AN EXAMINATION OF THE LATENT STRUCTURE OF THE NEUROPSYCHOLOGICAL TEST BATTERY OF THE WISCONSIN REGISTRY FOR ALZHEIMER'S PREVENTION

Maritza Dowling[1,2], PhD, Bruce P. Hermann[2], PhD, Asenath La Rue[2], PhD, Mark A. Sager[2], MD

[1] University of Wisconsin-Madison, Department of Biostatistics and Medical Informatics, Madison, WI, USA ; [2] Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

BACKGROUND

Neuropsychological test batteries used in the investigation of preclinical Alzheimer's disease (AD) typically include multiple tests that are complex and intercorrelated with an unknown latent structure that may vary across groups (e.g., gender, age, risk status, race and ethnic background). Empirical knowledge of the dimensionality of constructs measured by commonly-used cognitive tests in clinical diagnosis not only helps in **summarizing** and **categorizing** cognitive data, but also in the development of parsimonious causal models that focus on a valid set of constructs as outcomes. Knowledge of the underlying latent structure **reduces redundancy** and the risk of Type I error rates due to multiple comparisons and makes it easier to gain understanding about cognitive domains sensitive to impairment facilitating in turn the study of change in cognition across time.

The present study had a two-fold purpose. The first aim was to examine the validity of an underlying latent structure of a neuropsychological test battery with components frequently used in the investigation of preclinical AD. Once the latent model was defined, a second aim was to test for measurement invariance across subgroups.

PARTICIPANTS AND INSTRUMENTS

This study included 1,073 cognitively healthy adults enrolled in a prospective longitudinal cohort study entitled the Wisconsin Registry for Alzheimer's Prevention (WRAP) being conducted since 2001 by the Wisconsin Alzheimer's Institute (WAI). The analysis included individuals with valid baseline data on **17** widely-used psychometric tests (see Table 1) administered as part of the WRAP entry assessment protocol. The battery assesses major domains of cognitive function including intelligence, language, visuoperception, learning and memory, and executive function providing a comprehensive appraisal of constructs relevant to normal aging and the prediction, diagnosis, and treatment of cognitive impairment.

WRAP participants range in age from 36 to 67 (mean = 53 years, S.D. = 7), are predominatly white/Caucasian (98%), have an average education level of 16 years (college degree), and a gender composition of 70% female. The majority (74%) have a family history of Alzheimer's disease and 40% of these are also carriers of the Apolipoprotein (APOE) ε4 allele.

TABLE 1. NEUROPSYCHOLOGICAL TESTS INCLUDED IN THE ANALYSIS

Number of Tests	Test or Subscale Name
4	• Wechsler Abbreviated Scales of Intelligence (WASI) Vocabulary and Similarities subtests • Wide Range Achievement Test-3: Reading scale • Boston Naming Test
4	• WASI Block Design and Matrix Reasoning subtests • Judgment of Line Orientation • Wisconsin Card Sort -64 items: perseverative errors
4	• Wechsler Adult Intelligence Scale-III: Digit Span, Arithmetic, and Letter-Number Sequencing subtests • Controlled Oral Word Associations
2	• Rey Auditory Verbal Learning Test: Sum recalled on learning trials and delayed recall
3	• Stroop Color Word Test: Interference trial • Trail Making Test : A and B

METHODS

Samples

To ensure comparability of analytical samples, a two-step procedure was followed. First, the sample of participants with valid scores on the 17 psychometric tests at baseline (N= 1,073) was stratified by variables generally associated with cognitive latent processes such as IQ scores, years of education, age, and APOE ε4 status (1) . In the second step, participants were randomly assigned within strata to one of two groups: the exploratory factor analysis (EFA) group (N=535) and the confirmatory factor analysis (CFA) group (N=538). Samples were also balanced by gender distribution. (See Table 2.)

Exploratory and Confirmatory Factor Analysis

Tests were subjected to an EFA analysis to gain initial insights about scale dimensionality. Factors were extracted using a principal axis factoring algorithm and a correlation matrix as input. To facilitate interpretation and allow factors to be correlated, we performed an oblique (Promax) rotation of the latent factors. Variables were selected in the final solution if their factor loadings were greater than 0.29 and loaded high on a single factor. The fit of the factor structure produced by the EFA solution was assessed by a confirmatory model also allowing the latent constructs to co-vary with the other four latent constructs.

