Published in final edited form as:

Neuropsychology. 2010 November ; 24(6): 742–756. doi:10.1037/a0020176.

Latent Structure and Factorial Invariance of a Neuropsychological Test Battery for the Study of Preclinical Alzheimer's Disease

N. Maritza Dowling, Bruce Hermann, Asenath La Rue, and Mark A. Sager

University of Wisconsin-Madison

Abstract

Objective—To examine the latent structure of a test battery currently being used in a longitudinal study of asymptomatic middle-aged adults with a parental history of Alzheimer's disease (AD) and test the invariance of the factor solution across subgroups defined by selected demographic variables and known genetic risk factors for AD.

Method—An exploratory factor analysis (EFA) and a sequence of confirmatory factor analyses (CFA) were conducted on 24 neuropsychological measures selected to provide a comprehensive estimate of cognitive abilities most likely to be affected in preclinical AD. Once the underlying latent model was defined and the structural validity established through model comparisons, a multi-group confirmatory factor analysis model was used to test for factorial invariance across groups.

Results—The EFA solution revealed a factor structure consisting of 5 constructs: verbal ability, visuo-spatial ability, speed & executive function, working memory, and verbal learning & memory. The CFA models provided support for the hypothesized 5-factor structure. Results indicated factorial invariance of the model across all groups examined.

Conclusions—Collectively, the results suggested a relatively strong psychometric basis for using the factor structure in clinical samples that match the characteristics of this cohort. This confirmed an invariant factor structure should prove useful in research aimed to detect the earliest cognitive signature of preclinical AD in similar middle aged cohorts.

Keywords

Latent variable analysis; factorial invariance; partial measurement invariance; Alzheimer's disease; family history of dementia; neuropsychological test battery

INTRODUCTION

Alzheimer's disease (AD) is a chronic neurodegenerative disorder affecting an estimated 5.2 million Americans in 2008. According to the Alzheimer's Association, the number of

Correspondence concerning this article should be addressed to N. Maritza Dowling, Ph.D., Department of Biostatistics & Medical Informatics, University of Wisconsin School of Medicine and Public Health, J5/M160 Clinical Sciences Center, 600 Highland Avenue, Madison, WI. 53792-4675. Phone: (608) 263-4100. nmmdowlin@biostat.wisc.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/neu

persons with AD could reach 16 million by 2050 unless means are found to either delay or prevent the onset of the disease (Alzheimer's Association, 2009). Traditionally, AD has been considered a disease of older adults over age 65. However, there is increasing evidence that neurobiological changes consistent with AD can be found decades before a diagnosis is made (Bookheimer & Burggren, 2009). The early identification of disease pathology in younger asymptomatic persons provides the opportunity to intervene at early stages and potentially modify the disease course.

A major challenge in the early recognition of AD has been to characterize the pre-clinical cognitive changes occurring in persons at increased risk, including those who may have an APOE $\epsilon 4$ allele or a family history of AD. There is increasing evidence that studies of pre-clinical AD may be particularly valuable for those persons with a family history (Jarvik et al., 2008). Recent studies have found evidence for hippocampal dysfunction in neuroimaging (Bassett et al., 2006; Johnson, Schmitz, Trivedi, et al., 2006) and neurocognitive performance that are suggestive of pre-clinical AD in asymptomatic persons with a parental family history of AD (La Rue et al., 2008). In addition, elevated levels of plasma amyloid β have been observed in asymptomatic first-degree relatives of AD patients (Ertekin-Taner et al., 2008) suggesting that family history may be especially valuable in the study of preclinical AD.

Studies of preclinical AD focus on younger, relatively healthy, asymptomatic persons for whom brief cognitive screening batteries are not appropriate. As a consequence, these studies require more extensive and sensitive batteries of neuropsychological tests assessing a range of cognitive domains. These tests are often used under the assumption that underlying cognitive constructs are uniform or equivalent across groups and time. However, the tests may have a latent structure that may vary across demographic and AD risk factors (for example, gender, age, APOE genotype, or family history) seriously compromising the generalizability of findings across groups. The ability to define this latent structure, i.e., underlying cognitive domains, using a cluster of tests for each domain instead of individual tests, has the advantage of reducing redundancy and increasing the reliability of the summary measures. Assuming test measures with high quality psychometric characteristics, composite scores derived from a stable latent variable solution can be used as outcomes in further analyses reducing the risk of Type I error due to multiple tests. This may be particularly important in the early detection of AD and other dementias as well as in the monitoring of changes in cognitive functioning occurring over time in asymptomatic persons.

The examination of differences and similarities across subpopulations of interest on a given set of constructs also becomes more informative when the underlying latent structure of the test battery is invariant across patient groups. Invariance refers to the extent to which "under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992, p. 117). Lack of evidence supporting model invariance compromises comparison between groups because the meaning of the underlying constructs becomes group specific (Meredith, 1993).

Investigating construct comparability is a necessary approach to demonstrate that observed differences in scale scores represent true differences between groups and not differences due to systematic biases caused by non-equivalence of constructs. For example, a significant number of studies have found gender and age-related differences in verbal learning and memory (Bleecker, Bolla-Wilson, Agnew, & Meyers, 1988; Kramer, Delis, & Daniel, 1988; Geffen, Moar, O'Hanlon, Clark, & Geffen, 1990). Other studies (e.g., Revell & Schaie, 2004; Schaie, 2005; and Hofer et al., 2002) have found an association between cognitive performance in memory domains and the presence of an APOE $\epsilon 4$ allele in preclinical

samples, particularly for individuals with the $\epsilon 4/\epsilon 2$ and $\epsilon 4/\epsilon 4$ pairings. Given the importance of verbal learning and memory-related constructs as possible preclinical indicators of AD, supporting evidence of measurement invariance (or equivalence between observed and latent variables) across groups (e.g., age and gender, genotype) is critical for inferential purposes. Additionally, since the interplay of AD family history, APOE genotype, and middle-age populations have not been extensively studied, there is little understanding of the latent structure of comprehensive neuropsychological test batteries used to evaluate asymptomatic populations at various levels of risk and its validity across patient groups.

The present study was undertaken with a two-fold purpose. The first aim was to define the latent structure of a psychometric test battery currently being used in a study of pre-clinical AD in the Wisconsin Registry for Alzheimer's Prevention (WRAP; Sager, Hermann, & La Rue, 2005). WRAP is a longitudinal cohort study of an important, but seldom studied group, i.e., asymptomatic middle aged persons with a parental family history of AD, and a control group of persons whose parents lived to late life without AD. The primary goal of the WRAP study is to define the neurobiological course of preclinical AD which is a necessary first step in developing interventions to modify the disease course. Second, once the underlying latent model was defined and the structural validity established through model comparisons, we tested for factorial invariance across groups defined by selected demographic variables (age and gender) and known genetic risk factors for AD (parental family history and APOE genotype) using a multi-group confirmatory factor analysis model (Jöreskog, 1971).

We hypothesized that multiple, partially independent dimensions of cognitive performance would be identified through these analyses, including a secondary memory factor that may be sensitive to preclinical AD. In addition, given the relatively young mean age and clinically intact cognitive status of the sample, we predicted that the factor structure would be invariant for age, gender, and genetic risk subgroups.

METHOD

Participants

The study sample is from the WRAP investigation in which middle aged asymptomatic research volunteers are administered a battery of neuropsychological tests at 4 year intervals. Over two thirds of the participants (73%) have a parent with either autopsy-confirmed or probable AD as defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 1984). The remaining 27% of the WRAP sample are control participants without a parental history of AD.

WRAP enrollees with complete scores on a battery of psychometric tests administered as part of the entry assessment protocol were included in this study. The sample consisted of 1,288 participants, ranging in age from 36 to 67 (mean = 53 years, S.D. = 7), with an average education level of 16 years (S.D. = 2.71), and a gender composition of 70% female. The sample was predominantly Caucasian (97.7%). Approximately 44% of the participants with a family history of AD were also carriers of at least one Apolipoprotein (APOE) $\epsilon 4$ allele, a known risk factor for AD (see, for example, Etnier et al., 2007; Bartrés-Faz et al., 2002; Hyman et al., 1996). In contrast, approximately 18% of the participants without a family history of AD (WRAP registry controls) carried one or more $\epsilon 4$ alleles.

Neuropsychological Measures

The cognitive test battery consisted of standardized, widely used clinical neuropsychological measures administered according to standard procedures in a single 2.5 hour session. Tests

were selected to provide a comprehensive estimate of cognitive abilities, with an emphasis on those most likely to be affected in preclinical or early-stage AD (e.g., learning and memory and executive function). Table 1 summarizes the specific measures included in the present analyses, grouped according to the cognitive domains they were intended to assess. The individual trials of the AVLT were entered in the analyses as opposed to a single composite measure (total words recalled) because we anticipated that later learning trials and delayed recall would constitute a more reliable indicator of anterograde memory than the initial learning trials.

