattis Dementia Rating Scale

	Score	Dependent
Attention		
1. Digit span forward	max = 8	
2. Follow simple verbal command	max = 2	
3. Follow very simple verbal commands	4	*
4. Imitate simple movement	4	*
29. Counting A's in matrix of letters	6	
30. Counting A's in scrambled letter pattern	5	
33. Read list of five words 4×	4	*
35. Read list of five words 4× (diff than 33)	4	*
37. Complex design matching (set A)	4	*
Initiation and Perseveration		
5. Naming supermarket items	20	
6. Naming articles of clothing	8	
7. Say "Bee, Key, Gee" 4×	1	
8. Say "Bee, Ba, Bo" 4×	1	
9. Palm alternations	1	
10. Palm/open hand alternations	1	*
11. Tap left, tap right alternations	1	*
12. Copy simple line drawing	1	
13. Copy an "O"	1	*
14. Copy an "X"	1	*
15. Copy a string of "O" and "X"	1	*
Construction		
16. Copy diamond inside square	max=1	
17. Copy diamond next to square	max=1	*
18. Copy a diamond	1	*
19. Copy a square	1	*
20. Copy four short vertical lines	1	*
21. Write your name	1	*
Conceptualization		
22. Similarities: Apple/Orange, how same?	8	
23. Naming of things people eat, etc.	3	*
24. List of three things, which different	3	*
25. Similarities: mult. choice, which alike	8	*
26. Which geometric form is different	16	*
27. Make up and verbalize a sentence	1	
Memory		
28. Orientation items: Day, month, city . . .	9	
31. Recall an earlier presented sentence	4	
32. Recall the sentence made up earlier	3	*
34. Recognition of 5 words from #33	5	
36. Recall of 5 words from #35	5	*
38. Complex design matching (set B)	4	

Note.—Asterisks indicate that the item would be scored correct without test if the most difficult item within the group was scored correct.

however, the intercorrelations among the different scales can also increase to the point where correlations between different scales are practically as large as the alpha correlation. This is a sign that although the separate scales are internally consistent, they do not reliably measure different attributes—the subscales do not possess discriminant validity. Because mental status exams are sensitive to a limited range of ability, they can change from measuring several distinct abilities to measuring a single undifferentiated ability, depending upon the sample tested.

This problem is particularly interesting to study with the methods of factor analysis. The factor structure may not remain invariant across extreme groups, such as severely demented versus normal subjects. In fact, the factor structure can be expected to change with the progression of a dementing illness. For example, if a sample consisted of mildly to severely demented patients, the subgroup of severe cases could be expected to have extremely low scores on all measures, indicating a single factor, while in the mildly demented group there might be evidence of five distinct factors. Moreover, each of these factors would probably have a lower alpha than those found (for the same items) in the severely demented group. This illustrates why alpha alone can be a deceiving indicator of construct validity.

Changing factor structure can be seen in studies of the Extended Scale for Dementia (ESD; Hersch, 1979), an expanded version of the MDRS. In moderately impaired samples, Hersch (1979) and Helmes and colleagues (1992) found evidence for unidimensionality. Appell et al. (1990) reported a large first and a second small factor in a less impaired hospital sample. Thus while the theory of the MDRS implies that separate factors are internally consistent, stable, reliably independent and construct valid, and that evidence of convergent and divergent validity should obtain among severely demented people, as well as normals, there is inadequate evidence to support these assertions and there are reasons related to difficulty of achieving discriminable measurement for why they may not be true.

There is substantial evidence from the study of normals to suggest that reliably independent measures of cognitive processes, similar to those which the MDRS was designed to measure, can be obtained. As well, there is some evidence that the MDRS, whether regarded as a one-factor or a five-factor measure, reliably discriminates between severely demented and normal people, and perhaps between people with different dementing conditions. There is little evidence, however, to indicate which processes measured in the MDRS provide the basis for such discriminability. The purpose of this research is to provide empirical evidence for each of these matters.