Data analyses for the CFA model were conducted in two stages: **First**, the a priori model parameters were estimated for the whole sample and separately for each of eight groups of interest shown in **Figure 1, Step 2**. **Second,** factorial invariance tests were conducted to examine the assumption that the latent structure underlying the psychometric test scores was valid for making inferences across diverse groups. An advantage of performing these mutigroup analysis is that it provides a test for the significance of any difference that may exist across subgroups allowing the identification of the specific model parameters by which the two groups differ (2).

Figure 1: Data Analysis Flowchart

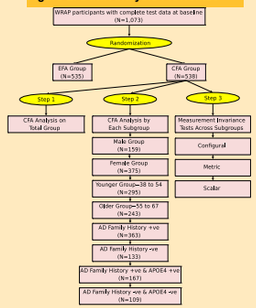


Table 2: Analytical Groups Comparison by Stratifying Variable

Stratifying Variables	*Analysis Type*	*N*	*Mean*	*SD*	*t*	*p-value*
Age	EFA	535	53.51	6.79	0.50	0.61
	CFA	538	53.30	6.55		
IQ (Scaled)	EFA	535	229.07	19.15	1.21	0.23
	CFA	538	227.64	19.7		
Gender=Male	EFA	531	0.30	0.46	-0.07	0.94
	CFA	534	0.30	0.46		
Education (yrs)	EFA	529	16.16	2.69	0.80	0.42
	CFA	532	16.03	2.87		
APOE4=Yes	EFA	523	0.36	0.48	-0.65	0.52
	CFA	528	0.38	0.49		
Group=AD	EFA	535	0.74	0.44	0.08	0.93
	CFA	538	0.73	0.44		

Factorial Invariance Tests

A hierarchy of models were fit to the data with increasingly stringent constraints imposed on the factor structure (3). The following three forms of invariance were examined:

- **full configural invariance**, which requires the same **patterns** of freed and fixed factor loadings across groups,
- **metric invariance,** requiring the same factor loading matrices across groups (same factor variable regressions) , and
- **scalar (strong) invariance**, which constrains factor loadings and test means (manifest variables) to be the same across groups.

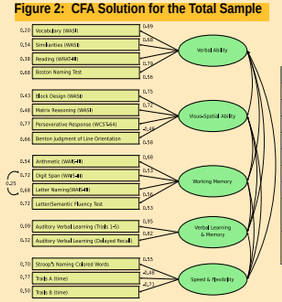
All CFA models were fitted using LISREL 8.8 (4). Models were tested using covariance matrices and model parameters were estimated using maximum likelihood (ML) estimation. Additionally, we used the asymptotic covariance matrix of the sample covariance matrix as a "weight matrix" to produce robust standard errors and chi-square values adjusted for non-normality. Model fit indices produced by the confirmatory factor solution were used to assess the extent to which the proposed model fitted the sample data. Criteria for an excellent model fit were CFI ≥.95 and RMSEA ≤ .06, while an adequate model fit was defined as CFI ≥

RESULTS

The **EFA** identified a five-factor structure with all 17 tests loading significantly high on a single factor (loadings varied, in absolute value, from 0.28 to 0.93). The strength of the factor loadings suggested that each of the subscales effectively measured its respective construct. The magnitude of the correlations between the factors (0.26 to 0.55) indicated a **high level of interconnectedness** among measures. The five-factor solution accounted for **65.5%** of the total variance among the tests.

Based on the EFA results, we hypothesized that the WRAP neuropsychological test battery consisted of **five** multi-test domains (see **Figure 2)** labelled as**: Verbal ability** with 4 tests; **Visuo spatial ability**, 4 tests; **Working memory**, 4 tests; **Verbal learning & memory**, 2 tests; and **Speed and flexibility,** 3 tests. The fit of the CFA model for the total sample provided support for the five-factor structure. The root mean square error of approximation (RMSEA) estimate was acceptable (0.067) yielding a 90% CI = (0.056, 0.078). Fit indices such as the Normed Fit Index (NFI), Comparative Fit Index (CFI), and Goodness-of-Fit Index (GFI), were within acceptable ranges (respectively, 0.94, 0.96, 0.91). Based on fit indices, CFA solutions for each of the eight subgroups was also within acceptable ranges (see summary in **Table 3**).

Figure 2: CFA Solution for the Total Sample



Note: Fit Indices for the Total CFA Sample: RMSEA=0.07; NFI=.094; CFI=0.96; and GFI=0.91.

