Cognitive Dimensions in Alzheimer's Disease, Mild Cognitive Impairment, and Normal Elderly: Developing a Common Metric

Robert M. Chapman[1,*], Mark Mapstone[2], John W. McCrary[1], Margaret N. Gardner[1], Laura E. Bachus[1], Elizabeth DeGrush[1], Lindsey A. Reilly[1], Tiffany C. Sandoval[1] and Maria D. Guillily[1]

[1]*Department of Brain and Cognitive Sciences and Center for Visual Science at the University of Rochester, Rochester, NY 14627, USA*

[2]*Department of Neurology at the University of Rochester Medical Center, Rochester, NY 14642, USA*

Abstract: The aim of this research was to assess similarity in cognitive factor structures underlying neuropsychological test performance of elders belonging to three clinical groups: Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), and normal elderly. We administered a battery of neuropsychological tests to 214 elderly participants in the groups. First, the underlying cognitive structure of a Combined-Set of AD, MCI, and Control subjects was determined by Principal Components Analysis (PCA), including quantitative relationships (loadings) between the test measures and the factors. The PCA resolved 17 neuropsychological test measures into 6 interpretable factors, accounting for 78% of the variance. This cognitive structure was compared with separate cognitive structures from an AD-Set, an MCI-Set, and a Control-Set (different individuals in each set) in additional PCA using Procrustes factor rotation. Analysis of congruence coefficients between each set and the Combined-Set by a bootstrapping statistical procedure supported the factor invariance hypothesis. These close similarities across groups in their underlying neuropsychological dimensions support the use of a common metric system (the factor structure of a Combined-Set) for measuring neuropsychological factors in all these elderly individuals.

Keywords: Neuropsychological tests, principal components analysis (PCA), cognitive structures, cognitive dimensions, common metric, procrustes factor rotation, Alzheimer's disease, mild cognitive impairment, elderly controls, congruence coefficients.

INTRODUCTION

Alzheimer's disease (AD) is a neurological illness with early cognitive and behavioral disruption. The early cognitive deficits are frequently in the domain of memory, especially retentive memory. However, researchers and clinicians are now appreciating the behavioral heterogeneity of AD cognitive deficits and recognize that early and isolated deficits in domains of language, visuospatial abilities, executive function, and even mood may represent nascent AD [1]. Mild Cognitive Impairment (MCI) is a recently described diagnostic entity which may represent a transition state between normal aging and AD [2]. MCI has been defined as memory complaint with objective memory impairment in the context of normal general cognitive function and intact activities of daily living. As in AD, non-mnemonic subtypes of MCI are recognized, including those primarily affecting language [3], visuoperception [4], and executive abilities [5].

It is difficult to discern if different groups of patients employ the same cognitive processes while performing standardized neuropsychological tests. This is tied to understanding what areas of cognition are impaired in AD and MCI and how these impairments are revealed through neuropsychological testing. Additionally, while the diversity of cognitive deficits among individuals with AD and MCI can make clinical diagnosis difficult, a more fundamental challenge arises from the use of neuropsychological tests to assess single domains of cognitive function. Most standardized tests used by clinical neuropsychologists rely on multiple cognitive capacities for successful completion of each task, and the inferences derived from test performance should be tempered by an understanding of the component processes involved.

As a formal, data-driven method, factor analysis has been applied to neuropsychological tests (for a survey, [6-9]) to derive underlying neuropsychological dimensions. Our approach to better understanding the functional cognitive structure underlying neuropsychological test performance employs Principal Components Analysis (PCA), which reduces a correlation matrix of test measures to a few factors that are implicit in the data [10, 11]. (In this article we will use the term "factor" instead of "component" as they are nearly analogous). As Harman [12] pointed out, following the principle of parsimony that is common to all scientific theory, a law or model should be simpler than the data upon which it

*Address correspondence to this author at the Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA; Tel: (585) 275-8679; Fax: (585) 271-3043; E-mail: rmc@cvs.rochester.edu

is based. Thus, the number of factors should be less than the number of variables (test measures), and, in the linear description of each variable, the complexity should be low.

PCA condenses a wide array of measures that are based on different metrics and cognitive functions into simpler and interpretable factors. When derived from neuropsychological tests, a factor can be considered as a cognitive dimension (e.g., a general memory dimension). PCA provides both factor loadings (which relate test measures to the cognitive dimensions) and factor scores (which pertain to an individual's performance on those cognitive dimensions). The factor loadings describe properties of the system in terms of weights (correlations) of neuropsychological test measures on the underlying factors. In other words, the factor loadings represent the tests' varying contributions to each dimension while the factor scores represent an individual's performance on each dimension. Each factor can be identified by its pattern of these test measure loadings. Larger loadings (more distant from zero) are more salient in this identification and interpretation. Additionally, PCA achieves data reduction, which can be an important advantage in subsequent analyses where degrees of freedom limitations may be present.

PCA is an additive factor model, where the performance measure of an individual on a test is the summation of each factor's contribution to that test measure. The contribution is the product of the factor loading on that test measure (which is a static structure) and the factor score (which varies with the individual). As said above, a neuropsychological test may involve multiple cognitive capacities that are difficult to separate. Through PCA, this distinction may be quantitatively expressed as a test measure's loadings on the associated factors for those cognitive dimensions.

