Published in final edited form as:

J Clin Exp Neuropsychol. 2017 September ; 39(7): 621–635. doi:10.1080/13803395.2016.1252725.

Latent Structure of Cognitive Performance in the Adult Children Study

Denise Head^{1,2,4}, Samantha Allison¹, Nathaniel Lucena¹, Jason Hassenstab^{1,2,3}, and John C. Morris^{2,3}

¹Department of Psychology, Washington University in St. Louis

²Knight Alzheimer's Disease Research Center, Washington University in St. Louis

³Department of Neurology, Washington University in St. Louis

⁴Department of Radiology, Washington University in St. Louis

Abstract

Objective—The Adult Children Study (ACS) at the Knight Alzheimer's Disease Research Center is a longitudinal investigation designed to identify and validate potential biomarkers of preclinical Alzheimer's disease (AD) in cognitively normal individuals with and without a family history of AD. The purpose of the current study was to validate the proposed latent structure of the ACS psychometric battery.

Method—Confirmatory factor analyses of baseline data in a sample of 229 (75 men) cognitively normal middle-aged to older adult individuals assessed a hypothesized 4-factor model of cognitive performance. Measurement invariance was investigated as a function of family history of AD and APOE status.

Results—This study confirmed *a priori* hypotheses of 4 latent cognitive domains in a unique longitudinal sample of cognitively normal adults. In addition, there was evidence of a similar factor structure for family history and APOE status groups.

Conclusion—These robust indicators of a broad range of cognitive domains will be used in future investigations to track the influence of family history of AD on cognitive performance over time. In addition, associations with fluid, structural, and molecular biomarkers of preclinical AD will be further examined, both cross-sectionally and longitudinally in this sample.

Keywords

psychometric test battery; confirmatory factor analysis; measurement invariance; cognitive aging

Introduction

Accumulating evidence from recent research indicates that Alzheimer's disease (AD) has a significant prodromal period during which neuronal pathology accumulates with no outward

symptoms, eventually reaching a level of damage sufficient to lead to cognitive impairment associated with symptomatic AD. The Adult Children Study (ACS) began collecting longitudinal data from a cohort of cognitively normal, middle-aged to older adults in 2005 to study this pre-symptomatic (preclinical) stage of AD. The ACS cohort contains individuals with and without a confirmed family history of AD. The ACS dataset consists of a broad array of psychometric, clinical, and biomarker data, including information on MRI-based measures of brain structure, cerebrospinal fluid and PET-based biomarkers, and candidate susceptibility genes such as apolipoprotein E (APOE). Longitudinal changes in cognitive ability and biomarkers in this sample of cognitively normal individuals may be indicators of accumulating AD pathology in the preclinical phase of the disease. Furthermore, the sample allows comparison of longitudinal changes in biomarkers and cognitive performance between those with and without a family history of AD. A recent assessment of the ACS cohort indeed found that family history of AD predicted accelerated change in several biomarkers associated with AD (Xiong et al., 2011).

The ACS psychometric battery was specifically designed to measure four cognitive domains that are particularly relevant to aging and AD, including episodic memory (for a review, see: Salmon, 2000), semantic memory (e.g., Aschenbrenner et al., 2015; Hodges, Salmon, & Butters, 1992), executive function (e.g., Bangert & Balota, 2012; Hutchison, Balota, & Duchek, 2010; for a review, see: Perry & Hodges, 1999), and visuospatial ability (for a review, see: Cronin-Golomb & Amick, 2011). The selected domains are also considered among the most relevant to incorporate into a general purpose cognitive test battery, such as the NIH toolbox (Nowinski, Victorson, Debb & Gershon, 2013). Tasks tapping multiple domains were included as past research indicates that a wide range of measures and domains may be affected early in the course of the disease and in at-risk individuals (Bäckman, Jones, Berger, Laukka, & Small, 2005; Small, Rosnick, Fratiglioni, & Bäckman, 2004; Twamley, Ropacki, & Bondi, 2006). In addition, a multivariate approach may enhance the ability to detect cognitive changes (Chapman et al., 2010; Loewenstein et al., 2001; Ownby, Loewenstein, Schram, & Acevedo, 2004). As no single task provides an error-free estimate of performance in a given cognitive domain, the ACS battery involves multiple measures for each cognitive construct. More specifically, three measures for each putative domain were included in an attempt to capture cognitive function broadly and increase the reliability of our estimate of each domain, but at the same time keep the length of the battery reasonable. We also sought to avoid tests that might have ceiling effects in this cognitively normal, middle-aged sample (e.g., Boston Naming Test).