Table 3: Summary of Fit Indices for the CFA Solution By Subgroup

		CFA Fit Indices			
Subgroup	N	RMSEA	NFI	CFI	GFI
Male	159	0.07	0.91	0.96	0.86
Female	375	0.06	0.95	0.97	0.92
Young (30-54)	295	0.07	0.93	0.96	0.90
Older (55-67)	243	0.07	0.92	0.95	0.88
AD Fam Hist +ve	363	0.07	0.93	0.95	0.89
AD Fam Hist -ve	133	0.04	0.92	0.98	0.89
AD Fam Hist +ve, APOE4 +ve	167	0.06	0.91	0.96	0.89
AD Fam Hist -ve, APOE4 -ve	109	0.04	0.90	0.98	0.88

As shown in **Table 4,** no significant deterioration in fit is observed across invariance tests. Overall, the **multigoup invariance** tests examined suggested reasonable model consistency supporting the utility of the factor structure as a measure of relatively stable traits across subgroups.

Table 4: Summary of Fit Indices By Subgroup and Type of Invariance Test

GROUP	Configural			Metric			Strong		
	RMSEA	NFI	CFI	RMSEA	NFI	CFI	RMSEA	NFI	CFI
Male & Female	0.07	0.93	0.96	0.07	0.92	0.96	0.077	0.91	0.94
Young & Old	0.07	0.91	0.95	0.07	0.91	0.95	0.066	0.91	0.94
AD FH +ve & AD FH -ve	0.076	0.92	0.95	0.06	0.92	0.95	0.07	0.91	0.95
AD FH +ve & APOE +ve AD FH -ve & APOE -ve	0.06	0.90	0.96	0.06	0.90	0.96	0.06	0.89	0.96

Note: X ² difference tests were also performed but not reported. Compliance with the invariance tests were based on the goodness of fit indices reported here.

CONCLUSIONS

The factor structure model identified by the EFA solution and examined in a series of CFA models lends empirical support to the validity of the constructs (both discriminat and convergent validity). Form a conceptual point of view, the subscales exhibited sufficient breadth of content to capture the domain of each of the five constructs. Results also provided evidence of the validity of a **reduced-dimensional representation** of a set of 17 neuropsychological scales commonly used in clinical evaluations and what specific scales can be meaningfully combined into a global measure representing a common trait of potential clinical relevance in studying cognitive aging. Studies like this serve to inform the practice of combining scores on multiple tests or sets of items across test subscales into constructs that help identify and diagnose different aspects of cognitive functioning. Higher precision can often be achieved by combining several measures of the same construct into a composite score.

The utility of a global measure depends on the extent to which the scales forming the composite are congeneric, that is, assess the same latent construct (5) with likely different units of measurement and precision. The hypothesized latent structure also allows the estimation of the reliability of a single summary measure or composite reliability and consequently how appropriately it captures the essence of examinee performance across the scales.

This analysis also demonstrated the generality of the proposed measurement model to make comparisons across subgroups of interest. Measurement invariance is necessary for valid inference and interpretation, particularly in studies comparing change over time in cognitive performance across different diagnostic groups. For example, satisfying configural invariance allows us to be confident about comparing two groups using the same factor structure since it guarantees that the pattern of factor loadings across the groups is similar. Metric invariance allows us to further claim that the actual loadings of each scale on the factors are the same across groups. That is, it allows us to evaluate whether or not the factors have the same meaning across the two groups. As a consequence, the comparison of variance across groups within a given factor becomes more meaningful. Finally, a test of scalar invariance is essentially a "differential item functioning test." That is, satisfying a test that fixes the factor loadings and the means or intercepts to be the same across groups further allow us to infer that the mean score on each factor component is similar across groups.

REFERENCES

1. Small, G. W., Chen, S. T., Komo, S., Ercoli, L, Bookheimer, S., Miller, K., Lavretsky, H., Saxena, S., Kaplan, A., Dorsey, S., Scott, K., Saunders, A.M., Haines, J. L., Roses, A.D., and Pericak-Vance, M. (1999). Memory self-appraisal in middle-aged and older adults with the apolipoprotein E-4 allele. American Journal of Psychiatry, 156, 1035-1038.

2. Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations, Psychometrika, 36, 409-426.

3. Meredith, W. (1993). Measurement invariance, factor analysis and factor invariance. Psychometrika, 58, 675-680.

4. Jöreskog, K. G., and Sörbom, D. (1996). LISREL8: Structural equation modeling with the SIMPLIS command language. Hillsdale, NJ: Erlbaum.5. Hu & Bentler (1999

5. Lord, F. M., & Novick, R. (1968). Statistical theories of mental test scores. Reading MA: Addison-

ACKNOWLEDGEMENTS

This research was supported by the NIA grant 5R01AG27161-2 (Wisconsin Registry for Alzheimer Prevention: Biomarkers of Preclinical AD).