While there are obvious quantitative differences among AD, MCI, and normal individuals in neuropsychological test performance, an important question remains concerning qualitative differences among these groups. These differences refer to the underlying dimensionality of test performance including the relations among test variables. Understanding these relations may reveal how these test variables measure cognitive performance in AD, MCI, and normal elderly. In terms of factor analysis, factor invariance between two factor structures derived from different groups should be established before meaningful comparisons with factor scores can be made [13]. This issue of studying the underlying cognitive structure of AD, MCI, and normal cognition has been addressed by other researchers [8, 9, 14-16]. Siedlecki *et al.* [9] studied neuropsychological invariance using exploratory and confirmatory factor analyses and determined that, generally, there was structural (but not metric) invariance among AD, MCI, and normal elderly but with unanswered questions pertaining to how delayed memory might be different in AD versus MCI and normal elderly.

There are difficulties inherent in trying to compare factors across solutions generated from different subject groups. There is an arbitrariness to factor rotation that can result in two factor solutions occupying the same factor space with markedly different orientations, which can result in misleadingly small congruence coefficients [17]. One would not expect the same factors to appear in the same order in each solution. Additionally, selecting the number of factors to retain is inherently an arbitrary process (one in which we

combine the Eigenvalue > 1 rule with interpretability). Particularly for the factors whose Eigenvalues are clustered around 1, due to noise there will be fluctuations concerning which are retained and which are not. Thus, it is not likely the exact same set of factors will appear in each data set. The method we employ in this paper allows better comparisons of factor solutions than generating an independent, orthogonal solution from each subject group and attempting to determine if each contains the same set of factors. By rotating a replication factor structure to a target factor structure, we can directly measure how similar the replication structure is to the target by the congruence coefficients without the difficulties sometimes encountered when using confirmatory factor analysis [13, 18, 19]. When used in conjunction with statistical tests of fit, orthogonal Procrustes rotation leads to the acceptance of models that are replicable and the rejection of those that are not [18].

We will explore the invariance issue with a different approach that utilizes PCA and orthogonal Procrustes rotation to reveal factor similarities. We will test the invariance between the factor structures of our subject sets by statistically evaluating three types of congruence coefficients through a bootstrap procedure that randomly permutes the factor matrices [13, 18]. We administered a battery of neuropsychological tests to elderly participants and will compare the underlying cognitive structures of three elderly sets of subjects (an AD-Set, an MCI-Set, and a Control-Set) with that of a Combined-Set (comprised of different AD, MCI, and Control subjects). We will determine whether each set shares a similar factor structure and how strong this similarity is by using the Combined-Set's structure as a target and rotating the other sets to best match this target. The important result of such a comparison lies in the common metric derived from the Combined-Set. A common metric ensures that the same measurement axes can be used for all the groups involved in the metric's creation. If the factor structures of the AD-Set, MCI-Set, and Control-Set are equivalent to that of the Combined-Set (within the bounds of sampling error), then it is reasonable to build a single factor structure simultaneously from all of these groups. Then, the factor scores created by PCA for each individual can determine group membership by placing the test performance tied to those scores upon the axes provided by the common metric. This could be a formal way of assessing where a novel patient belongs along those axes, as part of the AD group, as part of a Control group, or some place in the middle where the MCI group may lie. Such a diagnostic method could possibly aid in early detection of both AD and MCI.

METHODS

Study Sample

This study involved four sets of elderly individuals (Table **1**): an AD-Set (containing 38 diagnosed with AD), an MCI-Set (containing 62 diagnosed with Mild Cognitive Impairment), a Control-Set (containing 63 with normal cognition), and a Combined-Set of 51 (containing 17 AD, 17 MCI, and 17 Control subjects, all different from those in the other sets). The subjects were randomly placed in either the Combined-Set or one of the other sets depending on their clinical diagnoses. The term "separate sets" will be used to refer to the AD-Set, MCI-Set, and Control-Set.

Table 1.   Demographic Information for Each of the Subject Sets

Group	Gender[a]	Age[b]	Education[b]	MMSE[c]	GDS[d]	BDS[e]
Separate Sets						
AD-Set (n=38)	17:21	75.9 (7.7)	14.7 (3.3)	24.7 (3.8)	6.7 (5.4)	1.2 (1.1)
MCI-Set (n=62)	27:35	72.1 (8.5)	14.8 (2.5)	26.4 (3.1)	6.6 (5.9)	1.2 (1.3)
Control-Set (n=63)	39:24	69.0 (10.0)	15.9 (2.6)	28.7 (1.5)	5.1 (5.2)	0.6 (1.0)
Combined-Set (n=51)						
AD (n=17)	7:10	76.0 (7.0)	13.5 (2.3)	23.9 (3.1)	7.4 (5.3)	1.5 (1.7)
MCI (n=17)	8:9	78.8 (5.7)	14.4 (2.9)	25.9 (3.0)	7.1 (5.5)	1.2 (1.5)
Control (n=17)	8:9	75.2 (5.7)	16.2 (2.9)	28.3 (1.4)	4.4 (3.3)	0.7 (0.9)

Note. The Combined-Set contained different individuals than those in the AD-Set, MCI-Set or Control-Set. Values appear as mean (SD). The neuropsychological tests shown here were administered as part of the experimental battery used in this study.
a) Females:Males.
b) The age and education are shown as number of years.
c) Mini-Mental State Examination [20]. The highest score possible is 30 (a high score means better cognitive performance).
d) Geriatric Depression Scale [21]. The highest score possible is 30 (a high score means more depressive symptoms).
e) Blessed Dementia Scale – CERAD version [22 , 23]. The highest score possible is 17 (a high score means a greater impact of dementia on daily activities).