Summary composite scores for a given domain offer a more reliable estimate than any single task. Additionally, composite scores reduce the overall number of statistical comparisons, thus lowering the risk of Type 1 error. However, an important step prior to using composite measures is to establish the validity of the underlying hypothesized factor structure of the cognitive data. Numerous studies have examined the latent factor structure of psychometric batteries in older adult and AD samples using confirmatory factor analysis (CFA) (e.g., Davis, Massman, & Doody, 2003; Dowling, Hermann, La Rue, & Sager, 2010; Johnson, Storandt, Morris, Langford, & Galvin, 2008; Hayden et al., 2011; Loewenstein et al., 2001; Mitchell, Shaughnessy, Shirk, Yang, & Atri, 2012; Moleiro et al., 2013; Mungas, Widaman, Reed, & Tomaszewski Farias, 2011; Park et al., 2012; Robbins et al., 1994; Siedlecki,

Honig, & Stern, 2008). Many of these investigations were designed to assess a local center's test battery (Dowling et al., 2010; Johnson et al., 2008; Mitchell et al., 2012; Siedlecki et al., 2008), a multi-center battery (Hayden et al., 2011; Moleiro et al., 2013; Park et al., 2012), or a commonly used battery (Davis et al., 2003; Mungas et al., 2011; Robbins et al., 1994).

While the results from past studies have yielded factor structures ranging from three to six factors, they do generally converge in supporting the benefits of examining multiple factors to better understand the AD process (e.g., Johnson et al., 2008; Lowenstein et al., 2001; Ownby et al., 2004). The most common factors extracted across these studies include memory, verbal ability, visuospatial ability, executive function, attention, and processing speed. Although there is some overlap in the individual tasks used across studies, the number and the particular tasks do differ across studies. In general, a greater number of tasks or variables in CFA is associated with observing more latent factors. Furthermore, the investigations differ in sample size, the first language of participants, the disease status of the participants, and the methods used to screen for dementia. These differences highlight the need for confirming the latent factor structure of individual test batteries.

Thus, the primary aim of the current study was to confirm the *a priori* hypothesized latent factor structure of the psychometric battery using CFA in a group of well-characterized adults with extensive co-existing AD biomarker data. Specifically, we hypothesized that there would be four latent factors (episodic memory, semantic memory, executive function, and visuospatial ability), each with three indicator tasks. In post-hoc analyses, we also compared this hypothesized model to conceptually-driven three-factor and five-factor models.

As a particular goal of the ACS cohort is to investigate longitudinal changes in cognitive performance between groups (e.g., family history versus no family history of AD), another important step is to establish measurement invariance across groups. Measurement invariance refers to having a construct measured in a similar manner across groups or time. That is, assessment of measurement involves examining whether measurement properties are generalizable across groups or time. Determination of measurement invariance is important so that one can be confident that observed group differences at the construct level are not due to bias in the measurement of the construct for a particular group. Thus, in order to validly determine whether groups differ on a latent construct of interest, it is important to first establish that the constructs are being measured in a similar manner across groups. If they are not, then any observed group difference on the latent construct may actually be reflective of systematic bias in the estimation of the construct. For example, measurement bias would exist when one group performs more poorly on a particular scale due to such things as strategy or experience differences, even though the groups are equal on the underlying construct.

Some investigations have largely found measurement equivalence in comparisons between healthy older adults and symptomatic AD individuals (Chapman et al., 2010; Watts, Loskutova, Burns, & Johnson, 2013), between groups with and without a family history of AD (Dowling et al., 2010), and between groups with varying levels of cognitive impairment (Davis et al., 2003; Park et al., 2012). However, some studies have observed evidence of

measurement non-invariance in comparisons between healthy older adults and individuals with questionable dementia or symptomatic AD (e.g., Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003; Siedlecki et al., 2008). Given these findings, it is important to establish measurement invariance for the current psychometric battery to support the validity of comparisons across groups of interest. Thus, secondary aims were to test aspects of measurement invariance of the latent factor structure across relevant groups (i.e., individuals with and without a family history of AD; individuals with and without the presence of an APOE4 allele), and to assess whether any evidence of measurement bias observed was sufficiently substantive to invalidate group comparisons.