METHOD

Subjects

The data were gathered as part of a multidimensional longitudinal research project organized and conducted by the Alzheimer Disease Research Center (ADRC) of the University of Southern California. The sample consisted of 85 individuals with complete data:

1. $n = 19$ individuals diagnosed as having probable Alzheimer's Disease (AD)—based on NINCDS-ADRDA criteria (McKahnn, Drachman, Folstein, Katzman, Price, & Stadlan, 1984). The AD group was composed of 7 men and 12 women between the ages of 54.4 and 87.0 years (*M* age = 76.2 years, *SD* = 7.44). Their mean level of education was 11.74 (*SD* = 3.06) years.
2. $n = 17$ individuals with diagnosis indicating mixed dementia, in some cases due to multiple infarcts. The mixed dementia group was composed of 8 men and 9 women with a range in age from 59.9 to 93.8 years (*M* age = 75.05 years, *SD* = 9.30 years). Mean education level was 14.65 years (*SD* = 2.96).
3. $n = 49$ ostensibly healthy individuals, not known to be experiencing any form of cognitive impairment. Individuals were admitted to this group on the basis of similarity in age and SES to individuals in groups (1) and (2). The contrast group was composed of 21 men and 28 women with ages ranging from 38.7 to 89.1 (*M* age = 71.05, *SD* = 9.52). Mean education was 14.33 years (*SD* = 2.73).[1]

[1]Groups formed in this way are often said to be "matched control groups." This is misleading. Such groups are not known to "control" for any relevant variance. Certainly they do not control for the most relevant variance—all influences other than the ones of interest that could produce the hypothesized relationships—and, rather than control, can introduce spurious variance (Baltes, 1973; Campbell & Fiske, 1959; Edwards, 1968; Fisher, 1946; Horn, 1979; Horn & Donaldson, 1976; Humphreys, 1978; Reichardt, 1979). The widespread practice of forming groups that are comparable in terms of age, SES and gender can be justified in design on the grounds that they provide blocks of subjects that can be contrasted in terms of many variables other than age, SES and gender in respect to which groups can differ.

In order to assess differences between groups, an ANOVA was performed on the variables age, education, and gender. AD subjects had significantly fewer years of education than the mixed dementia and the contrast group, $F(2.82) = 6.58, p < .05$. Differences in age and gender were not significant.

A significant difference in education between demented and other groups is a common finding among studies of dementia scales. This raises questions about the extent to which the MDRS reflects differences in education rather than in level of dementia. While Jorm, Scott, Henderson, and Kay (1988) found no evidence for test bias against the poorly educated on a scale highly similar to the MDRS, the Mini Mental Status Examination (Folstein, Folstein, & McHugh, 1975), other researchers call for adjusting the dementia ratings scale scores for level of education (e.g., Kittner, et al., 1986).

In the present study, analyses were performed with educational differences partialled from the test scores to eliminate the group differences. The partialling of education variance could attenuate the effects of individual differences (such as dementia or development) and reduce chances of finding distinct factors (Horn & Cattell, 1982). For example, partialling age would produce this effect because age variance is associated with declines in some abilities. Similarly, level of education can have an effect on the measurement of dementia. Better educated individuals might not show decline on tests of mental ability as soon as others with less education, and the dementia measures might be less sensitive to the decline of higher level knowledge or learned skills. It was thought important, in this study, to be able to attribute group differences to dementia and not to a possibly confounding demographic variable such as education.

Procedure

The Mattis Dementia Rating Scale was administered as part of a battery of cognitive and neuropsychological tests during a two hour test period with a trained tester. Data collection took place in the Los Angeles area from 1986 to 1988. Demographic data were collected during a clinical interview prior to testing.

Scoring dependence: Problems for structural research. The MDRS test protocol produces dependence among the items, potentially creating problems for examining the internal consistency and independence of the separate scales. Groups of test items within a subscale are arranged hierarchically, the most difficult item administered first. In three of the five scales, if the subject passes the first item, it is assumed that he or she will pass the less difficult items of the scale. For example, if a subject passes item 2 in Table 1, the test instructions stipulate that the subject receive maximum credit on items 3 and 4. This scoring rule only holds for the first, most difficult item. Such scoring assumes there is perfect theory-specified (e.g., tetrachoric, biserial) correlation of item 2 with items 3 and 4. Such scoring increases the empirical correlations among items 2, 3, and 4 and tends to force a factor among these items and other items correlated with any of these three. By assumption, then, not by free test, such scoring "stacks the cards" in favor of finding a factor among the items of a scale. If no convergence results, it is strong evidence that there is no factor among the items. If the factor does not appear with the help of operational dependence, it will not appear.

Variables scored in a dependent manner are said to be operationally dependent (e.g., Nunnally, 1978). Operationally dependent items are indicated in Table 1 by an asterisk. The larger-numbered items within a scale are not presented to the subject and maximum credit is given if there is pass performance on the smallest-numbered item within the scale. In the MDRS, this operational dependence in scoring follows from the nature of the items. For example, the less difficult construction test items are essentially smaller parts of the most difficult construction item. If all items were presented to all subjects, these dependencies would still exist.