The participants were recruited primarily from the Geriatric Neurology and Psychiatry Clinic at the University of Rochester and affiliated University of Rochester clinics. All AD and MCI subjects were evaluated by memory-disorders physicians. All AD subjects met standard criteria for AD (NINCDS-ADRDA) [24] and DSM-IV-TR criteria for Dementia of the Alzheimer's Type [25] and were considered early in the course of the disease. All MCI subjects met current consensus criteria for amnestic MCI [2, 26, 27]. Control subjects were also from area clinics, were spouses or friends of the AD or MCI subjects, or were volunteers from the community. No Control subject met criteria for AD or MCI. The clinical diagnosis of MCI and AD was based on a detailed patient history, relevant physical and neurological examinations and laboratory findings, and imaging studies routinely performed as part of the clinical assessment of dementia. Limited cognitive testing was performed by the memory-disorders physicians to assist with their diagnosis. With the exception of the Mini-Mental State Examination (MMSE) [20], a clock face drawing, and a category fluency task (animal naming), no cognitive test used in clinical decision making was repeated as part of our experimental test battery described below. Exclusion criteria for all groups included clinical (or imaging) evidence of stroke, Parkinson's disease, HIV/AIDS, and reversible dementias, as well as treatment with benzodiazepines, antipsychotic, or antiepileptic medications.

Demographic information for each of the subject sets is in Table **1**. The mean MMSE score of each subject group was appropriate for its diagnosis. For the MMSE, the ANOVA group effect was statistically significant within the Combined-Set (F (2, 48) = 11.6, p <0.001) and among the separate sets (F(2, 160) = 26.76, p<0.001). The mean score on the Blessed Dementia Scale (BDS) showed no significant group effect within the Combined-Set, but there was a significant group effect among the AD-Set, MCI-Set, and Control-Set (F(2, 160) = 5.61, p<0.01). There was no significant group effect for comorbid depressive symptoms measured by the Geriatric Depression Scale (GDS) within the Combined-Set or among the separate sets. In general, the mean scores for the GDS for each set were considered "normal" for depressive symptoms [28]. The number of years of education differed slightly, but significantly, among the groups within the Combined-Set (F(2, 48) = 4.5, p<0.05), as did the ages among the separate sets (F(2, 160) = 6.8, p<0.01). However, the influence of these demographic variables was removed by transforming each subject's scores to standard scores using published, corrected normative data, as described below. At the time of testing, 48 of the 55 AD subjects in the AD-Set and the Combined-Set (28 males, 20 females) and 44 of the 79 MCI subjects in the MCI-Set and the Combined-Set (23 males, 21 females) were taking cholinesterase inhibitors and/or memantine.

Our study received IRB approval from the University of Rochester Research Subjects Review Board, and informed consent was obtained from each subject.

Neuropsychological Assessment

The neuropsychological battery administered to each subject contained 15 common tests (Table **2**) that target different cognitive domains, particularly memory. We designed the battery to produce a comprehensive sample of cognitive processes and their degeneration in AD. Among others, the tests included measures of memory encoding, retrieval, and retention, generative fluency, executive function, and visuospatial abilities. All measures based on the amount of time the subject took to accomplish a task were inverted to become speed measures [29]. The effects of demographic variables, such as education and age, can make results difficult to interpret [6]. Therefore, the test measures of each participant were transformed to z-scores (mean 0, variance 1) using established age/education corrected normative data.

Neuropsychological Factors Derived from Principal Components Analysis (PCA)

The data used for PCA were 17 neuropsychological test measures (variables) obtained from 51 aged individuals (ob-

servations) in a Combined-Set of 17 ADs, 17 MCIs, and 17 Controls. Starting with the correlation matrix of the 17 measures, principal factors were extracted from the Combined-Set. The number of factors retained (6) was mainly determined by the Kaiser-Guttman rule [30] of the number of roots (Eigenvalues) greater than unity. The loadings of the 17 variables on each factor were measured. The FACTOR procedure (METHOD = PRINCIPAL) of SAS 9.1.3 [31] was used to conduct this analysis.

Measuring Factor Similarity Between Each of the Separate Sets and the Combined-Set

After developing the Combined-Set solution and determining the number of factors to retain, separate PCAs were made on the AD-Set, MCI-Set, and Control-Set in order to assess the similarity of their underlying multidimensional structures to that of the Combined-Set where these sets each contained a unique set of subjects. This separation of subjects was done to ensure new and different data would be used to measure the similarity of the Combined-Set factor solution to the separate sets. While performing the PCA on these smaller sets reduced the sample size used to create the target structure, it was important to test the reliability of the factor structure with entirely novel data.

Using the factor solution of the Combined-Set as the target structure, a new PCA was computed on the three separate sets and each rotated *via* the orthogonal Procrustean method [32] to best fit this target. In brief, this rotation method forces a replication structure to match a target structure as much as possible. As McCrae *et al.* [18] describe, "Factors are rotated to minimize the sums of squares of deviations from a target matrix, under the constraint of maintaining orthogonality. The technique realigns the position of the axes in the factor space without affecting the relative positions, just as multiple visual perspectives on a rigid object such as a table give different views without in the least changing the shape of the table."