ANALYSES

Samples

An overview of the analysis plan is presented in the Supplemental Materials which is available online. The analyses were performed using a split-sample or cross-validation approach. To ensure comparability of analytical subsamples, a two-step procedure was followed. First, the sample of participants with complete data on the 24 psychometric tests at baseline ($N = 1,288$) was stratified by variables generally associated with cognitive latent processes such as IQ scores, years of education, age, and APOE $\epsilon 4$ status (see, for example, Small et al., 1999). Gender and AD family history status were also included as stratifying variables to balance the groups. In the second step, we used urn randomization (Wei & Lachin, 1988) to assign participants within strata to one of two groups: the exploratory factor analysis (EFA) group ($N = 638$) and the confirmatory factor analysis (CFA) or “testing” group ($N = 650$). Table 2 shows the composition of the resulting subsamples by stratifying characteristics.

EFA, EFA within CFA Framework, and CFA

The examination of the latent structure of the WRAP cognitive test battery was conducted following a sequence of analysis steps with increasing levels of restrictions imposed on the solutions, namely, EFA, which uses minimal identification constraints, EFA within the CFA framework (EFA-CFA, Jöreskog, 1969), CFA, and multi-group CFA. The initial step employed a linear multi-factor model, henceforth EFA, to study the nature and appropriate number of factors underlying the neuropsychological test measures and characteristics of cross loadings. Factors for the exploratory latent structural analysis or EFA were extracted using robust maximum likelihood estimation. To facilitate interpretation and allow factors to be correlated, we performed an oblique (Promax) rotation of the latent factors. Alternative measurement models with a different number of latent factors were evaluated and compared using fit indices. Test indicators were selected in the final solution if their primary factor loadings were greater than 0.31 (Tabachnick & Fidell, 2002) but below 0.95 and loaded high on a *single factor*. Alternative measurement models with a different number of latent factors were evaluated and compared using the standardized root mean square residual (**SRMR**), the root-mean-squared error of approximation (**RMSEA**), and its 90% confidence interval as model fit indices. A final model was selected based on the following five criteria: (1) overall fit, (2) a structure characterized by factors with more than two indicators, (3) interpretability (simple structure), (4) theoretical significance, and (5) parsimony.

As part of the exploratory phase and prior to moving to more restrictive CFA models, we also conducted an EFA-CFA analysis as described in Jöreskog (1969), which further allows us (a) to test whether factor cross-loadings and correlations are statistically significant by providing standard error estimates and (b) to determine the presence of large residual covariances (which generally suggest minor factors or method effects) by providing modification indices for indicator error covariances. The latter is a source of misfit that is not detected by EFA (Brown, 2003).

In the next step, we assessed the fit of the factor structure of the measurement model produced by the exploratory factor analysis using a confirmatory model. Data analyses for the CFA model were conducted in two stages: First, the covariance among observed variables was estimated for the whole cross-validation sample ($N = 650$) and the a priori or hypothetical baseline model was fitted. Second, covariances among observed variables were computed for each of eight groups of interest followed by CFA parameter estimation.¹ The eight single-groups were defined as: (1) male ($N = 200$), (2) female ($N = 450$); (3) younger (36 to 54 years old, $N = 330$), (4) older (55 to 67 years old, $N = 320$); (5) AD family history ($N = 469$), (6) controls or no AD family history ($N = 181$); (7) AD Family history and positive APOE $\epsilon 4$ allele ($N = 187$), and finally, (8) no AD family history and negative APOE $\epsilon 4$ ($N = 154$). The last two pairs of groups represent a comparison between those at the highest and lowest genetic risk for AD.

All models were fitted using latent variable software programs (LISREL 8.8, Jöreskog & Sörbom, 2007; MPLUS 5.2, Muthén & Muthén, 1998–2008). The models were tested using sample variance-covariance matrices (sample matrix is provided in the Supplemental Materials) as input and parameters were estimated using robust maximum-likelihood minimization functions. Models fitted in LISREL 8.8 used the asymptotic covariance matrix of the sample covariance matrix as a “weight matrix” to produce robust standard errors (less biased) and reported chi-square (χ^2) values adjusted for non-normality in the distribution of the factor indicators (Satorra & Bentler, 1988; 1994; Hu & Bentler, 1995).

To increase the reliability of model solution evaluation, we used multiple indices of fit: the normed and the non-normed fit indexes (**NFI** and **NNFI**, the comparative fit index (**CFI**), and the root-mean-squared error of approximation (**RMSEA**) and its 90% confidence interval (MacCallum, Browne, & Sugawara, 1996). For the first three of these common indices, values above 0.90 indicate a good fit (Schumacker & Lomax, 1996). RMSEA values below 0.05 denote “good fit” and under 0.08 can be considered acceptable (Browne & Cudeck, 1993). Hu & Bentler (1999) recommend a $CFI \geq 0.95$ and $RMSEA \leq 0.06$ for an “excellent” model fit and $CFI \geq 0.90$ and $RMSEA \leq 0.08$ for an “adequate” model fit. We defined a model as acceptable if the following criteria were met: $CFI > 0.90$; NFI and $NNFI > 0.90$; and $RMSEA < 0.08$, 90% $CI < 0.08$.

The χ^2 statistic is a traditional measure of overall fit in covariance structure models. However, sample sizes above 150 tend to produce a significant χ^2 resulting in the rejection of the proposed model, which was generally the case in our analysis. This sample size dependency of the χ^2 statistic, has led to a large number of alternative approaches to assessing model fit (Kaplan, 2009). In this case, it is useful to rely on indices such as NFI and $NNFI$ that compare the χ^2 for the full unconstrained model with that of the “null” or independence model (i.e., a model that postulates uncorrelated indicator variables). These indices reflect the proportion of reduction in the χ^2 of the full model over the null model (Bentler, 1980). The ratio χ^2/df can also be computed as another reference for model fit. Yet, the problem of sample size dependency cannot be completely eliminated by this procedure (Bollen, 1989, p. 278). Several thresholds have been recommended for the ratio χ^2/df ranging from a value as low as 1 and as high as 5 (see for example, Marsh & Hocevar, 1985; Byrne, 1991). In this study, a ratio between 2 and 4 for χ^2/df was considered a satisfactory data-model fit.

Finally, in multi-group CFA testing where competing models were nested, comparative fit was evaluated with χ^2 difference tests in addition to the comparative Akaike information

¹Education is a well-known risk factor for AD; however, because the majority of WRAP participants are well educated, we did not compare education subgroups.

criterion (CAIC) (Bozdogan, 1987). Decisions concerning model modification were guided by a combination of expected change (EC) statistics and modification indices (MIs; also known as Lagrange multipliers or estimates of the reduction in χ^2 if the constrained parameters were estimated (cf. Saris, Satorra, & Sörbom, 1987). The interpretation of the size and direction of the ECs in tandem with MIs are necessary, particularly when sample sizes are large (cf. Kaplan, 2009).

Testing For Factorial Invariance

Once it is established that single-group models fit the data well, another important step in validating a hypothesized factor structure is the assessment of whether numerical values under consideration have the same measurement scale to allow meaningful comparisons across groups with respect to a given trait (Drasgow, 1984). In this study, factorial invariance examined the assumption that the latent structure underlying the psychometric test scores was valid for making inferences across subgroups of the same sample (Alwin & Jackson, 1981). This was accomplished through a hierarchy of models fit to the data with increasingly stringent constraints imposed on the factor structure (Meredith, 1993). There are different forms of factorial invariance in CFA that can be distinguished according to the pattern of constraints (see, for example, Byrne, Shavelson, & Muthén, 1989; Meredith, 1993; Horn & McArdle, 1992; Widaman & Reise, 1997; Steenkamp, 1998). Our analysis followed the steps for testing invariance proposed by Steenkamp (1998) and focused primarily on the following five tests of group invariance: (1) configural invariance (or identical factor structure), which requires the same patterns of freed (nonzero) and fixed (zero) factor loadings across groups, (2) metric invariance, requiring the same factor loading matrices across groups (i.e., the regression coefficients or “slopes” of underlying factors are constrained to be the same across groups), (3) scalar (strong) invariance, which constrains *both* factor loadings (slopes) and intercepts of manifest variables (scale means) to be the *same* across groups, (4) factor variance-covariance invariance, which restricts the latent variables (factors) covariances and correlations to be the same, and (5) strict invariance (error variance invariance), which further constrains the measurement error covariance matrices of observed variables to be the same across groups.