Methods

Participants

The current sample included data from the baseline assessment for 229 cognitively normal, community-dwelling participants from the ACS cohort (see Table 1 for demographic data). All clinical and cognitive assessments are obtained in 3-year increments after baseline until age 65 or older, after which participants are assessed annually. Eligibility criteria for inclusion in the ACS sample were baseline age 45 to 75, willingness to provide an informant who knew the participant well, Clinical Dementia Rating (CDR; (Morris, 1993)) of 0 indicating normal cognition, and willingness to participate in all study procedures. Comorbid conditions (e.g., depression) were allowed if the participant was considered stable at the time of baseline enrollment. Exclusion criteria included terminal conditions that would prevent longitudinal assessment (e.g., end-stage cancer), or membership in a family with either dominantly inherited AD or a known causal mutation for AD. Each participant was characterized as positive (n=140) or negative for family history. A positive family history was defined as at least 1 biological parent with age at onset for symptomatic AD of less than 80 years, and a negative family history was defined as both biological parents living to age 70 or longer without symptomatic AD. If a parent living to age 70 without symptomatic AD later developed symptomatic AD by age 80, the participant was reassigned to the positive family history group.

Approximately one-third of participants were recruited as children of parents who were enrolled in a pre-existing longitudinal study of aging and dementia at the Washington University Alzheimer's Disease Research Center. Approximately one-half of the participants had increased risk for developing AD due to the presence of an APOE4 allele. As expected, the proportion of individuals with an APOE4 allele was significantly higher in the group with family history than without family history ($X^2=22.12$, $p < 0.01$). The Washington University Human Research Protection Office approved all study procedures, and data were obtained in compliance with the Helsinki Declaration. Written consent was obtained from all participants prior to completing any study procedures.

Clinical Assessment

Presence or absence of dementia, as well as severity of dementia when present, were assessed using the CDR (Morris, 1993). The CDR is based on a rating by an experienced clinician (i.e., nurse clinicians, physicians). The rating uses information both from an

examination of the participant and from the participant's informant to judge whether an individual's performance in cognitive abilities and daily activities is at his or her previously attained level. The CDR evaluation is completed independently of psychometric test performance. The CDR is highly reliable (Burke et al., 1988; McCulla et al., 1989; Morris et al., 1997), and sensitive to very mild cognitive impairment resulting from AD (Morris et al., 1991; Morris et al., 2001; Storandt, Grant, Miller, & Morris, 2006).

APOE Genotyping

APOE genotyping was performed on DNA from peripheral blood sampling using previously described procedures (Talbot et al., 1994). Individuals were classified as APOE4+ (44, 34, 24) or APOE4- (33, 23, 22).

Cognitive Assessment

All participants completed a 2-hour psychometric testing session within 2 weeks of the clinical assessment. The cognitive battery included 12 measures across four domains: episodic memory, semantic memory, executive function, and visuospatial ability. The raw score for each task was used in the analyses, and all scores were coded so that higher numbers indicated better performance.

The episodic memory domain included tasks based on the three major methods used to assess memory: a) prose recall was assessed with the delayed subtest from the Wechsler Memory Scale-III Logical Memory task (split-half reliability, $r=0.75$) (Wechsler, 1997a), b) paired associates learning was assessed with the immediate subtest from the Wechsler Memory Scale-III Verbal Paired Associates task (split-half reliability, $r=0.92$) (Wechsler, 1997a), and c) list learning was assessed with the Free Recall subtest from the Free and Cued Selective Reminding Test (SRT Free Recall) (Cronbach's $\alpha=0.85$) (Grober, Buschke, Crystal, Bang, & Dresner, 1988; Grober, Ocepek-Welikson, & Teresi, 2009).

The measures of the semantic memory domain included a) the Information subtest from the Wechsler Adult Intelligence Scale-III (split-half reliability, $r=0.92$) (Wechsler, 1997b), b) the Similarities subtest from the Wechsler Adult Intelligence Scale-III (split-half reliability, $r=0.84$) (Wechsler, 1997b), and c) Animal Naming (test-retest reliability, $r=0.83$) (Goodglass & Kaplan, 1983; Solomon et al., 1998). All of these measures require retrieval of semantic information learned in the past.

For the visuospatial ability domain, both visuomotor and motor-free tasks were selected. Specifically, the measures for this domain included a) the Block Design subtest from the Wechsler Adult Intelligence Scale-III (split-half reliability, $r=0.89$) (Wechsler, 1997b), b) Benton Judgment of Line Orientation (test-retest reliability, $r=0.90$) (Benton, Hamser, Varney, & Spreen, 1983; Franzen, 2000), and c) the Visual Relations subtest from the Woodcock-Johnson-III (test-retest reliability, $r=0.81$) (Woodcock, McGrew, & Mather, 2001).