After deriving these rotated factor solutions for the AD-Set, MCI-Set, and Control-Set, we measured their degrees of similarity to the target (the Combined-Set) through congruence coefficients. These are similar to correlation coefficients in that a coefficient of ±1 indicates perfect congruence while a coefficient of 0 indicates no congruence. However, unlike correlation coefficients, congruence coefficients are not adjusted for the means of the samples being compared [12]. Congruence coefficients are often used to measure invariance between a target factor structure and a replication structure (the data being rotated to match the target structure) [12, 13, 18]. We examined three types of congruence coefficients: variable congruence (which measures the extent of agreement in the pattern of loadings one variable has across all factors between two solutions and can be an indicator of the variable causing poor fit when there is a lack of factor congruence), factor congruence (which measures the extent of agreement in the pattern of loadings between corresponding factors in two solutions), and total congruence (an overall index of the degree of similarity) [13]. Thus, for each set there was a congruence coefficient for each variable (17), a congruence coefficient for each factor (6), and one total congruence coefficient. Congruence coefficients do not have known sampling distributions; therefore, bootstrapping is a

solution to determining statistical significance. Bootstrapping with replacement generates a distribution of congruence coefficients from randomly sampling the target and replication data and performing the PCA with Procrustes rotation. In other words, we randomized the subjects of Combined-Set and each of the separate sets (AD-Set, MCI-Set, and Control-Set) with replacement (which randomly allows subjects to potentially appear in the same group more than once), performed the PCAs, rotated each of the randomized separate sets to match the randomized Combined-Set, and measured the resultant congruence coefficients. This was done for every iteration of the bootstrapping technique. Performing this randomization and subsequent analysis numerous times creates a sampling distribution of congruence coefficients.

Two factor structures are judged to be equivalent (invariant) if their congruence coefficients are larger than critical values defined at a certain alpha level. For a more complete discussion of these concepts and the SAS IML code used to conduct this analysis, see Chan *et al.* [13] and Paunonen [17]. We performed 5,000 replications to generate critical values (measured from the bootstrapped sampling distribution) for each of the three randomized separate sets compared with the randomized Combined-Set (15,000 replications total). Our alpha level was 0.05.

Each congruence coefficient (24 total) for each comparison (AD-Set with Combined-Set, MCI-Set with Combined-Set, and Control-Set with Combined-Set) was judged to be statistically significant against the critical values generated from the bootstrapping analysis. A coefficient higher than its critical value supported invariance. These analyses were performed using the SAS IML procedure [31].

RESULTS

Neuropsychological Factors

From the PCA on the Combined-Set of three subject groups, we retained six factors. Four factors were retained by the Eigenvalue > 1 rule. An additional two factors had Eigenvalues nearly equal to 1 and were interpretable and thus were retained. Together these six factors accounted for 78% of the total variance. Table **2** shows the factor loadings with Varimax rotation [33], which was done for interpretability and simplicity. One factor was comprised of a single salient measure from a specific cognitive domain and required little interpretation (factor 6 in Table **2**), while all the remaining factors entailed two or more salient loadings (e.g., factor 2 in Table **2**). The factors encompass a wide range of cognitive skills, including general memory, generative fluency, visuospatial orientation, and speeded executive function. The factor solution (prior to Varimax rotation) became the target for comparisons with the AD-Set, the MCI-Set, and the Control-Set.

Cognitive Structures Compared for the AD, MCI, and Control Sets

In order to assess the similarities of the cognitive structures in the AD-Set, the MCI-Set, and the Control-Set, a PCA was done on each set separately and results of each set were compared with those of the Combined-Set. Because the Combined-Set PCA yielded six factors, a six factor solution was selected for each PCA performed. These turned out to

Table 2.   Factor Loadings for the 6-Factor Combined-Set Solution

Test Measure	Factors					
	1–General Memory	2–Speeded Executive Function	3–Learning/Object Naming	4–Generative Fluency	5–Visuo-construction	6–Visuospatial Orientation
Rey-Osterrieth Complex Figure [34, 35]						
Delayed Recall Score	.74	.30	.26	-.06	.06	.18
Mini-Mental State Examination [20]						
Score	.62	.19	.13	.28	.38	-.05
WMS-III Digit Span [36]						
Letter-Number Score	.27	.63	-.12	.46	.08	.08
WMS-R Logical Memory I [37, 38]						
Total Recall Score	.78	.38	.10	.13	.11	.22
WMS-R Logical Memory II [37, 38]						
Total Recall Score	.90	.20	.05	.01	.09	.09
Clock Face Drawing [39]						
Score	.40	.20	.06	.14	.79	-.04
Stroop Test [40]						
Color-Word Score	-.02	.68	.18	.07	.44	.20
Brief Visuospatial Memory Test-Revised [41]						
Total Recall Score	.30	.08	.81	.09	.20	.25
Discrimination Index	.79	-.07	.31	.26	.10	.02
Controlled Oral Word Association Test [42]						
Total Score	.15	.18	.11	.89	.12	.11
Category Fluency [42]						
Animal-Naming	.44	.63	.14	.36	.03	-.18
Hopkins Verbal Learning Test [43]						
Total Recall Score	.48	.36	.46	.29	.26	-.10
Discrimination Index	.69	.10	.37	.19	.29	-.12
Boston Naming Test [44, 45]						
Score	.17	.28	.73	.01	-.06	-.21
Standardized Road-Map Test of Direction [46]						
Score	.13	.14	-.01	.09	-.02	.92
Trail Making Test [47]						
A Speed†	.17	.66	.39	>.01	.19	.23
B Speed†	.33	.66	.44	.10	-.04	.03

Note. Red indicates values above an arbitrary threshold of 0.43 that was selected to highlight more salient loadings. This analysis was from the Combined-Set, which included 17 early-stage AD, 17 MCI, and 17 Controls. The table shows the PCA orthogonal Varimax [33] rotated factor pattern. Performing Varimax rotation produced a simple, more interpretable structure. The six factors explained 78% of the total variance.

† *z* scores for these test measures were generated from study data since published age/education corrected normative data were not available.

be similar as all solutions had at least five factors above the Eigenvalue = 1 threshold, and, in each case, the sixth factor had an Eigenvalue very close to 1 (in the Control-Set, this factor had a value greater than 1). The six factors accounted for nearly the same percent of variance in the AD-Set, the MCI-Set and the Control-Set (78%, 75%, and 72%, respectively). The Eigenvalues of pre-rotation Factor 6 were 0.96

for the AD-Set, 0.91 for the MCI-Set, and 1.05 for the Control-Set, as compared with 0.73 for the Combined-Set. These Eigenvalues are in the context of a trace of 17.