Extension Analysis (Horn, 1973) on the non-operationally dependent items was used to extend factor analytic results to these dependent items. The results were somewhat supportive of the complete item analyses; however, the factor loadings were more complex. Results for the complete score analyses only are reported here.

Rationale and overview of statistical procedures. Factor analysis of test items provides relevant information for evaluating the hypothesis that separate subscales measure distinct attributes. Evidence of distinct subscales is neither necessary nor sufficient to rule in, or rule out, all hypotheses relating to the (construct) validities of measures, but it is sufficient to establish sound arguments regarding convergence (internal consistency) among items of a scale and discriminant validity between scales. Evidence to support the hypothesis of convergence would be given by results showing that the items scored to represent a particular feature of the MDRS—the items of Attention, for example—come together in a factor obtained by objective methods (Horn, 1967; Horn & Knapp, 1973, 1974). Evidence to support the hypothesis of divergent subtest validity would be adduced by results showing that the items scored for a particular feature correlate primarily with only one factor and not with other factors.[2]

Discriminant analysis (DA) tests the clinical utility of the MDRS for discriminating between sample populations of Alzheimer patients, mixed dements, and normal subjects. Interpretation of the discriminant weights permits further understanding of which items or constructs best characterize each group by maximally discriminating between groups. This also provides evidence for construct validity of the scale.

Finally, of interest, is whether the information gained from the factor analyses can lead to a more reliable classification of subjects in a discriminant analysis. Factor analytic results were used to guide construction of composite variables as indicators of the factors. Selection was based on simple structure criteria—choosing those items which provide the clearest interpretation for a factor. A simple unit-weighted sum of the standardized item scores was used in the calculation of each composite. This method of estimating factor scores produces similar results as using weighted combinations of items (see Wackwitz & Horn, 1971). These analyses aim to more clearly define the constructs which provide the best discrimination between groups.

RESULTS

Discriminant Analysis

Heterogeneity of covariance matrices can affect tests of equality of group mean vectors (Dillon & Goldstein, 1984). Inspection of the distributions of scores across groups confirms that violation of the multivariate normality assumption is likely and may lead to bias in both estimated classification error rates and significance tests. Moreover, unequal sample sizes in each group can bias classification in favor of the larger group. In an effort to decrease this bias, prior probabilities for the groups were set to the sample proportions.

Two discriminant functions were calculated using all 38 education partialled items. The combined chi-square fit over both functions was (76) = 166.36, $p < .01$. The second discriminant was significant after removal of the first function, $X^2(37) = 66.96$, $p < .01$. The two functions account for 67% and 33%, respectively, of the between group variability. The first maximally separates AD and mixed dementia subjects from the contrast group. The second discriminant function separates the mixed dementia group from the AD with the contrast group falling in between.

The loading matrix of correlations between the predictor variables and the discriminant functions is shown in Table 2a. The variables that best distinguish the AD/mixed dementia

---

[2]Evidence in support of obtaining separate factor scores or sum-scores for each feature is given by demonstration that the internal consistency of each feature is substantially larger than the squared multiple correlation of the measure of that feature with the measures of all other features (Horn et al., 1987). Analyses are directed at providing this kind of evidence.

Table 2a
Structure Matrix of Pooled Within-Groups Correlations Between MDRS Items and Canonical Discriminant Functions

Item	Function 1	Function 2
5	.53*	.07
32	.47*	.09
31	.44*	.11
24	.36*	.14
17	.36*	.05
36	.35*	.19
12	.35*	−.14
18	.34*	.05
6	.34*	.22
28	.60*	.12
15	.33*	−.06
10	.33*	.03
22	.32*	.24
23	.31*	.13
34	.31*	.12
9	.28*	.17
38	.27*	.14
16	.27*	.04
19	.26*	.04
20	.25*	.04
26	.24*	.05
37	.23*	.09
8	.20*	.05
21	.20*	.16
14	.19*	.05
33	.16*	−.00
13	.16*	.08
30	.16*	.05
7	.15*	.08
35	.12*	.08
27	.11*	.02
1	.20	.27*
25	.16	.27*
29	.14	.16*
11	.07	.13*
2	.06	.13*
4	.08	.11*
3	.05	.11*