To test for factorial invariance across the two gender groups, two age groups, two types of family history groups, and two types of family history and APOE status groups, we employed a multi-group confirmatory factor analytic model (Jöreskog, 1971; Sörbom, 1974) that estimates parameters and test hypotheses about both groups in a single analysis. An advantage of performing a multi-group analysis over separate single group analyses, besides improving the accuracy of parameter estimates, 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.

RESULTS

Exploratory Factor Analysis

A sequence of EFA solutions, varying the number of factors extracted from four to six, identified some tests performing poorly as factor indicators. Six tests or subscales (*Clock drawing*, *WAIS-Arithmetic*, *WMS-Faces Immediate & Delay*, and *AVLT-Trials 1 & 2*) were dropped from the analysis for several reasons including (a) the presence of salient cross-loadings (*WAIS-Arithmetic*), (b) consistent low loadings across factors (below 0.30; *Clock drawing & WMS-Faces*), and/or (c) loading on a given factor with only one additional indicator (*AVLT-Trials 1 & 2*). Across solutions, the Clock drawing subtest contributed to all factors, but not sufficiently enough to be included in any. Similarly, and consistent with other factor analytic studies, Faces showed relatively low correlations, not only with

memory tests, but also with other domains measured by the scales included in the analysis (Holdnack & Delis, 2004; Millis, Malina, Bowers, & Ricker, 1999).

Thus, an EFA was rerun using a reduced pool of 18 subscales. (A summary of the fit indices for the sequence of EFA analyses is provided in the online Supplemental Materials.) None of the solutions provided a poor fit. The six-factor solution provided the best fit, $\chi^2(60) = 80.28$, RMSEA = 0.02 (0.01, 0.04), and SRMR = 0.01. Factor determinacies (i.e., correlations between estimated factor scores from the observed indicators and related latent factors) were above the cut-off of 0.80 (Gorsuch, 1983, p. 260) ranging from 0.84 to 0.95. Factor determinacies can be used as a measure of the validity of estimated factor scores (Vittadini, 1989; Grice, 2001). Given a high factor determinacy, it is possible to use factor scores as “substitutes” of the factor itself in situations when a latent structural analysis is not a feasible option (Brown, 2003; cf. Grice, 2001).

Despite the satisfactory model fit characteristics described above, the six-factor model produced a small negative residual variance for the WASI-Vocabulary scale (−0.03). Negative error variances or “Heywood cases” may be due to multiple causes, with model misspecification just being one of them (van Driel, 1978; Bollen, 1989). The extraction of more factors than what the data can afford may also result in Heywood cases (Sato, 1987; Loehlin, 2004). The five-factor solution, however, produced no Heywood cases and also provided a good fit to the data ($\chi^2(73) = 124.5$, RMSEA = 0.03 (0.02, 0.04), SRMR = 0.02). The solution also produced factors explaining the correlations amongst the test scales that were more meaningful from the clinical point of view than the solution produced by the six-factor model. Additionally, factor determinacies were relatively high ranging from 0.87 to 0.93. Consequently, the five-factor model, explaining 63.3% of the total variance among the scales, was the preferred solution from EFA. Table 3 shows the tests producing primary factors loadings for a given factor. Primary factor loadings varied, in absolute value, from 0.33 (fluency) to 0.89 (WASI-Vocabulary). The intercorrelations among the five factors ranged from 0.22 to 0.45 justifying a *Promax* rotation. The five factors were labeled as indicated in Table 3.

Exploratory Factor Analysis within a Confirmatory Factor Analytic Framework

We used the five-factor EFA solution as reference to select the anchor and fixed items (i.e., test scales) for each factor to perform the EFA/CFA analysis. The model produced a good fit to the data ($\chi^2(73) = 124.5$, RMSEA = 0.03 (0.023, 0.043, $p < 0.05 = 0.998$), CFI = 0.99, NNFI = 0.98). As illustrated in Table 3, the size of primary factor loadings varied, in absolute value, between 0.34 and 0.94 indicating moderate to strong factor-indicator relationships. All the loadings were statistically significant with *z-estimates* ranging from 3.02 to 17.88. Factor determinacies were also relatively high (0.86 to 0.96). Factor intercorrelations ranged between 0.02 and 0.52. The following correlations were statistically significant: verbal ability with working memory (Factors 1 and 4, $\hat{\rho} = 0.51$, $p = 0.01$) and speed & executive function with verbal learning & memory (Factors 3 and 5, $\hat{\rho} = 0.47$, $p = 0.02$). Finally, an examination of the MIs pertaining to correlations of indicator residuals did not reveal substantial evidence of a misspecification area in the model.

Confirmatory Factor Analysis

Figure 1 contains the graphical representation of the five-factor conceptual CFA model submitted to analysis. Table 4 shows the CFA solutions using the cross-validation sample ($N = 650$) and the following eight subgroups within the same sample: (1) male ($N = 200$), (2) female ($N = 450$), (3) young ($N = 330$), (4) older ($N = 320$), (5) AD family history +ve ($N = 469$), (6) AD family history −ve ($N = 181$), (7) AD Family history +ve and an APOE ε4 +ve ($N = 187$), and (8) AD family history −ve and APOE ε4 −ve ($N = 154$).

Total group—The overall fit of the CFA model provided support for hypothesized cognitive structure (χ^2 (125) = 450.16 ($p < 0.00$), $\chi^2/df = 3.60$, CFI=0.97, NNFI=0.96, NFI=0.95). The RMSEA estimate was acceptable (0.06) yielding a 90% CI = (0.057; 0.069). All of the model fit indices, except the χ^2 statistic, were satisfactory and within recommended thresholds. All standardized loadings were relatively high, ranging in absolute value, from 0.49 to 0.88 and statistically significant, which confirmed the convergent validity of the hypothesized five-factor CFA model. The composite reliability for the *total scale*, estimated using a covariance structure modeling procedure with nonlinear constraints outlined by Raykov (1998; 2001), was 0.89. Composite reliabilities for individual factors were also satisfactory: 0.70 (speed & executive function), 0.74 (visuo-spatial ability), 0.76 (working memory), 0.82 (verbal ability), 0.91 (verbal learning & memory). Furthermore, the quality of the factor score estimates, as measured by factor determinacy coefficients, was reasonably high ranging from 0.89 (speed & executive function) to 0.96 (verbal learning & memory).

The 10 estimated correlations between factors were positive and statistically significant ($p < 0.05$) and had magnitudes varying from 0.24 (verbal ability with speed & executive function) to 0.56 (visuo-spatial ability with speed & executive function) (see Figure 1). While the observed degree of interrelatedness among cognitive factors is typical and expected, the “low” to “moderate” magnitude of the intercorrelations supports the relative distinctiveness or discriminant validity of the five factors in measuring cognitive processes (Kline, 2005). The average variance extracted for each factor, ranging from 0.48 to 0.71, was higher than the square of the correlations among the five factors (spanning from 0.06 to 0.31). This provided further evidence of discriminant validity (cf. Fornell & Larcker, 1981).

Single groups—Overall, fit indices indicated a well-fitting and theoretically viable baseline model to conduct invariance tests (see Table 5). Primary factor loadings were all significant ($ps < 0.05$) and maintained a relatively stable strength both across most groups within the same latent factor and across most of the latent factors within the same group. Factor loadings across groups showed most stability for the verbal learning & memory construct (see Table 4), likely due to the fact that all scores on this factor were derived from a single clinical test. All single group results indicated a reasonable fit of the five-factor model. Fit indices (except χ^2 tests) and factor determinacies were well within recommended thresholds. Composite scale reliabilities were also relatively high varying from 0.86 (AD–ve & APOE4–ve group) to 0.89 (male group and AD+ve group). There were no standout MIs requiring specific attention.

Factorial Invariance

Table 5 summarizes the results of the sequence of factorial invariance tests across all four pairs of groups. As noted in this table, some of the omnibus invariance tests failed using χ^2 difference tests (p -values < 0.05). Yet, none of the alternative model fit criteria were compatible with this result. In fact, all model fit indices remained within acceptable thresholds suggesting no major degrade in fit with the added constraint. Given this discrepancy, in cases where the invariance test did not hold (i.e., a significant decrease in fit occurred per χ^2 difference test), we proceeded to inspect MIs and ECs for prominent indicators of misfit and tested *partial invariance* accordingly. That is, we freed some of the parameters that appear to differ across groups and re-tested the psychometric equivalence of the model across groups. Results by multi-group comparison are presented next.