After Procrustes rotation, the similarities between the Combined-Set and the factor structures separately derived from the AD-Set, the MCI-Set, and the Control-Set were assessed by their congruence coefficients (Table **3**). Values

Table 3.   Congruence Coefficients Between the Procrustes-rotated Solution for Each Subject Set and the Combined-Set

Congruence Type		Subject Set		
		AD	**MCI**	**CON**
Variable Congruence	Rey Delayed Recall Score	.81 (.65)	.83 (.62)	.63 (.43)
	MMSE Score	.85 (.68)	.56 (.30)	.88 (.67)
	Digit Span Letter-Number Score	.88 (.68)	.75 (.59)	.44 (.27)
	LM I Total Recall Score	.94 (.81)	.94 (.69)	.88 (.64)
	LM II Total Recall Score	.97 (.83)	.97 (.79)	.95 (.77)
	Clock Face Drawing Score	.61 (.39)	.82 (.41)	.55 (.20)
	Stroop Color-Word Score	.89 (.66)	.84 (.70)	.85 (.67)
	BVMT-R Total Recall Score	.57 (.18)	.90 (.53)	.85 (.63)
	BVMT-R Discrimination Index	.50 (.25)	.65 (.16)	.52 (.15)
	COWAT Total Score	.86 (.61)	.82 (.65)	.89 (.73)
	Category Fluency Score	.75 (.49)	.84 (.64)	.73 (.50)
	HVLT Total Recall Score	.88 (.67)	.76 (.59)	.72 (.45)
	HVLT Discrimination Index	.92 (.70)	.86 (.73)	.60 (.37)
	Boston Naming Test Score	.85 (.68)	.17 (.05)	.83 (.63)
	Road-Map Score	.91 (.61)	.93 (.56)	.97 (.85)
	Trail Making A Speed	.75 (.53)	.78 (.65)	.88 (.71)
	Trail Making B Speed	.84 (.63)	.90 (.70)	.85 (.67)
Factor Congruence	Factor 1	.96 (.89)	.98 (.95)	.94 (.90)
	Factor 2	.87 (.62)	.83 (.43)	.66 (.46)
	Factor 3	.92 (.47)	.51 (.34)	.63 (.47)
	Factor 4	.50 (.32)	.78 (.29)	.85 (.41)
	Factor 5	.42 (.29)	.23 (.29)	.33 (.25)
	Factor 6	.67 (.22)	.61 (.18)	.64 (.19)
Total Congruence		.81 (.71)	.79 (.71)	.77 (.70)

Note. Each subject set column shows the congruence coefficients (perfect congruence = 1; no congruence = 0). After it the critical value at a one-tailed uncorrected alpha level of 0.05 corresponding to each congruence coefficient is displayed in parenthesis. These critical values were generated from 5,000 bootstrapping replications with replacement for each set.

in Table **3** were computed separately for each of the 17 variables (variable congruence) and for each of the six factors (factor congruence). An overall measure of congruence (total congruence) was also computed for each set (.81 for the AD-Set, .79 for the MCI-Set, and .77 for the Control-Set). For each set, these congruence coefficients were large enough such that, given the critical values derived from the bootstrapping procedure, factor invariance could not be rejected. This indicated the Combined-Set factor solution was approximated very well in the AD-Set, the MCI-Set, and the Control-Set solutions as graphically shown by the overlapping factor loadings patterns in Fig. (**1**).

The three types of congruence coefficients (variable congruence, factor congruence, and total congruence) total 24 for each separate set. All of the congruence coefficients were statistically larger than the critical values at alpha = 0.05.

Thus our findings support the conclusion of factor invariance.