Note.—MDRS items are ordered by size of correlation with function. Primary correlations between items and discriminant functions are marked by asterisks.

groups from the contrast group are short-term apprehension and recall items (Horn, 1985). These items have the four largest correlations ($r > .44$) with the first discriminant function. The most highly correlated item measures very basic "awareness" of date, time, and place, while the second is a measure of word fluency, the ease with which one can name items belonging to a distinct set. The next two items measure elementary encoding and retrieval: recall a sentence read earlier and one that the subject made up. These items measure apprehension (awareness) and retrieval functions so basic to all other abilities that they rarely indicate individual differences. Only when abilities are severely affected, as in AD, do they indicate differences between such affected people and others. The more a measure indicates this elementary failing in cognitive capability, the better it discriminates along the first discriminant that maximizes the difference between AD and others. Table 2b contains the group centroids for both discriminant functions.

Table 2b
Canonical Discriminant Functions Evaluated at Group Means (Group Centroids)

Group	Function 1	Function 2
Alzheimer	−2.10	−2.02
Mixed	−2.37	2.11
Contrast	1.64	.05

The second discriminant function, which best distinguishes the multiple-condition brain damage individuals from the people with AD, correlates most highly with items 1 (digit span), 25 (multiple choice conceptual task), 29 (find and count As), 11 (finger tapping), and 2, 4, and 3 (follow simple commands), respectively. These items, too, require only very basic cognitive capabilities (which do not distinguish among normals). This suggests that the ability to follow very simple commands (as well as elementary short term memory) is to a greater extent spared in the brain damage of the multiple condition group—i.e., they are more like the contrast (normal) individuals in these abilities.

Classifications based on the discriminant functions for the sample of 85 subjects are presented in Table 3a. In this sample, 81 (95.3%) subjects were classified correctly. Three mixed dementia subjects and one AD subject were misclassified.

Although often reported, this classification rate is misleadingly high because it was calculated on the same sample of subjects on which the best-fitting discriminant functions were determined. Cross-validation in an independent sample is required to indicate how well the discrimination can be made when there is no capitalization on chance variation in determining the functions.

The small sample size in this study did not allow for split-half cross-validation of the discriminant function; therefore, the U-method (Dillon & Goldstein, 1984; as implemented in BMDP; Dixon, 1988) was used to reduce classification bias and increase test coefficient stability. In this procedure a discriminant function is computed with one observation omitted from the sample and the left-out case classified based on the resulting discriminant function. This process is repeated until all cases have been classified. This method provides increased confidence in the results based on small samples and yields nearly unbiased estimates of misclassification probabilities.

Under U-method cross-validation, classification rates were reduced, to a more reliable rate of 62.4%. Most of the errors in classification occurred between the AD and mixed dementia groups, as can be seen in Table 3b. For comparison purposes, classification tables are reported using both the cross-validation and normal classification procedures.

Factor Analysis

A scree plot (Cattell, 1966; Horn & Engstrom, 1979) was used to indicate the number of factors to be retained. The scree based on the eigenvalues from the residualized correla-

Table 3a
Classification Table for 38 MDRS Items Using Prior Group Probabilities

		Number of Cases Classified Into Group		
Group	Percent Correct	Alzheimer	Mixed	Contrast
Alzheimer	94.7	18	0	1
Mixed	82.4	1	14	2
Contrast	100.0	0	0	49
Total	95.3	19	14	52

Table 3b
Cross-Validation Classification Table for 38 MDRS Items Using Prior Group Probabilities

Group	Percent Correct	Number of Cases Classified Into Group		
		Alzheimer	Mixed	Contrast
Alzheimer	21.1	4	7	8
Mixed	17.6	10	3	4
Contrast	93.9	3	0	46
Total	62.4	17	10	58

tion matrices (with ones in the diagonal) indicated eight factors. Analyses such as those of Horn and Engstrom (1979) suggest that this scree-indicated number is usually the upper boundary of the number of factors one can expect to find evidence for under cross-validation in the same kind of sample. Solutions with four to eight factors were therefore considered. Common factor (CF) analyses (estimating communalities with squared multiple correlations, SMC) followed by Equamax rotation were performed on the education residualized correlation matrix. Five factors provided the clearest, most parsimonious simple structure. Factor loadings for the complete set of items are presented in Table 4.