AD FH+ve - ADFH –ve groups—The series of two-group five-factor CFA models to evaluate invariance between the AD family history +ve and the AD family history –ve groups produced acceptable fit indices for all five forms of invariance examined: (1)

configural or equal *pattern* of factor loadings, (2) metric or equal factor loadings (slopes), (3) full scalar, (4) equal factor variance-covariance, and (5) strict or equal indicator variance-covariance. The χ^2 difference tests revealed that increasing levels of constraints did not significantly deteriorate the overall fit of the model (all p -values > 0.05). CAIC values also decreased steadily. In fact, the equal error variance solution produced a subtle improvement in the parsimony goodness-of-fit indices, for example, for the equal factor variance-covariance, RMSEA=0.0572 (0.051, 0.064) and CAIC = 1250.46 vs. RMSEA=0.056 (0.049, 0.062) and CAIC = 1136.11 for the equal error variance.

Younger-older groups—The simultaneous group analysis of invariance between the young and old groups yielded solutions that were consistent with good model fit with the exception of one test: full scalar invariance, which constrained item (scales) intercepts and factor loadings to equality between groups. As indicated in Table 5, this model produced a significant decrease in goodness of fit, $\chi^2_{diff}(13) = 48.11$ ($p < 0.01$), suggesting mean differences for some scales across the two groups. That is, for fixed levels of the latent construct, certain abilities measured by the scales appear to be more prevalent on one age group than the other. Still, model fit indices revealed an overall good fit to the data (RMSEA = 0.06 (0.054, 0.067), CFI = 0.96, NNFI = 0.96, NFI = 0.94, and CAIC=1334.31). The MIs and respective ECs indicated some group differences in favor of a) “younger” participants in Stroop CW and Fluency test scales and (b) “older” participants in Benton JLO. Therefore, using a partial scalar invariance model (Steenkamp, 1998), these scales were allowed to be freely estimated across the young and old groups. This model represents a compromise between metric and strict invariance and allows the study of what scales perform differentially across groups (Byrne et al., 1989). The resulting model produced a non-significant decrease in fit ($\chi^2_{diff}(10) = 16.56$, $p = 0.08$) indicating that partial scalar invariance was valid between the young-old groups. This finding suggests that at the same factor level, younger individuals tended to obtain higher scores on Stroop CW and Fluency, whereas, “older” individuals obtained higher scores in Benton, JLO. Both full factor variance-covariance and strict invariance did not degrade the fit of the solution.

Female-male groups—The examination of the degree of factorial invariance between female and male groups produced significant χ^2 difference tests for two forms of invariance: (1) full equality constraints of intercepts (scalar invariance) and (2) factor variance-covariances. That is, female and male groups appeared to differ in factor mean scores and factor intercorrelations. As shown in Table 5, the test of full scalar invariance, in particular, produced a highly significant increase in chi-square between the model of full metric and the model of full scalar invariance ($\chi^2_{diff}(13) = 121.70$, $p < 0.01$). However, other practical fit diagnostics did not provide evidence of misfit in the solution. An inspection of the MIs and ECs suggested that the intercepts of some of the scales were not invariant across sexes. On average, females performed slightly better than males on WASI-Matrix Reasoning, Fluency, and Rey AVLT3. Conversely, males fared better than females on Boston naming and Benton JLO. After successively allowing these intercept parameters to be freely estimated, the fit of the *partial scalar* invariance model improved substantially compared to the full scalar invariance ($\chi^2_{diff}(8) = 12.28$, $p = 0.14$).

The next step imposed full factor variance-covariance invariance on the model maintaining the intercepts that were allowed to be free in the previous step. The model produced a significant increase in chi-square ($\chi^2_{diff}(10) = 25.32$, $p < 0.01$) indicating that full factor variance-covariance was not supported across the female and males groups. The MIs suggested a difference in the factor variance between visuo-spatial ability and verbal ability and between visuo-spatial ability and speed & executive function. Allowing the covariances between these factors to be estimated improved the fit of the model ($\chi^2_{diff}(8) = 14.65$, $p = 0.07$). Finally, the first model specifying invariance of error variances given the previous

relaxation of constraints, was not significant ($\chi^2_{diff}(10) = 25.32, p < 0.01$). Hence, error invariance across gender was supported.

AD FH+ve & APOE4+ve versus AD FH–ve & APOE4–ve groups—Overall, invariance-test models across groups of individuals with a family history of AD and also carriers of an APOE $\epsilon 4$ allele and those with no family history of AD and APOE4 $\epsilon 4$ negative produced satisfactory practical fit indices. The chi-square test for strict factorial invariance constraining indicator error variances and covariances to be the same across group was statistically significant ($\chi^2_{diff}(18) = 32.54, p = 0.02$). The magnitude of the MIs for the error variance covariance matrix indicated a difference across groups in the scale Trails A. Further removing the invariance constraint on Trails A and re-fitting the model yielded a non-significant chi-square result ($\chi^2_{diff}(17) = 26.25, p = 0.07$) supporting partial error variance invariance.

DISCUSSION

There is now worldwide interest in identifying early cognitive signs of Alzheimer's disease prior to the development of conditions such as Mild Cognitive Impairment (see, for example, Petersen, 2003; Tuokko & Hultsch, 2006). These studies by necessity must focus on younger, relatively healthy, asymptomatic persons for whom brief cognitive screening batteries are not appropriate, and therefore, require more extensive and sensitive batteries of neuropsychological tests that assess a range of cognitive domains. Neuropsychological batteries are typically heterogeneous in terms of the tests administered but have considerable shared variance across tests and cognitive domains of interest. The focus of this investigation was to define the latent structure of a comprehensive test battery currently being used in a longitudinal study of preclinical Alzheimer's disease, the Wisconsin Registry for Alzheimer's Prevention (Sager et al., 2005). Once the latent structure was defined by a sequence of factor analysis models, we then tested the invariance of the constructs across subgroups of the cohort defined by various known risk factors for AD (age, gender, parental family history, and APOE genotype). The overarching goal of these efforts was to provide information that may prove helpful and accelerate research in preclinical AD by identifying meaningful and reliable dimensions of cognitive performance at a presymptomatic stage and by establishing the comparability of cognitive structure across groups of interest at baseline.

Latent structure of the neuropsychological test battery

The results of the exploratory and subsequent confirmatory factor analyses indicated a reliable 5-factor solution that was both statistically sound and clinically meaningful. This solution explained 63.3% of the total variance among scales. The five latent factors (Figure 1) are readily interpretable within the context of clinical neuropsychology and included three factors known to be at higher risk for change in the earliest stages of AD (verbal learning & memory, working memory, speed & executive function), as well as abilities less vulnerable to early effects of this disease (verbal and visuo-spatial skills) (Backman et al., 2005). The composite reliability estimates of the 5-factor scale for the total sample and across groups were fairly acceptable ranging from 0.86 to 0.89. Collectively, the results suggested a relatively strong psychometric basis for using the factor structure in clinical samples that match the characteristics of this cohort.

The composition of the verbal, visuo-spatial, and working memory factors identified in our analyses parallels widely-recognized classifications of tests (Strauss, Sherman, & Spreen, 2006; Lezak et al., 2004) based on prior factor analytic research (e.g., the loading of verbal and performance subtests from the WASI onto verbal ability and visuo-spatial factors,

respectively). Tests that required verbal or psychomotor speed (Stroop Color Word and Trail Making) alone (Trails A) or in combination with response inhibition (Stroop Color Word) or attentional switching (Trails B), loaded on a common factor that is probably measuring attention and aspects of executive function, but may also be sensitive to individual differences in speed per se.

While not an intent of this investigation, the results do bear an interesting relationship to classic attempts to characterize the taxonomy of human cognition (cf, McGrew, 2009). The factors identified here reflect cognitive processes that have been reliably identified in prior broader factor analytic studies (Cattell-Horn-Carroll), pointing out the potential relevance of that literature to the efforts represented in this investigation. Some components of the Cattell-Horn-Carroll models will of course be more important than others depending on the disease process under investigation (e.g., AD), but a fuller characterization of the comparative impact of various disease processes on empirically determined characterizations of cognition would be valuable.

Learning trials 3 through 5 and delayed recall of the Rey Auditory Verbal Learning Test formed a separate factor, providing a highly reliable measure of secondary, episodic memory. This finding concurs with prior studies in normal elderly controls (Siedlecki, Honig, & Stern, 2008; Delis, Kramer, Kaplan, & Ober, 2000), but contrasts with findings for patients with questionable dementia or probable AD, where delayed recall often forms a separate factor. The differential factor structure of memory tests in persons with AD most likely reflects the mesial temporal pathology that undermines the ability to store new information. The failure to find a separate delayed memory factor in our study suggests that mesial temporal function of most WRAP participants was qualitatively intact at study entry. The failure of AVLTL trials 1 and 2 to load on the learning and memory factor is explained by the fact that initial attempts to learn a supraspan list, especially trial 1, depend primarily on short-term or immediate memory span (Lezak et al., 2004; Mitrushina, Boone, Razani, & D'Elia, 2005) rather than on secondary memory processes.