Measures for the executive function domain were selected to tap a range of skills included under this rubric, including working memory, attention-shifting, and cognitive flexibility. Specifically, the measures for this domain included a) the Letter-Number Sequencing subtest

from the Wechsler Memory Scale-III (split-half reliability, $r=0.79$) (Wechsler, 1997a), b) Auditory Consonant Trigrams (test-retest reliability= 0.71) (Mertens, Gagnon, Coulombe, & Messier, 2012; Peterson & Peterson, 1959), and c) Trail Making B minus Trail Making A (test-retest reliabilities, $r=0.89$ and $r=0.79$, respectively) (Dikmen et al., 1999). This difference score was used to better isolate the executive component of Trail Making B.

Alternative Models

Three competing models were also evaluated. These models were generated based on conceptual considerations. A three-factor model consisting of factors for memory (i.e., all six memory tasks), visuospatial ability, and executive functions was examined in consideration of the conceptualization that the data may be better modeled with one memory domain rather than two separate domains. There were also two five-factor models. Both of these models had the same episodic memory and visuospatial factors as the proposed four-factor model, and both had WAIS-III Similarities and WAIS-III Information as indicators of the semantic memory factor. For one of the five-factor models (Model 5A), Auditory Consonant Trigrams and Letter-Number Sequencing were indicators of a working memory factor. Both are conceptualized as working memory tasks, and past factor analytic work has observed both loading on the same factor (Mertens, Gagnon, Coulombe, & Messier, 2006). Animal Naming and Trail Making B-A were indicators of a speeded processing factor as both of these are timed tasks that can be considered to have an executive component (Mitrushina, Boone, Razani, & D'Elia, 2005; Strauss, Sherman, & Spreen, 2006). For the other five-factor model (Model 5B), we considered whether there might be separable working memory and attention factors. Thus, Animal Naming and Letter-Number Sequencing were indicators of a working memory factor, and Auditory Consonant Trigrams and Trail Making B-A were indicators of an attention factor. The basis for the working memory factor was evidence that verbal fluency measures have a working memory component (e.g., Unsworth, Spillers, & Brewer, 2011). In addition, Auditory Consonant Trigrams has been considered a divided attention task and Trail Making B-A as an attention switching task (Mitrushina, Boone, Razani, & D'Elia, 2005; Strauss, Sherman, & Spreen, 2006).

Statistical Analysis

A CFA of the ACS psychometric battery was performed to establish that our *a priori* factor structure adequately fit the data. Extraction of *a priori* latent factors was implemented in MPLUS 7.0 (Muthén & Muthén, 1998–2009) using the robust maximum likelihood estimator. This estimator was selected as most of the variables evidenced non-normal distributions (see Tables 2 and 3). MPLUS uses all data that are available to estimate models in data sets with missing data points. In the current sample, .007% of the 2,748 total data points were missing.

Model fit was evaluated using the following: χ^2 test for overall model, the comparative fit index (CFI; Bentler, 1990), the root mean square error of approximation (RMSEA; Steiger, 1990), and the standardized root mean square residual (SRMR; Bentler, 1995; Jöreskog & Sörbom, 1981). Hu and Bentler (1999) recommended CFI close to 0.95, RMSEA close to 0.06, and SRMR close to 0.08 as adequate fit. CFI values above 0.90 can be considered to

represent the lower bound (Brown, 2006; Hoyle, 2011; Schumaker & Lomax, 1996). Browne and Cudeck (1993) suggest that RMSEA values below .08 represent adequate fit, and values below 0.06 represent good fit. For the current report, the following values indicated acceptable model fit: X^2 with an associated p-value .05; CFI 0.90; RMSEA 0.08; SRMR 0.08.

The X^2 difference test (Satorra & Bentler, 1999) was used to compare the nested three-factor and four-factor models. In addition, the Akaike's information criterion (AIC; Akaike, 1987) and the Bayesian information criterion (BIC; Schwarz, 1978) were examined in comparing the four-factor model with the five-factor models, as well as for the three-factor and four-factor comparison. For both of these criteria, a smaller value is indicative of the preferred model.

A secondary goal of the current work was to assess for measurement invariance across groups with and without a family history of AD, and across groups with and without an APOE4 allele. This analysis examines whether the obtained measurement model generalizes and provides valid indicators of cognitive functioning across groups. Factorial invariance is often tested within a multiple-groups CFA framework using a multi-step approach (Steenkamp & Baumgartner, 2010), which includes assessment of multiple sources of invariance (i.e., configural, metric, scalar, factor variance-covariance, and error-variance invariances). However, multiple-group analysis in the CFA framework requires sufficiently large sample sizes per group to obtain reliable parameter estimates and test statistics. In instances where cell sizes may not be sufficiently large, a Multiple Indicators Multiple Cause (MIMIC) model may be a more practical method for assessing aspects of measurement invariance, although fewer parameters related to invariance can be estimated with MIMIC models versus multiple-groups CFA (Muthén, 1988). MIMIC models can test the factor structure with the covariates included in the model while also testing whether groups differ in mean levels of latent variables (i.e., population heterogeneity), and whether the intercepts for observed indicator variables (test scales) are the same across groups at a given level of a latent variable (i.e., measurement invariance). In terms of population heterogeneity, an association between the grouping variable (e.g., family history status) and a latent variable indicates that the groups differ in the latent cognitive construct, which influences responses to the observed indicator variables. Measurement non-invariance or bias exists when scores on a particular scale differ across groups, even though the groups are equated on the underlying construct. Given the relatively small sample size in some group-level cells, we chose to use the MIMIC approach to test for measurement bias across family history of AD and APOE status.