DISCUSSION

Cognitive Structures of AD, MCI, and Normal Elderly: a Common Metric

We examined whether the factor structure derived from a Combined-Set of AD, MCI, and Control subjects is appropriate for each type of subject. This is of interest for both theoretical and practical reasons. Theoretically, invariance between the factor structures derived from separate groups allows comparisons of scores based on them to be meaningful. This is akin to being certain that the cognitive processes assessed by a neuropsychological test battery are the same
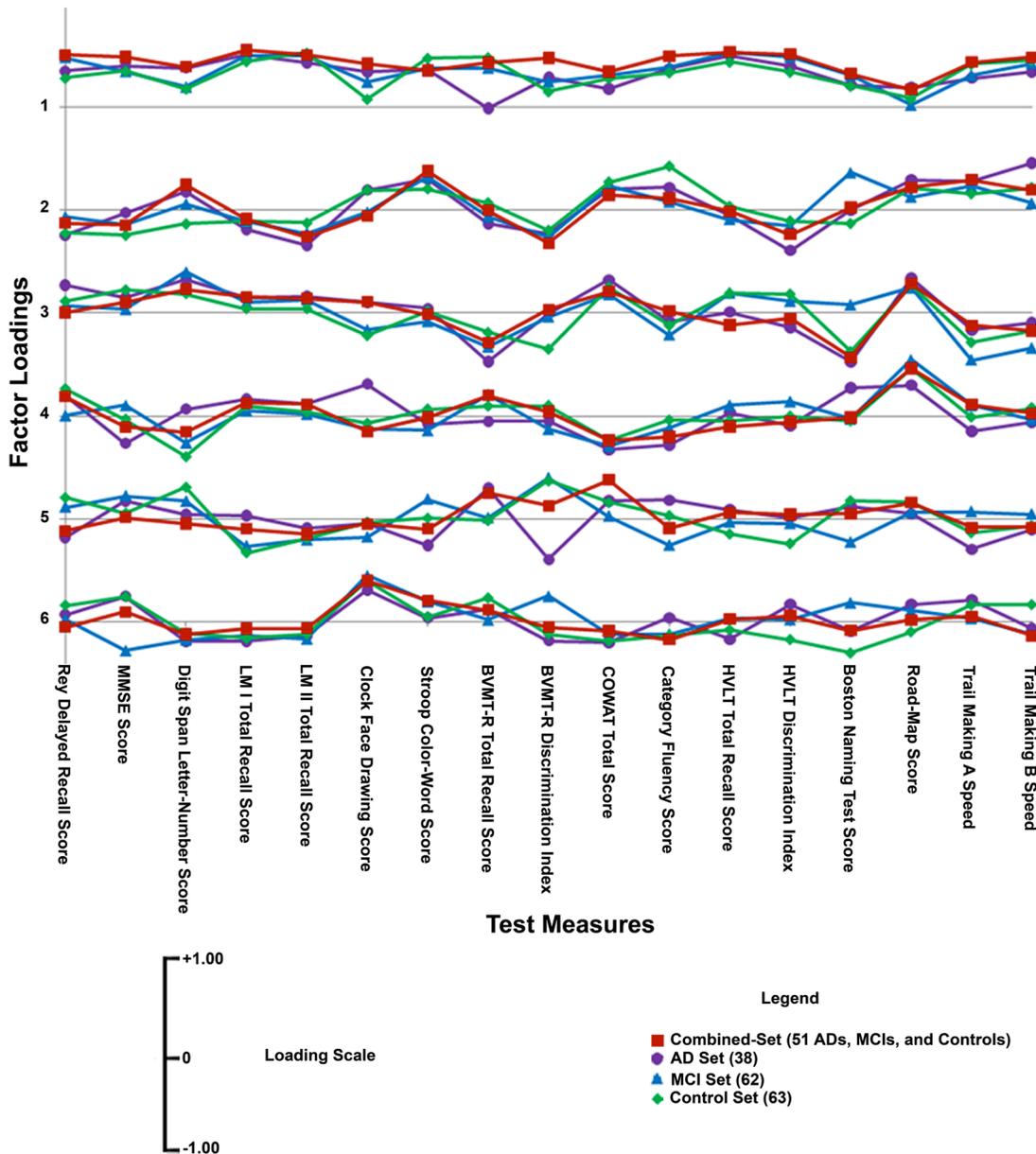
Fig. (1). Superimposed factor loadings for the Combined-Set, the AD-Set, the MCI-Set, and the Control-Set. The factors were generated from PCA; the Combined-Set had an orthogonal constraint applied, and the separate sets were rotated *via* orthogonal Procrustes rotation to best match the target (Combined-Set) structure. Loadings of the six factors are plotted across the 17 test measures (shown here in the order the tests were administered).

among different subject groups. On the practical level, it is advantageous for these groups to share the same underlying neuropsychological dimensions and thus justify the use of a common metric system (the factor solution) in measuring all the individuals. Thus knowing a novel individual's diagnosis is not necessary when applying the factor structure to obtain factor scores, and these factor scores can be easily used in subsequent statistical analyses for many purposes, including analyzing group differences, diagnosing individuals with AD [48], and predicting MCI progression to AD in individuals. These latter issues concerning analyses at the individual level are particularly important for early identification of patients requiring therapeutic and pharmacologic interventions. Additionally, this allows the clinician to avoid assum-

ing a single test score represents a single cognitive domain and instead permits remapping a patient's test battery to the fewer, more interpretable factor scores.

We used the factor solution of the Combined-Set as the target and computed separate PCAs on the AD-Set, MCI-Set, and the Control-Set with Procrustes rotation to best fit this target. The congruence coefficients between each set and the target (Table **3**) supported factor invariance. The majority of the 24 variables and 6 factors had large congruence coefficients. For all factors and variables in each two-set comparison, the congruence coefficients were larger than the critical values calculated from the bootstrapping analysis. The analyses supported the conclusion that the factor structures are congruent (invariant) (for a more in depth discussion of

this issue, see [13]). The overlaid pattern of factor loadings shown in Fig. (**1**) also visually revealed good agreement among the separate sets and the Combined-Set. Here the loadings of each factor in each analysis were plotted across the 17 test measures (the factors were grouped by their numerical order in each solution, as was done in Table **3**). Their superimposed patterns of loadings were extremely similar. Although it may appear that some factors (factor 5, in particular) showed low congruence coefficients and weaker visual agreement, this is due to sampling error rather than dissimilarity (as indicated by the congruence coefficient for factor 5 being larger than its critical value). The AD-Set was slightly more similar to the Combined-Set than the Control-Set was. The total congruence coefficient was slightly higher for the AD-Set (0.81) than for the MCI-Set (0.79) than for the Control-Set (0.77).

Although there may be concerns about PCA "over-factoring" the solution [49], the finding that even the sixth factor showed good agreement by congruence coefficients (all >0.60) that were all larger than their corresponding critical values confirms the original retention of six factors. There are multiple mathematical methods to measure latent constructs in a dataset, but we chose PCA because it operates with relatively few prior assumptions about the resultant structure and, though not used in this paper, easily generates factor scores. The choice of how to measure the latent constructs generally does not greatly affect the results [49] and sample size as a function of the number of variables is not an important factor for stability [50].

There is a technical advantage to using a variety of groups in the development of an underlying factor structure. Using data from only one group risks restricting the range in the test measures and attenuating correlations among variables. This can result in falsely low estimates of component loadings [51]. This risk is reduced by involving data from multiple groups of individuals.