Invariance of the latent structure

Another major finding was that the factorial structure of the 5-factor solution identified by the latent variable models was invariant across different groups defined by gender, age, parental family history of AD, and APOE genotype. It is important to note that establishing invariance for at least two items (scales) per underlying construct is a sufficient condition for meaningful comparisons across groups (see, for example, Byrne et al. 1989; Meredith, 1993; Steenkamp, 1998). Therefore, as a result of the full and partial scalar test, observed variable means can also be significantly compared across groups of interest. This is an important finding, assuring us that observed differences in test scores among subgroups of interest represent true differences that cannot be attributed to biases caused by the nonequivalence of constructs. Error invariance further guarantees that tests scales are equally reliable across groups. The invariance of a factor structure solution is rarely considered in investigations such as these, and it eliminates a potential source of misinterpretation of results in subgroup comparisons. We are not aware of any similar analyses in studies of preclinical AD.

Our analyses of invariance suggest that the neuropsychological battery used in WRAP is tapping comparable cognitive processes across key demographic and genetic risk groups at the point of study entry. That is, the same basic cognitive structures appear to be present at midlife for persons with and without a parental family history of AD and for APOE $\epsilon 4$ carriers and non-carriers, and also apply across selected demographic factors (age, gender) that are known to have significant quantitative effects on test performance—but as shown here, not on the latent structure (i.e., qualitative aspects) of test performance. This is a

critical point as it implies that the latent solution reported here is widely applicable in the broad age span of midlife. For WRAP and other prospective studies of AD, establishing the comparability of key comparison groups at midlife on basic cognitive architecture will make any future changes that emerge as a function of family history or other risk factors easier to interpret.

A few research groups using related, but different, approaches to compare the latent structure of cognitive performance across clinical groups, including normal aging and dementia, have obtained mixed results concerning the validity of a single factor structure (e.g., Johnson, Storandt, Morris, Langford, & Galvin, 2008; Siedlecki et al., 2008; Jones & Ayers, 2006; Kane, Balota, Storandt, McKeel, & Morris, 1998). For example, Siedlecki et al. (2008) identified differing factor structures in the memory performance of cognitively-normal controls compared to clinical samples with questionable dementia or probable AD. This suggested that the same memory tests may be measuring qualitatively different constructs in normally-aging persons compared to those with dementia. In such cases, quantitative group comparisons may be difficult to interpret and may fail to detect underlying differences in neurocognitive processes. Conversely, Johnson, et al. (2008) found that a 3-factor hybrid model of general and specific cognitive domains (attention, executive function, and working memory) had configural invariance across individuals with and without dementia. This finding provided support for using a common factor structure, and therefore, a common neuropsychological test battery, to study clinical and cognitive performance across diverse clinical samples. However, the battery used by Johnson et al. did not include secondary memory tests. Investigations that contrast cognitive structure in normal aging vs. AD are a source of hypotheses with regard to the cognitive changes most likely to emerge in preclinical stages of the disease. Prior studies suggest that prodromal changes in secondary memory can be detected years or even decades prior to clinical AD (Elias et al., 2000; Kawas et al., 2003), and if so, such memory change could manifest itself not only in quantitative differences across preclinical groups, but in changes in underlying cognitive structure as well. As a result, it will be important to evaluate the stability of the latent structure over time as our study population ages and as the development of MCI and AD becomes more likely.

Strengths and limitations

A limitation of these analyses is that the observed comparability of latent cognitive structures across key demographic and genetic subgroups may only hold true for traditional clinical summary scores derived from the tests included in our battery or similar batteries. We have previously shown differences in serial position profiles on the AVLT for WRAP participants with a family history of AD compared to controls (La Rue et al., 2008), and others have found differences by APOE genotype on sensitive experimental cognitive assays in middle-aged asymptomatic samples (e.g., Negash et al., 2008). Further studies should also focus on applying confirmatory factor analytic techniques to test competing models both within single populations and across populations.

There are other limitations related to the study sample. Although sample sizes were relatively large, this is a convenience sample recruited throughout the state of Wisconsin and to a lesser degree throughout the Midwest. While there is diversity in the distribution of urban versus rural residents and a reasonable range of intellectual ability, participants are generally highly educated, with above-average socioeconomic status, and almost exclusively non-Hispanic Caucasian. As a result, our findings may not generalize to other study populations. Future research to test the psychometric properties of neuropsychological test batteries should include more diverse samples and incorporate factorial invariance tests across racial and ethnic groups. In addition, although the overall fit of the CFA models was reasonable, the sample size of some of the subgroups (e.g., the AD FH-ve & APOE4-ve

group) in the invariance tests was below 200. It is possible that this might have affected factor loading estimates and, therefore, the stability of multi-group comparisons. However, most of the factor loadings across factors were relatively high indicating reasonably stable factors. Finally, there are additional variables of interest that we did not include in our factorial invariance comparisons, such as education. We will be analyzing education effects in WRAP in conjunction with other experiential factors (e.g., occupation and continued mental activity) that may influence midlife cognitive performance and eventual AD risk.

A strength of WRAP is its focus on an important at-risk group, i.e., AD children. Odds of developing AD are increased in offspring of parents with AD (Cupples et al., 2004; cf. Jarvik et al., 2008), and asymptomatic AD children may exhibit unique patterns on biomarkers by the time they are middle-aged. Strong family history effects, independent of APOE $\epsilon 4$ genotype, were found in fMRI activation patterns of AD children from the WRAP sample (Johnson, Schmitz, Moritz, et al., 2006a) and independently among asymptomatic children of autopsy-confirmed AD patients (Bassett et al., 2006). We have also found distinctive serial position patterns on verbal list learning for AD children compared to controls, suggesting less reliance on hippocampally-based memory functions in this genetically at-risk group (La Rue et al., 2008). Our study and a handful of others now focusing on AD relatives will eventually enable a better understanding of the interplay between genetic risk and other factors in determining the development of AD and may aid in identifying groups most likely to benefit from preventive interventions.

CONCLUSION

These analyses address often ignored but critical considerations in research employing extensive neuropsychological test batteries. The ability to define the latent structure minimizes redundancy, increases the reliability of measures, and reduces the risk of Type 1 error due to multiple comparisons. In addition, these analyses illustrate the importance of evaluating the invariance of a derived latent structure in subgroups defined by variables of interest. This is particularly important in prospective studies such as WRAP, in which a large number of at risk subjects are evaluated in an attempt to identify the earliest cognitive changes that may characterize pre-clinical Alzheimer's disease. While the reported latent structure is extremely sound in the age range represented here—the degree to which it will hold over long prospective time periods (e.g., decades) remains to be determined. We suspect that if a substantial proportion of persons developed AD or other neurological disorders (e.g., stroke), the underlying latent trait structure would change. This will be a topic of interest as we follow this cohort prospectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Alwin, DE.; Jackson, DJ. Applications of simultaneous factor analysis to issues of factorial invariance. In: Jackson, DD.; Borgatta, EP., editors. Factor analysis and measurement in sociological research: A multidimensional perspective. Beverly Hills, CA: Sage; 1981. p. 249-280.
- Alzheimer's Association. 2009 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* 2009;5:234–270.
- Backman L, Jones S, Berger AK, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology* 2005;19:520–531. [PubMed: 16060827]
- Bartrés-Faz D, Junqué C, Moral P, López-Alomar A, Sánchez-Aldeguer J, Clemente I. Apolipoprotein E gender effects on cognitive performance in age-associated memory impairment. *Journal of Neuropsychiatry and Clinical Neurosciences* 2002;14:80–83. [PubMed: 11884660]