Based on the five factor solution, the factors can be described as follows:

Long-term memory (verbal recall)/Verbal Fluency. A similar factor has been found in studies of normal adults (Horn, 1972, 1986; Horn & Hofer, 1992), where lower processes of association were seen together with processes of recall and recognition. In the present factor, in contrast, the association processes have been largely split off from the recall and retrieval processes: all the variables salient in the factor require either recall from store of knowledge or recognition of stored knowledge. (Where storage, as in previous work, can be consolidation measured as little as five minutes after presentation). Verbal fluency can be thought of as a type of recall from a long term storage, more specifically an ability to access such a store.

Construction. This dimension aligned closely with Mattis' measure of construction. It mainly indicates the ability to physically reconstruct visual images. In addition, both design matching items load on this factor (one with minor loading), indicating visual pattern recognition. The construction factor was robust, with similar configurations across analyses with different numbers of factors extracted.

Memory SAR (short-term apprehension retrieval)/TSR (broad recognition). An aspect of this factor has been identified in several studies of non-demented adults (Horn, 1972, 1986; Horn & Hofer, 1992). It represents the processes of first becoming fully enough aware of stimuli to be able to use them in thinking or to reproduce them within less than two minutes after presentation. Digit span is the best exemplar of this aspect of this factor. In combination with the elementary apprehension of short-term memory and the recall from long-term storage (verbal recall), this factor indicates processes of recognition of nonverbal stimuli. The word recognition of this factor could be primarily recognition of the shape of the word rather than comprehension of its meaning. This primary process is shown in such diverse tasks as counting A's, recognizing previously presented words, and creating sentences.

Initiation/Perseveration. This factor is similar to Mattis' measure of initiation. It is composed of items such as continuous design copying, palm alternation, and oral repetition of syllables. A design matching item loads highly on this factor as well as on the construction factor.

Simple Commands/Attention. This factor indicates basic abilities demanding attention, such as following simple verbal commands, imitating simple motor movements, and identifying simple concept similarities.

Table 4
Factor Pattern Matrix Based on Five Factor Equamax Rotation of Education Partialled Data

		Rotated Factor Pattern				
Description	Item	F1	F2	F3	F4	F5
Memory/Verbal Fluency						
Similarities	22	74				
Supermarket	5	74			36	
Recall Sent.	31	67	39			
Orientation	28	67	41		42	
Different items	24	66	33	38		
Word Recognition	34	61		50		37
Recall Sent.	32	60	54			
Naming (eat, ride)	23	58				
Naming (clothes)	6	55	37	45		
Word Recall	36	55		31	50	
Construction/Copy						
Diamond/Square	17	32	85			
Diamond/Square	16		79			
Diamond	18		72	37		
Square	19		62		30	
Four Lines	20		62	35		
Design Matching	37		54		47	
Memory (Short term)						
Read Word List	35			81		
Read Word List	33		35	70		
Digit Span	1			64	38	
Write Name	21		45	57		
Count A's (reg)	29			54	30	36
Count A's (mix)	30			53	44	36
Copy "O"	13		36	52		40
Copy "X"	14	34		50		40
Makeup Sentence	27			49		
Say "Bee, Key, Gee"	7			44		
Initiation/Perseveration						
Copy "XOXOXO"	15		35		70	
Palm Alteration	9		47		62	
Design Match.	38	35	32		62	
Palm/Open Hand	10	34			62	
Finger Tapping	11				55	42
Geometric Differences	26	38		39	54	32
Simp. Line Draw	12	42	36		47	−34
Simple Commands						
Imitate Movement	4					73
Verbal Commands	2					61
Verbal Commands	3					47
Simil. (mult choice)	25				45	45
Say "Bee, Ba, Bo"	8					−48

Note.—Factor loadings under .30 and decimal points were omitted from table.

The six factor extraction produced similar results, except that a factor composed of simple copying of X's and O's (items 13 and 14) and counting As within a regular pattern of letters became differentiated from the memory factor (factor three) of the five factor solution.

Discriminant Analysis Based on "Factor Score" Composites

After Z-score transformation of the items, unit-weighted composite variables were created based on the results of the five-factor 38-item orthogonal solution. All 38 items were used to form five composite variables, each computed by summing the Z-transformed item scores for those items which loaded high (>.30) on a factor.

Discriminant analysis of the five composite variables resulted in a single significant discriminant function (DF) with X^2 (10) = 81.33, $p < .01$. The discriminant function accounted for 94% of the between group variability. Classification based on U-method cross-validation was 72%, shown in Table 5. Without cross-validation, classification was 75.3%. Contrast group subjects were always correctly classified, with errors occurring when classifying the AD patients (about 50% correct) and mixed dementia group subjects (about 25% correct).