However, we believe the stronger, practical advantage to be had in a common metric lies in the utility of the factor scores measured by it. A common metric ensures that the same measurement axes can be used for all the groups involved in the metric's creation. Sometimes separate factor structures have been derived for each group of subjects [9], but this leads to a problem of how to compare the resulting factor scores when the factor structures used to create them are different. When all the clinical groups of interest simultaneously contribute to the structure, it is a far easier task to use the factor scores derived from the factor structure to compare the neuropsychological performance of groups and even individuals [48]. The common metric demonstrated by the Combined-Set was influenced by both group and individual differences. There is no need to "translate" the scores of a Control individual into the factor structure of the AD group, for example. The factor scores of all individuals lie on the same dimensions and are thus easily compared and manipulated in subsequent analysis. Because our Combined-Set included AD, MCI, and Controls, the neuropsychological dimensions derived from the PCA performed on the Combined-Set can represent axes reflective of both demented and normal cognition. These dimensions might thus symbolize a gradient between neuropsychological performance in AD and normal elderly, and an individual's factor score on that

dimension can possibly lie at either end or some place between (as might be the case with MCI patients). For a greater discussion and demonstration of what can be gained from the implementation of a common factor metric, see [48].

These data support the conclusion that the AD, MCI, and Control groups likely share similar underlying cognitive dimensions. This does not mean that these groups would tend to have similar factor scores. It is expected that different groups of subjects would tend to obtain different factor scores on one or more of these factors. Here we are asking if the underlying neuropsychological dimensions are different, rather than if the locations of various groups on the dimensions are different.

Further research is needed to determine how stable the Combined-Set solution is. Given a larger number of subjects and more test measures, it is possible to refine the factor structure. Because the Combined-Set contained twice as many impaired individuals as normal individuals, its factor structure was slightly more reflective of cognitive processing in AD/MCI rather than normal elderly (as may be seen in our congruence coefficients). Additionally, it would be of great clinical and research interest to determine if these common factors underlie neuropsychological test performance in other types of dementia. If so, the empirically derived factor scores from the common metric might help differentiate between AD and other cognitive diseases.

CONCLUSIONS

1. Using neuropsychological test measures as input, principal components analysis (PCA) allows development of fewer, more interpretable underlying cognitive dimensions that provide both relations between the test measures and the dimensions (factor loadings comprising a factor structure) and measures of individual performance on those dimensions (factor scores).

2. The underlying factor structures of an AD-Set, an MCI-Set, and a Control-Set are congruent with the factor structure derived from a Combined-Set of those subject types (as judged through orthogonal Procrustes rotation and the resultant high congruence coefficients between the factors and variables of each separate set and a Combined-Set). This suggests the common metric of the Combined-Set adequately represents the cognitive dimensions of AD, MCI, and normal elderly.

3. A common metric system (the Combined-Set) is a useful and powerful measurement tool for gauging neuropsychological performance in AD, MCI, and normal elderly. This metric is made more stable by contributions from varied groups of individuals as the range in values is larger due to group differences as well as individual differences. Additionally, because the common metric reflects the influences of AD, MCI, and Control subjects, the factor scores derived from this metric may be used to diagnose patients with AD or predict progression to AD in MCI patients.

ABBREVIATIONS

AD   =   Alzheimer's Disease

BDS   =   Blessed Dementia Scale

BVMT-R	=	Brief Visuospatial Memory Test-Revised
COWAT	=	Controlled Oral Word Association Test
GDS	=	Geriatric Depression Scale
HVLT	=	Hopkins Verbal Learning Test
MCI	=	Mild Cognitive Impairment
MMSE	=	Mini-mental State Examination
LM I/LM II	=	Logical Memory I/Logical Memory II
PCA	=	Principal Components Analysis
SAS	=	Statistical Analysis System
WMS-R	=	Weschler Memory Scale-Revised

REFERENCES

[1] Weintraub S, Mesulam M. Four neuropsychological profiles in dementia. In: Boller F, Grafman J, Eds. Handbook of Neuropsychology: Elsevier 1993.

[2] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999; 56: 303-8.

[3] Nelson AP, O'Connor MG. Mild cognitive impairment: a neuropsychological perspective. CNS Spectr 2008; 13(1): 56-64.

[4] Mapstone M, Steffenella T, Duffy C. A visuospatial variant of Mild Cognitive Impairment: getting lost between aging and Alzheimer's disease. Neurology 2003; 60: 802-8.

[5] Traykov L, Raoux N, Latour F, et al. Executive functions deficit in mild cognitive impairment. Cogn Behav Neurol 2007; 20(4): 219-24.

[6] Carroll JB. Human cognitive abilities: a survey of factor-analytic studies. New York: Cambridge University Press 1993.

[7] Jones SN, Ayers CR. Psychometric properties and factor structure of an expanded CERAD neuropsychological battery in an elderly VA sample. Arch Clin Neuropsychol 2006; 21: 359-65.

[8] Loewenstein DA, Ownby RM, Schram L, Acevedo A, Rubert M, Argelles T. An evaluation of the NINCDS-ADRDA Neuropsychological Criteria for the assessment of Alzheimer's disease: a confirmatory factor analysis of single versus multi-factor models. J Clin Exp Neuropsychol 2001; 23(3): 274-84.

[9] Siedlecki KL, Honig LS, Stern Y. Exploring the structure of a neuropsychological battery across healthy elders and those with questionable dementia and Alzheimer's disease. Neuropsychology 2008; 22(3): 400-11.

[10] Chapman RM, McCrary JW. EP component identification and measurement by principal components analysis. Brain Cogn 1995; 27: 288-310.