- Bassett SS, Yousem DM, Cristinzio C, Kusevic I, Yassa MA, Caffo BS, et al. Familial risk for Alzheimer's disease alters fMRI activation patterns. *Brain* 2006;129:1229–1239. [PubMed: 16627465]
- Bentler PM. Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology* 1980;31:419–456.
- Benton AL. Neuropsychological assessment. *Annual Review of Psychology* 1994;45:1–23.
- Benton, AL.; Hamsher, K.; Sivan, AB. *Multilingual Aphasia Examination*. Iowa City, IA: AJA Associates; 1976.
- Bleecker ML, Bolla-Wilson K, Agnew J, Meyers DA. Age-related sex differences in verbal memory. *Journal of Clinical Psychology* 1988;44:403–411. [PubMed: 3384968]
- Bollen, KA. *Structural equations with latent variables*. New York: Wiley; 1989.
- Bookheimer S, Burggren A. APOE-4 genotype and neurophysiological vulnerability to Alzheimer's and cognitive aging. *Annual Review of Clinical Psychology* 2009;5:343–362.
- Bozdogan H. Model selection and Akaike's information criteria (AIC): The general theory and its analytical extensions. *Psychometrika* 1987;52:345–370.
- Brown TA. Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy* 2003;41:1411–1426. [PubMed: 14583411]
- Byrne B. The Maslach Burnout Inventory: validating factorial structure and invariance across intermediates, secondary, and university educators. *Multivariate Behavioral Research* 1991;26:583–605.
- Byrne B, Shavelson RJ, Muthén B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin* 1989;106:456–466.
- Cahn-Weiner D, Sullivan E, Shear P, Fama R, Lim K, Tinklenberg J, et al. Brain structural and cognitive correlates of clock drawing performance in Alzheimer's disease. *Journal of the International Neuropsychological Society* 1999;5:502–509. [PubMed: 10561930]
- Cudeck R, Browne MW. Cross-validation of covariance structures. *Multivariate Behavioral Research* 1983;18:147–167.
- Cupples LA, Farrer LA, Sadovnick AD, Relkin N, Whitehouse P, Green RC. Estimating risk curves for first-degree relatives of patients with Alzheimer's disease: the REVEAL study. *Genetics in Medicine* 2004;6:192–196. [PubMed: 15266206]
- Drasgow F. Scrutinizing psychological tests: Measurement equivalence and equivalence relations with external variables are central issues. *Psychological Bulletin* 1984;95:134–135.
- Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of Alzheimer disease: a 22-year prospective study of the Framingham Cohort. *Archives of Neurology* 2000;57:808–813. [PubMed: 10867777]
- Ertekin-Taner N, Younkin LH, Yager DM, Parfitt F, Baker MC, Asthana S, et al. Plasma amyloid β protein is elevated in late-onset Alzheimer disease families. *Neurology* 2008;70:596–606. [PubMed: 17914065]
- Etnier JL, Caselli RJ, Reiman EM, Alexander GE, Sibley BA, Tessier D, et al. Cognitive performance in older women relative to ApoE-epsilon4 genotype and aerobic fitness. *Medicine and Science in Sports and Exercise* 2007;39:199–207. [PubMed: 17218903]
- Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 1981;18:39–50.
- Geffen G, Moar KJ, O'Hanlon AP, Clark CR, Geffen LB. Performance measures of 16- to 86-year-old males and females in the Auditory Verbal Learning Test. *Clinical Neuropsychologist* 1990;4:45–63.
- Gorsuch, RL. *Factor analysis*. 2. Hillsdale, NJ: Erlbaum; 1983.
- Grice JW. Computing and evaluating factor scores. *Psychological Methods* 2001;6:430–450. [PubMed: 11778682]
- Hofer SM, Christensen H, Mackinnon AJ, Korten AE, Jorm AF, Henderson AS, et al. Change in cognitive functioning associated with ApoE genotype in a community sample of older adults. *Psychology and Aging* 2002;17:194–208. [PubMed: 12061406]

- Holdnack JA, Delis DC. Parsing the recognition memory components of the WMS-III face memory subtest: Normative data and clinical findings in dementia groups. *Journal of Clinical and Experimental Neuropsychology* 2004;26:459–483. [PubMed: 15512935]
- Horn JL, McArdle J. A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research* 1992;18:117–144. [PubMed: 1459160]
- Hu, L.; Bentler, P. Evaluating model fit. In: Hoyle, RH., editor. *Structural equation modeling. Concepts, issues, and applications*. London: Sage; 1995. p. 76-99.
- Hyman BT, Gomez-Isla T, Briggs M, Chung H, Nichols S, Kohout F, et al. Apolipoprotein E and cognitive change in an elderly population. *Annals of Neurology* 1996;40:55–66. [PubMed: 8687193]
- Jarvik L, La Rue A, Blacker D, Gatz M, Kawas C, McArdle JJ, et al. Children of persons with Alzheimer's disease: What does the future hold? *Alzheimer Disease & Associated Disorders* 2008;22:6–20. [PubMed: 18317242]
- Johnson DK, Storandt M, Morris JC, Langford ZD, Galvin JE. Cognitive profiles in dementia Alzheimer disease vs healthy brain aging. *Neurology* 2008;71:1783–1789. [PubMed: 19029518]
- Johnson SC, Schmitz TW, Moritz M, Meyerand H, Rowley A, Alexander K, et al. Activation of brain regions vulnerable to Alzheimer's disease: The effect of mild cognitive impairment. *Neurobiology of Aging* 2006a;27:1604–1612. [PubMed: 16226349]
- Johnson SC, Schmitz TW, Trivedi MA, Ries ML, Torgerson BM, Carlsson CM, et al. The influence of AD family history and APOE4 on mesial temporal lobe activation. *Journal of Neuroscience* 2006b;26:6069–6076. [PubMed: 16738250]
- Jones SN, Ayers CR. Psychometric properties and factor structure of an expanded CERAD neuropsychological battery in an elderly VA sample. *Archives of Clinical Neuropsychology* 2006;21:359–365. [PubMed: 16766158]
- Jöreskog KG. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 1969;34:183–202.
- Jöreskog KG. Simultaneous factor analysis in several populations. *Psychometrika* 1971;36:409–426.
- Jöreskog, KG.; Sörbom, D. LISREL 8.8 [Computer software]. Chicago: Scientific Software International; 2007.
- Kanne SM, Balota DA, Storandt M, McKeel DW, Morris JC. Relating anatomy to function in Alzheimer's disease: neuropsychological profiles predict regional neuropathology 5 years later. *Neurology* 1998;50:979–985. [PubMed: 9566382]
- Kaplan, E.; Goodglass, H.; Weintraub, S. *The Boston Naming Test*. 2. Philadelphia: Lea & Febiger; 1983.
- Kaplan, D. *Structural Equation Modeling: Foundations and Extensions*. 2. Newbury Park, CA: SAGE Publications; 2009.
- Kawas CH, Corrada MM, Brookmeyer R, Morrison A, Resnick SM, Zonderman AB, et al. Visual memory predicts Alzheimer's disease more than a decade before diagnosis. *Neurology* 2003;60:1089–1093. [PubMed: 12682311]
- Kline, RB. *Principles and practices of structural equation modeling*. 2. New York: Guilford Publications, Inc; 2005.
- Kramer JH, Delis DC, Daniel M. Sex differences in verbal learning. *Journal of Clinical Psychology* 1988;44:907–915.
- La Rue A, Hermann BL, Jones J, Johnson SC, Asthana S, Sager M. Effect of parental family history of Alzheimer's disease on serial position profiles. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association* 2008;4:285–290.
- Lezak, MD.; Howieson, DB.; Loring, DW. *Neuropsychological assessment*. 4. London: Oxford University Press; 2004.
- Loehlin, JC. *Latent variable models: An introduction to factor, path, and structural equation analysis*. 4. Mahwah, NJ: Lawrence Erlbaum; 2004.
- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 1996;1:130–149.

- McGrew KS. CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 2009;37:1–10.
- Marsh HW, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin* 1985;97:562–582.
- McKhann G, Drachman D, Folstein M, 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;34:939–944. [PubMed: 6610841]
- Meredith W. Measurement invariance, factor analysis and factor invariance. *Psychometrika* 1993;58:675–680.
- Millis SR, Malina AC, Bowers DA, Ricker JH. Confirmatory factor analysis of the Wechsler Memory Scale–III. *Journal of Clinical and Experimental Neuropsychology* 1999;21:87–93. [PubMed: 10421004]
- Mitrushina, M.; Boone, KB.; Razani, J.; D'Elia, LF. *Handbook of Normative Data for Neuropsychological Assessment*. 2. NY: Oxford University Press; 2005.
- Muthén, L.K.; Muthén, B.O. *Mplus 5.2* [Computer software]. Los Angeles: Author; 1998–2009.
- Negash S, Greenwood PM, Sunderland T, Parasuraman R, Deda YE, Knopman DS, et al. The influence of apolipoprotein E genotype on visuospatial attention dissipates after age 80. *Neuropsychology* 2008;23:81–89. [PubMed: 19210035]
- Petersen, RC. Conceptual issues. In: Petersen, RC., editor. *Mild Cognitive Impairment: Aging to Alzheimer's Disease*. New York, NY: Oxford University Press; 2003. p. 1–14.
- Reitan, R.M.; Wolfson, D. *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation*. 2. Tucson: Neuropsychology Press; 1993.
- Raykov T. Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement* 1998;22:375–385.
- Raykov T. Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology* 2001;54:315–323. [PubMed: 11817096]
- Revell, A.J.; Schaie, K.W. Domain-specific cognitive deficits associated with ApoE genotype by age group in community-dwelling elderly. Poster presented at the biennial Cognitive Aging Conference; Atlanta, GA. New York: Oxford University Press; 2004 Apr.
- Sager MA, Hermann BL, La Rue A. Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. *Journal of Geriatric Psychiatry and Neurology* 2005;18:245–249. [PubMed: 16306248]
- Saris, W.E.; Satorra, A.; Sörbom, D. The detection and correction of specification errors in structural equation models. In: Clogg, C.C., editor. *Sociological Methodology*. San Francisco: Jossey-Bass; 1987. p. 105–129.
- Sato M. Pragmatic treatment of improper solutions in factor analysis. *Annals of the Institute of Statistical Mathematics* 1987;39:443–455.
- Satorra, A.; Bentler, P.M. Proceedings of the Business and Economic Statistics Section of the American Statistical Association. 1988. Scaling Corrections for Chi-square Statistics in Covariance Structure Analysis; p. 308–313.
- Satorra, A.; Bentler, P.M. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye, A.; Clogg, C.C., editors. *Latent variable analysis: Applications for developmental research*. Thousand Oaks, CA: Sage; 1994. p. 399–419.
- Schaie K.W. Developmental influences on adult intelligence: The Seattle Longitudinal Study. 2005
- Schumacker, R.E.; Lomax, R.G. *A guide to SEM*. New Jersey: Lawrence Erlbaum; 1996.
- 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;22:400–411. [PubMed: 18444718]
- Small GW, Chen ST, Komo S, Ercoli L, Bookheimer S, Miller K, et al. Memory self-appraisal in middle-aged and older adults with the apolipoprotein E-4 allele. *American Journal of Psychiatry* 1999;156:1035–1038. [PubMed: 10401448]