The matrix of correlations between predictor variables and discriminant functions is provided in Table 6a. Inspection of these correlations suggests that the primary constructs associated with the first DF are Memory/Verbal Fluency, Initiation/Perseveration, and Construction. Coupled with the information about group centroids for the composite variables, in Table 6b, this information provides some understanding of the abilities on which the groups differ. Memory/Verbal Fluency, Initiation/Perseveration, and Construction best distinguish AD subjects from the contrast subjects, the mixed dementia group falling between. The second discriminant function is correlated with the attention/simple commands factor. This second function maximally separates the mixed dementia group from both AD and contrast subjects.

Group means for each of the constructs are presented in Table 7. Contrast subjects have higher scores on long-term memory (recall) tasks than do mixed dementia individuals who in turn perform at a higher level than AD patients. Factor three (broad recognition) also contributes to the distinction between groups on the first DF. Together, these two factors indicate that the brain-damaged individuals are distinguished from the contrast subjects primarily in respect to very basic, simple processes of being able to recognize and recall information previously learned or assumed to have been learned in the testing.

Discussion

Measurement Issues

MDRS item scores were generally not normally distributed. Dementia scales, designed to measure at the low end of cognitive functioning, can not be expected to have normal distributions in a sample taken from both clinical and normal adult populations. On these scales, normal adults show ceiling effects while those with dementia perform at the lower end, sometimes at floor level, producing U-shaped distributions. The groups exhibit mean differences which, if the scale had a greater range, might allow normally distributed scores within groups. In this combined sample of subjects, the item distributions ranged from near normal to sawtooth distributions to U-shaped distributions, with AD and normal contrast subjects having oppositely skewed distributions. As is the case here, non-normality is usually directly related to small and unequal sample sizes.

On the basis of item distributions, the MDRS seems to function well when used as a tool to initially differentiate dementia patients from a normally functioning population (i.e., gross group discrimination). Item or test distributions which exhibit skews on both ends can

Table 5
Cross-Validation Classification Matrix from Analysis of Five Composite Constructs Using 38 MDRS Items

		Classified Into Group		
Group	Percent Correct	Alzheimer	Mixed	Contrast
Alzheimer	42.1	8	7	4
Mixed	23.5	6	4	7
Contrast	100.0	0	0	49
Total	71.8	14	11	60

Table 6a
Structure Matrix of Pooled Within-Groups Correlations Between MDRS Items and Canonical Discriminant Functions Using All Items to Create Factors Based on Five-Factor Equamax Solution

Factor Score Composite	Function 1	Function 2
Memory/Verbal Fluency	0.91*	0.12
Initiation/Perseveration	0.59*	−0.23
Construction	0.59*	−0.24
Short-Term Memory	0.38*	−0.28
Simple Commands	0.34	0.71*

Note.—Primary correlations between factor score composites and discriminant functions are marked by asterisks.

allow good discriminability between groups. A rectangular distribution would allow for the best discrimination among subjects, each item providing incremental distributional properties. Rectangular distributions might be attainable if the items were tested in the sample for which the measure was designed.

Factor Analyses

The stability of several of the factors is worth note. The Construction, Initiation/Perseveration, and Attention/Simple Commands factors were structurally invariant across the number of factors extracted and across the partialling of education and group variables.

Interpreting the memory, verbal fluency, and pattern recognition (count As) factors is more difficult. Several of the items shifted under different extractions and partialling, decreasing confidence in the interpretation of these factors.

Discriminant Analyses

Discriminant analysis provides evidence for the utility of the scale in distinguishing various types of dementia from normal elderly populations. Caution must be observed when interpreting the discriminant analyses in this study since the overall score of the MDRS was used, along with a number of other neuropsychological tests, in the initial diagnosis (grouping) of the patients. In other words, the individuals in this study were grouped, in part, from the information gained from this rating scale. Despite this circularity, something could be learned about the types of abilities which best differentiate these three groups, especially between Alzheimer and mixed dementias.