[11] Chapman RM, Nowlis GH, McCrary JW, et al. Brain event-related potentials: Diagnosing early-stage Alzheimer's disease. Neurobiol Aging 2007; 28: 194-201.

[12] Harman HH. Modern factor analysis. Chicago: University of Chicago Press 1976.

[13] Chan W, Ho RM, Leung K, Chan DK-S, Yung Y-F. An alternative method for evaluating congruence coefficients with Procrustes

rotation: a bootstrap procedure. Psychol Methods 1999; 4(4): 378-402.

[14] Lekeu F, Magis D, Marique P, et al. The California verbal learning test and other standard clinical neuropsychological tests to predict conversion from mild memory impairment to dementia. J Clin Exp Neuropsychol 2009; i: 1-12.

[15] Tabert MH, Manly JJ, Liu X, et al. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. Arch Gen Psychiatry 2006; 23: 916-24.

[16] Tierney MC, Yao C, Kiss A, McDowell I. Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. Neurology 2005; 64: 1853-9.

[17] Paunonen SV. On chance and factor congruence following orthogonal Procrustes rotation. Educ Psychol Meas 1997; 57: 33-59.

[18] McCrae RR, Zonderman AB, Costa PT, Bond MH, Paunonen SV. Evaluating replicability of factors in the revised NEO personality inventory: confirmatory factor analysis versus procrustes rotation. J Pers Soc Psychol 1996; 70(3): 552-66.

[19] Raykov T, Little TD. A note on Procrustean rotation in exploratory factor analysis: a computer intensive approach to goodness-of-fit evaluation. Educ Psychol Meas 1999; 59(1): 47-57.

[20] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12(3): 189-98.

[21] Yesevage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1983; 17(1): 37-49.

[22] Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in cerebral grey matter of elderly subjects. Br J Psychiatry 1968; 114: 797-811.

[23] Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 1989; 39(9): 1159-65.

[24] McKhann G, Drachman D, Folstein MF, Katzman R, Price D, Stadlan EM. 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 1984; 34: 939-44.

[25] American Psychiatric Association. Diagnostic and statistical manual of the mental disorders. 4th, text rev. ed. Washington, DC: 2000.

[26] Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameters: early detection of dementia: mild cognitive impairment (an evidence-based review). Report on the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001; 56: 1133-42.

[27] Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004; 256: 183-94.

[28] Hickie C, Snowdon J. Depression scales for the elderly: GDS, Gilleard, Zung. Clin Gerontol 1987; 6: 51-3.

[29] Davidson WM, Carroll JB. Speed and level components in time-limit scores, a factor analysis. Educ Psychol Meas 1945; 45: 29-35.

[30] Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas 1960; 20: 141-51.

[31] SAS Institute Inc. SAS OnlineDoc 9.1.3. Cary, NC: SAS Institute Inc.; 2002 [cited 2007]; Available from: http://support.sas.com/onlinedoc/913/docMainpage.jsp

[32] Schönemann PH. A generalized solution of the orthogonal Procrustes problem. Psychometrika 1966; 31: 1-10.

[33] Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika 1958; 23(3): 187-200.

[34] Rey A. L'examen psychologique dans les cas d'escéphalopathie traumatique. Archives de Psychologie 1941; 28: 286-340.

[35] Osterrieth PA. Le test de copie d'une figure complexe. Archives de Psychologie 1944; 30: 206-356.

[36] Wechsler D. Wechsler Adult Intelligence Scale-III (WAIS-III) manual. New York: The Psychological Corporation 1997.

[37] Wechsler D. A standardized memory scale for clinical use. J Psychol 1945; 19: 87-95.

[38] Wechsler D. Wechsler memory scale-revised. San Antonio: The Psychological Corporation 1987.

[39] Tuokko H, Hadjistavropoulos T, Miller JA, Beattie BL. The clock test: a sensitive measure to differentiate normal elderly from those with Alzheimer's disease. J Am Geriatr Soc 1992; 40: 579-84.

[40]   Golden CJ. Stroop Color and Word Test: a manual for clinical and experimental uses. Chicago: Stoelting 1978.

[41]   Benedict RHB, Groninger L. Preliminary standardization of a new visuospatial memory tests with six alternate forms. Clin Neuropsychol 1995; 9: 11-6.

[42]   Benton AL, Hamsher KS. Multilingual aphasia examination. Iowa City: University of Iowa 1976.

[43]   Brandt J. The hopkins verbal learning test: development of a memory test with six equivalent forms. Clin Neuropsychol 1991; 5: 125-42.

[44]   Kaplan EF, Goodglass H, Weintraub S. The Boston naming test. Philadelphia: Lea & Febiger 1978.

[45]   Mack WJ, Freed DM, Williams BW, Henderson VW. Boston naming test: shortened versions for use in Alzheimer's disease. Journal of Gerontology: Psychol Sci 1992; 47: 154-8.

[46]   Money JA. A standardized road map of directional sense, manual. San Rafael: Academic Therapy Publications 1976.

[47]   Reitan RM. Validity of the trail making test as an indicator of organic brain damage. Percept Mot Skills 1958; 8: 271-6.

[48]   Chapman RM, Mapstone M, Porsteinsson AP, *et al.* Diagnosis of Alzheimer's disease using neuropsychological testing improved by multivariate analyses. J Clin Exp Neuropsychol 2010; In press.

[49]   Velicer WF, Jackson DN. Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. Multivariate Behav Res 1990; 25(1): 1-28.

[50]   Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988; 103: 265-75.

[51]   Fabrigar LR, MacCullum RC, Wegener DT, Stahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods 1999; 4: 272-99.