- Sörbom D. A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology* 1974;27:229–239.
- Spreen, O.; Strauss, E. *A Compendium of Neuropsychological Tests*. 2. New York: Oxford University Press; 1998.
- Steenkamp JE, Baumgartner H. Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research* 1998;25:78–90.
- Strauss, EM.; Sherman, EM.; Spreen, O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3. Oxford: Oxford University Press; 2006.
- Tabachnick, BG.; Fidell, LS. *Using Multivariate Statistics*. Boston: Allyn and Bacon; 2001.
- Trenerry, M.; Crosson, B.; DeBoe, J.; Leber, L. *Stroop neuropsychological screening test*. Odessa, FL: Psychological Assessment Resource; 1989.
- Tuokko, HA.; Hultsch, DF. *Mild Cognitive Impairment: International Perspectives*. New York, NY: Psychology Press; 2006.
- van Driel OP. On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika* 1978;43:225–243.
- Vittadini G. Indeterminacy problems in the LISREL model. *Multivariate Behavioral Research* 1989;24:397–414.
- Wechsler, D. *Wechsler Abbreviated Intelligence Scale*. San Antonio: The Psychological Corporation; 1999.
- Wechsler, D. *Wechsler Adult Intelligence Scale*. 3. San Antonio: The Psychological Corporation; 1997a.
- Wechsler, D. *Wechsler Memory Scale*. 3. San Antonio: The Psychological Corporation; 1997b.
- Wei LJ, Lachin JM. Properties of the urn randomization in clinical trials. *Controlled Clinical Trials* 1988;9:345–364. [PubMed: 3203525]
- Widaman, KF.; Reise, SP. Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In: Bryant, KJ.; Windle, M.; West, SG., editors. *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research*. American Psychological Association; Washington DC: 1997.
- Wilkinson, G. *WRAT3 Administrative Manual*. Delaware: Wide Range; 1993.
- Yuan, KH.; Bentler, PM. Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. In: Sobel, ME.; Becker, MP., editors. *Sociological Methodology*. Washington, D.C: ASA; 2000. p. 165-200.
- Zumbo, BD.; Sireci, SG.; Hambleton, RK. Re-Visiting Exploratory Methods for Construct Comparability: Is There Something to be Gained from the Ways of Old?. Paper presented at NCME; April 2003; Chicago, Illinois. 2003.

Figure 1. Conceptual model for the confirmatory factor analysis. Factor correlations correspond to the full sample (N= 650).

Table 1**Summary of Neuropsychological Tests Grouped by Conceptual Domain**

Verbal ability and language	
•	Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999)
	Vocabulary
	Similarities
•	<i>Wide Range Achievement Test (WRAT-III; Wilkinson, 1993)</i>
	Reading subtest
•	<i>Boston Naming Test</i> (Kaplan, Goodglass, & Weintraub, 1983)
•	<i>Multilingual Aphasia Examination</i> (Benton, Hamsher, & Sivan, 1976)
	Controlled Oral Word Association (Lexical Fluency)
Visual spatial ability	
•	<i>Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999)</i>
	Block design
	Matrix Reasoning
•	<i>Benton Judgment of Line Orientation (JLO; Benton, 1994)</i>
•	<i>Clock Drawing Test (CDT; Cahn-Weiner, Sullivan, Shear, Fama, Lim, Yesavage, Tinkleberg, & Pfefferbaum, 1999)</i>
Working, visual memory, & executive function	
•	<i>Wechsler Adult Intelligence Scale—3rd edition (WAIS-III; Wechsler, 1997a)</i>
	Digit Span
	Arithmetic
	Letter-Number Sequencing
•	<i>Stroop's Naming Colored Words (NCW; Trenerry, Crosson, DeBoe, & Leber 1989)</i>
•	<i>Trail Making Tests</i> (Reitan & Wolfson, 1993)
	Trails A Time
	Trails B Time
Episodic memory	
•	<i>Wechsler Memory Scale-III (WMS-III; Wechsler, 1997b)</i>
	Faces I - Immediate
	Faces II - Delay
•	<i>Rey Auditory Verbal Learning Test (AVLT; Spreen & Strauss, 1998)</i>
	Trials 1 to 5
	Delayed Recall

Table 2

Comparison of Analytical Groups by Stratifying Variable

Stratifying Variables	EFA (N=638) ¹	CFA (N=650) ²	<i>t</i>	<i>p-value</i> ³
Age (years)				
<i>M</i>	53.18	53.8	-1.93	0.06
<i>SD</i>	6.82	6.47		
<i>Range</i>	36–67	36–67		
WASI ⁴ (full scaled score)				
<i>M</i>	112.82	112.88	-0.12	0.96
<i>SD</i>	8.96	9.86		
<i>Range</i>	81–140	72–140		
Education (years)				
<i>M</i>	16.06	16.12	0.43	0.67
<i>SD</i>	2.77	2.64		
<i>Range</i>	11–24	12–24		
Male				
<i>N</i>	184	200	-0.76	0.45
<i>%</i>	29	31		
APOE ε4				
<i>N</i>	240	214	1.76	0.08
<i>%</i>	38	33		
AD Family History				
<i>N</i>	473	469	0.80	0.42
<i>%</i>	74	72		

¹ Exploratory factor analysis.² Confirmatory factor analysis.³ All *p*-values were compared to a Bonferroni-adjusted nominal $\alpha = 0.0083$.⁴ Wechsler Abbreviated Scale of Intelligence.

Table 3

Factor Loadings for the Common Exploratory Factor Analysis and the Exploratory Factor Analysis Conducted within the Confirmatory Factor Analysis Framework after Promax Rotation

Scale	Verbal Ability			Visuo-Spatial Ability			Speed & Executive Function			Working Memory			Verbal Learning & Memory		
	EFA	EFA/CFA	EFA	EFA	EFA/CFA	EFA	EFA	EFA/CFA	EFA	EFA	EFA/CFA	EFA	EFA	EFA/CFA	EFA/CFA
WASI-Vocabulary	0.89		0.80												
WRAT-Reading	0.74		0.66												
WASI-Similarities	0.70		0.69												
Boston Naming	0.50		0.48												
Fluency	0.33		0.34												
WASI-Block Design				0.74		0.73									
Benton JLO				0.65		0.63									
WASI-Matrix Reasoning				0.50		0.53									
Stroop CW						0.53		0.56							
Trail Making A						-0.64		-0.62							
Trail Making B						-0.79		-0.78							
WAIS-Digit Span-Forward									0.75		0.83				
WAIS -Digit Span-Backward									0.76		0.84				
WAIS -Letter-number sequencing									0.65		0.67				
Rey AVLT Trial 3												0.82		0.93	
Rey AVLT Trial 4												0.86		0.90	
Rey AVLT Trial 5												0.84		0.94	
Delayed Recall												0.83		0.93	
Factor Determinacies	0.93	0.92	0.87	0.87	0.86	0.88	0.87	0.91	0.90	0.91	0.95	0.96	0.96	0.96	0.96

Note: Robust maximum likelihood was used as the extraction method. The five-factor solution accounted for **63.3%** of the total variance among the tests.