The U-method cross-validation procedure increases confidence in the results and, in some cases, drastically lowers classification estimates. This method of classification decreases bias due to chance and should always be used when a small sample analysis is performed and an independent sample is not available. In this study, when all 38 items were included, there was a larger drop in the number of correct classifications than when the factor analytic-based composite variables were used. In the latter case, the difference between ordinary classification and cross-validation was much smaller, indicating less chance variation within the sample. Combining items to form scales increased the reliability therefore decreasing chance variation. One should use caution when considering these classifi-

Table 6b
Canonical Discriminant Functions Evaluated at Group Means (Group Centroids)

Group	Function 1	Function 2
Alzheimer	−1.72	−0.35
Mixed	−1.06	0.51
Contrast	1.03	−0.04

Table 7
Group Means and Standard Deviations on Composite Variables Derived from Factor Analysis

	Alzheimer	Mixed	Contrast
Means			
Construct			
Memory/Word Fluency	−8.80	−4.96	5.13
Construction	−3.71	−2.99	2.47
Short-term Memory	−4.99	−1.73	2.53
Initiation/Perseveration	−4.03	−3.20	2.67
Simple Commands	−2.27	.02	.88
Standard Deviations			
Construct			
Memory/Word Fluency	7.08	8.85	2.67
Construction	6.20	5.84	1.24
Short-Term Memory	12.57	6.21	1.35
Initiation/Perseveration	6.97	5.98	1.26
Simple Commands	4.82	1.88	1.71

cation results for clinical use since there is likely capitalization on chance due to the relatively small sample sizes.

Errors in classifying the mixed dementia subjects occurred with the greatest frequency, understandably so since it is a diverse group of individuals with compound diagnoses. As well, the range of cognitive decrement was much greater in this group, leading some to be classified as normals, others as AD. Differences in etiologies clearly make it difficult to treat these individuals as a single, distinct group.

The diversity of the mixed dementia group is likely the greatest reason for the low levels of correct classification in the discriminant analyses. Despite these low levels of classification, some information was gained as to how these individuals differed from normal elderly and Alzheimer patients.

These results must be tentative until a true cross-validation study is performed. Clinical decisions based upon reported results, interpretations and reworkings of MDRS subscales should be made with some caution. However, the use of results reported here may increase the reliability of assessment decisions and may help to clarify the underlying qualities of the subscales used in the measurement of decline.

Conclusions

The main goal of the study was to test hypotheses regarding the convergent and divergent construct validity in the MDRS. Factor analyses provided evidence for the separate subscales of the MDRS. In addition, evidence for construct validity was obtained through test of the MDRS' discriminating power. The U-method cross-validation discriminate analyses show that the MDRS does discriminate well in the sample tested, especially when coupled with the factor analytic results.

The factor analyses indicate an interpretable but complex structure underlying the MDRS. Some of the difficulty in interpretation may be due to the sample or to the scale itself. Further study using a larger sample, with greater sampling of mildly demented individuals would be necessary to test factorial invariance across groups and to increase confidence in these results. Overall, the psychometric properties of the MDRS at least approximate Mattis' design and supports its use as an initial screening device.

References

Appell, A., Kazarian, S.S., Mazmanian, D., & Ross, T. (June, 1990). *Factorial structure, internal consistency, convergent validity and discriminant validity of the Extended Dementia Scale*. Presented at the annual meeting of the Canadian Psychological Association, Ottawa.

BALTES, P.B. (1973). Life-span models of psychological aging: A white elephant? *Gerontologist, 13*, 458–492.

CAMPBELL, D.T., & FISKE, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

CATTELL, R.B. (1957). *Personality and motivation structure and measurement*. New York: Harcourt, Brace & World.

CATTELL, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.

CRAIK, F.I.M. (1977). Age differences in human memory. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 384–420). New York: Van Nostrand Reinhold.

CRONBACH, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

CRONBACH, L.J. (1989). Construct validation after thirty years. In R. Linn (Ed.), *Intelligence: Measurement, theory and public policy* (pp. 147–171). Urbana, IL: University of Illinois Press.

DILLON, W.R., & GOLDSTEIN, M. (1984). *Multivariate analysis: Methods and applications*. New York: John Wiley & Sons.

DIXON, W.J. (1988). *BMDP Statistical Software Manual*. Los Angeles: University of California Press.

ECKSTROM, R.B., FRENCH, J.W., & HARMAN, M.H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*. Fort Worth, Texas Christian University Press.

EDWARDS, A.L. (1968). *Experimental design in psychological research* (3rd ed.). New York: Holt, Rinehart & Winston.

FISHER, R.A. (1946). *Statistical methods for research workers* (10th ed.). Edinburgh: Oliver & Boyd.