Latent Structure of the 18 Cognitive Scales: Confirmatory Factor Analysis Solution Using the Cross-Validation Sample (N =650) and Subgroups of Interest within the Same Sample

Table 4

Scale	Factor Loadings																
	Verbal Ability							Visuo-Spatial Ability							Speed & Executive Function		
	All	Male	Fem	Young	Old	AD+	AD-	All	Male	Fem	Young	Old	AD+	AD-	AD+	AD-	AD-
WASI -Vocabulary	0.88	0.89	0.87	0.84	0.90	0.88	0.86	0.88	0.87								
WRAT-Reading	0.79	0.83	0.76	0.79	0.79	0.78	0.79	0.80	0.76								
WASI -Similarities	0.71	0.77	0.68	0.65	0.76	0.71	0.71	0.60	0.77								
Boston Naming	0.54	0.53	0.54	0.57	0.51	0.55	0.48	0.53	0.44								
Fluency	0.49	0.56	0.46	0.48	0.51	0.47	0.55	0.48	0.55								
WASI -Block Design																	
Benton JLO																	
WASI -Matrix Reasoning																	
Stroop CW																	
Trail Making A																	
Trail Making B																	
WAIS -Digit Span-Forward																	
WAIS -Digit Span-Backward																	
WAIS -Letter-number sequencing																	
Rey AVLT Trial 3																	
Rey AVLT Trial 4																	
Rey AVLT Trial 5																	
Delayed Recall																	
Factor Determinacies	0.95	0.95	0.93	0.93	0.95	0.94	0.93	0.94	0.94	0.90	0.90	0.89	0.91	0.87	0.91	0.88	0.88
Factor Loadings																	
Working Memory																	
Verbal Learning & Memory																	
All	Male	Fem	Young	Old	AD+	AD-	AD+	AD-	AD+	AD-	AD+	AD-	AD+	AD-	AD+	AD-	AD-
WASI -Vocabulary																	

Scale	Factor Loadings												
	Working Memory							Verbal Learning & Memory					
	All	Male	Fem	Young	Old	AD+	AD-	AD+ APOE+	AD- APOE+	AD+ APOE+	AD- APOE+	AD+ APOE-	AD- APOE-
WRAT-Reading													
WASI -Similarities													
Boston Naming													
Fluency													
WASI -Block Design													
Benton JLO													
WASI -Matrix Reasoning													
Stroop CW													
Trail Making A													
Trail Making B													
WAIS -Digit Span-Forward	0.70	0.70	0.70	0.75	0.64	0.73	0.71	0.67	0.67				
WAIS -Digit Span-Backward	0.74	0.76	0.72	0.76	0.73	0.71	0.76	0.77	0.75				
WAIS -Letter-number sequencing	0.71	0.76	0.68	0.69	0.73	0.72	0.68	0.77	0.65				
Rey AVLT Trial 3								0.83	0.75	0.77	0.80	0.79	0.80
Rey AVLT Trial 4								0.89	0.87	0.87	0.89	0.88	0.88
Rey AVLT Trial 5								0.91	0.85	0.87	0.88	0.87	0.88
Delayed Recall								0.82	0.80	0.80	0.83	0.84	0.85
Factor Determinacies	0.90	0.91	0.88	0.90	0.90	0.880.90	0.89	0.90	0.88	0.96	0.96	0.95	0.96

Table 5

Summary of Tests of Factorial Invariance by Groups

	<i>Composite Reliability</i>	χ^2	<i>df</i>	χ^2/df	$\Delta\chi^2$	Δdf	<i>p-value</i>	<i>RMSEA (90% CI)</i>	<i>CFI</i>	<i>NFI</i>	<i>NNFI</i>	<i>CAIC</i>
Single Group Model												
AD+ (n=469)	0.89	362.48**	125	2.90				0.06 (0.056, 0.071)	0.97	0.95	0.96	
AD- (n=181)	0.87	206.36**	125	1.65				0.06 (0.045, 0.074)	0.96	0.91	0.96	
Invariance												
Full Configural		568.36**	250	2.27				0.06 (0.055, 0.069)	0.97	0.94	0.96	1525.41
Full Metric		577.60**	268	2.16	9.24	18	0.95	0.06 (0.053, 0.066)	0.97	0.94	0.96	1400.07
Full Scalar		589.22**	281	2.10	11.62	13	0.56	0.06 (0.051, 0.065)	0.97	0.94	0.96	1314.49
Full Factor Variance-Covariance		599.97**	291	2.06	10.75	10	0.38	0.06 (0.051, 0.064)	0.97	0.94	0.97	1250.46
Full Strict (Error Variance)		620.20**	309	2.01	20.23	18	0.32	0.06 (0.049, 0.062)	0.97	0.94	0.97	1136.11
Single Group Model												
Younger (n=330)	0.88	320.61**	125	2.56				0.07 (0.057, 0.078)	0.96	0.93	0.95	
Older (n=320)	0.88	216.94**	125	1.74				0.05 (0.037, 0.058)	0.98	0.95	0.98	
Invariance												
Full Configural		536.41**	250	2.15				0.06 (0.052, 0.066)	0.97	0.94	0.96	1493.46
Full Metric		560.93**	268	2.09	24.52	18	0.14	0.06 (0.051, 0.065)	0.97	0.94	0.96	1383.40
Full Scalar		609.04**	281	2.17	48.11	13	0.001**	0.06 (0.053, 0.067)	0.96	0.94	0.96	1334.31
Partial Scalar Invariance		577.49**	278	2.08	16.56	10	0.08	0.06 (0.051, 0.064)	0.97	0.94	0.96	1325.19
Full Factor Variance-Covariance		595.66**	288	2.07	18.17	10	0.05	0.06 (0.051, 0.064)	0.97	0.94	0.96	1268.59
Full Strict (Error Variance)		607.93**	306	1.99	12.27	18	0.83	0.06 (0.049, 0.062)	0.97	0.94	0.97	1146.27
Single Group Model												
Females (n=450)	0.88	294.23**	125	2.35				0.05 (0.047, 0.063)	0.97	0.95	0.97	
Males (n=200)	0.89	219.71**	125	1.76				0.06 (0.048, 0.075)	0.97	0.94	0.97	
Invariance												
Full Configural		513.73**	250	2.05				0.06 (0.050, 0.064)	0.97	0.95	0.97	1201.61

	Composite Reliability	χ^2	df	χ^2/df	$\Delta \chi^2$	Δdf	p-value	RMSEA (90% CI)	CFI	NFI	NNFI	CAIC
Full Metric		538.28**	268	2.01	24.55	18	0.14	0.06 (0.049, 0.063)	0.97	0.95	0.97	1360.77
Full Scalar		659.98**	281	2.35	121.70	13	0.001**	0.06 (0.058, 0.071)	0.96	0.93	0.96	1385.25
Partial Scalar Invariance		550.56**	276	1.99	12.28	8	0.14	0.06 (0.049, 0.06)	0.97	0.94	0.97	1313.21
Full Factor Variance-Covariance		575.88**	286	2.01	25.32	10	0.005**	0.06 (0.049, 0.063)	0.97	0.94	0.97	1263.76
Partial Factor Variance Covariance		566.24**	285	1.99	15.68	9	0.07	0.06 (0.049, 0.062)	0.97	0.94	0.97	1261.60
Error Variance		590.57**	302	1.96	25.36	18	0.12	0.05 (0.048, 0.061)	0.97	0.94	0.97	1158.82
Single Group Model												
AD+ & APOE4+ (n=187)	0.88	230.03**	125	1.84				0.07 (0.053, 0.080)	0.96	0.92	0.95	
AD- & APOE4- (n=154)	0.86	199.21**	125	1.59				0.06 (0.045, 0.078)	0.96	0.90	0.95	
Invariance												
Full Configural		429.09**	250	1.72				0.07 (0.054, 0.075)	0.96	0.91	0.95	1303.57
Full Metric		436.83**	268	1.63	7.74	18	0.98	0.06 (0.050, 0.071)	0.96	0.91	0.96	1188.34
Full Scalar		457.84**	281	1.63	21.01	13	0.07	0.06 (0.050, 0.071)	0.96	0.90	0.96	1120.54
Full Factor Variance-Covariance		469.22**	291	1.61	11.38	10	0.32	0.06 (0.050, 0.070)	0.96	0.90	0.96	1063.60
Full Strict (Error Variance)		501.76**	309	1.62	32.54	18	0.020*	0.06 (0.051, 0.070)	0.96	0.90	0.96	973.16
Partial Error Variance		495.47**	308	1.61	26.25	17	0.07	0.06 (0.050, 0.070)	0.96	0.90	0.96	973.71

Notes.

** p < 0.01;

* p < 0.05;

p-values correspond to $\Delta \chi^2$.