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

HAKSTIAN, A.R., & CATTELL, R.B. (1974). The checking of primary ability structure on a broader basis of performances. *British Journal of Educational Psychology, 44*, 140–154.

HELMES, E., MERSKEY, H., HACHINSKI, V.C., & WANDS, K. (1992). An examination of psychometric properties of the extended scale for dementia in three different populations. *Alzheimer Disease and Associated Disorders, 6*(4), 236–246.

HERSCH, E.L. (1979). Development and application of the Extended Scale for Dementia. *Journal of the American Geriatric Society, 27*, 348–354.

HORN, J.L. (1967). Intelligence—Why it grows, why it declines. *Trans-Action, 5*, 23–31.

HORN, J.L. (1972). The structure of intellect: Primary abilities. In R.M. Dreger (Ed.), *Multivariate personality research* (pp. 451–511). Baton Rouge, LA: Claitor's Publishing Company.

HORN, J.L. (1973). On extension analysis and its relation to correlations between variables and factor scores. *Multivariate Behavioral Research, 8*, 477–489.

HORN, J.L. (1979). Trends in the measurement of intelligence. *Intelligence, 3*, 229–240.

HORN, J.L. (1985). Remodeling old models of intelligence: Gf-Gc theory. In B.B. Wolman (Ed.), *Handbook of intelligence* (pp. 267–300). New York: Wiley.

HORN, J.L. (1986). Intellectual ability concepts. In R.L. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3; pp. 35–77). Hillsdale, NJ: Lawrence Erlbaum.

HORN, J.L. (1988). Thinking about human abilities. In J.R. Nesselroade (Ed.), *Handbook of multivariate psychology* (pp. 645–685). New York: Academic Press.

HORN, J.L., & CATTELL, R.B. (1982). Whimsy and misunderstandings of Gf-Gc theory: A comment on Guilford. *Psychological Bulletin, 91*, 623–633.

HORN, J.L., & DONALDSON, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist, 31*, 701–719.

HORN, J.L., & ENGSTROM, R.O. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research, 14*, 283–300.

HORN, J.L., & HOFER, S.M. (1992). Major abilities and development in the adult period. In R. Sternberg & C. Berg (Eds.), *Intellectual development*. (pp. 44–99). Cambridge: Cambridge University Press.

HORN, J.L., & KNAPP, J.R. (1973). On the subjective character of the empirical base of Guilford's structure-of-intellect model. *Psychological Bulletin, 80*, 33–43.

HORN, J.L. & KNAPP, J.R. (1974). Thirty wrongs do not make a right: Reply to Guilford. *Psychological Bulletin, 81*, 502–504.

HORN, J.L., WANBERG, K., & FOSTER, M. (1987). *Guide to the Alcohol Use Inventory*. Minnetonka, MN: National Computer Systems.

HUMPHREYS, L.G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology, 70*, 873–876.

JORM, A.F., SCOTT, R., HENDERSON, A.S., & KAY, D.W. (1988). Educational level differences on the Mini-Mental State: The role of test bias. *Psychological Medicine, 18*, 727–731.

KITTNER, S.J., WHITE, L.R., FARMER, M.E., WOLZ, M., KAPLAN, E., MOES, E., BRODY, J.A., & FEINLEB, M. (1986). Methodological issues in screening for dementia: The problem of education adjustment. *Journal of Chronic Disease, 39*, 163–170.

MATTIS, S. (1976). Mental status examination for organic mental syndrome in the elderly patient. In L. Bellak & T.B. Karasu (Eds.), *Geriatric psychiatry* (pp. 77–121). New York: Grune & Stratton.

MCDOWD, J.M., & BIRREN, J.E. (1990). Aging and attentional processes. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging*. (pp. 222–233). New York: Academic Press.

MCKAHNN, 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.

NUNNALLY, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw Hill.

REICHARDT, C.S. (1979). The statistical analysis of data from nonequivalent group designs. In T.D. Cook & D.J. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings* (pp. 147–205). Chicago: Rand McNally.

VITALIANO, P.P., BREEN, A.R., RUSSO, J., ALBERT, M., VITIELLO, M.V., & PRINZ, P.N. (1984). The clinical utility of the dementia rating scale for assessing Alzheimer patients. *Journal of Chronic Diseases, 37*, 743–753.

WACKWITZ, J.H., & HORN, J.L. (1971). On obtaining the best estimates of factor scores within an ideal simple structure. *Multivariate Behavioral Research, 6*, 389–408.