The procedure for using MIMIC models to assess for measurement bias followed the steps described by Muthén (1988). The base MIMIC model examined the effect of a covariate on the latent variables with the direct effect of the covariate on the intercepts of the indicator variables set to zero. This base MIMIC model represented a no-measurement bias model. Modification indices were examined to determine a direct effect that, if freely estimated in the model, would significantly improve model fit (i.e., test scales with measurement bias). In a subsequent model, the identified direct effect was allowed to be freely estimated (Direct Effects models) with all other paths remaining fixed at zero. Improvement of fit was

assessed using the X^2 difference test (Satorra & Bentler, 1999). If there was a significant improvement in fit, then the identified path was retained in the model. This was done in an iterative manner such that each additional direct effect identified by the modification indices would be added to the previous model. This iterative process continues until there are no longer significant improvements in model fit. Simulation studies suggest Type 1 error inflation when implementing the MIMIC procedure to assess for measurement invariance (Kim, Yoon, & Lee, 2012). Thus, the Oort adjustment to the critical X^2 value was adopted to control for Type 1 error inflation (Kim, Yoon, & Lee, 2012).

Results

Total Sample

For the *a priori* model all fit indices except the X^2 statistics were within acceptable ranges for good model fit ($X^2(48)=70.062$, $p=.021$; CFI=0.970; SRMR=0.044; RMSEA=0.045, 90% C.I.=0.018–0.066). To address model misspecification that may have contributed to the significant X^2 , we first sought to account for methodological contributions in order to minimize adjustments to the hypothesized factor structure. The Similarities and Information tasks are both subtests of the WAIS-III and are more methodologically similar compared to the third task for the Semantic Memory domain (i.e., Animal Naming). In addition, examination of the modification indices supported the addition of correlated measurement error. Thus, we added correlated errors for the Similarities and Information tasks of the Semantic Memory latent factor (see Figure 1). For the re-specified model, all fit indices were within acceptable ranges for good model fit ($X^2(47)=63.149$, $p=.058$; CFI=0.978; SRMR=0.044; RMSEA=0.039, 90% C.I.=0.000–0.062). Standardized loadings were significant across all indicators (range=0.49–0.80; see Figure 1), thus supporting construct validity of the *a priori* four-factor model. Intercorrelations between latent factors ranged from 0.49 to 0.80 (see Figure 1).

In post-hoc analyses, the final four-factor model with correlated errors was compared to the alternative models. It was observed that fit indices for the three-factor model with correlated errors for Similarities and Information were all within acceptable ranges except for the X^2 test ($X^2(50)=82.840$, $p=.002$; CFI=0.955; SRMR=0.050; RMSEA=0.054, 90% C.I.=0.032–0.074). Furthermore, there was a significant improvement in fit for the four-factor model (X^2 difference (3)=20.273, $p<.001$). The AIC and BIC values were also smaller for the four-factor model suggesting that this is the preferred model (Three-factor model: AIC=17097.089 and BIC=17234.438; Four-factor model: AIC=17083.441 and BIC=17231.091).

Fit indices for Model 5A were all within acceptable ranges ($X^2(44)=59.481$, $p=.060$; CFI=0.979; SRMR=0.037; RMSEA=0.039, 90% C.I.=0.000–0.063). However, Model 5A did not provide an improved fit compared to the four-factor model (Model 5A: AIC=17085.511 and BIC=17243.462; Four-factor model: AIC=17083.441 and BIC=17231.091). In terms of Model 5B, fit indices were all within acceptable ranges except for the X^2 test ($X^2(44)=65.153$, $p=.021$; CFI=0.955; SRMR=0.050; RMSEA=0.054, 90% C.I.=0.032–0.074). However, the correlation between the working memory and attention factors was above 1 indicating an inadmissible model.

MIMIC models

Family History—Fit indices for the base MIMIC model with Family History status added as a covariate for latent factors were all within acceptable ranges except for the X^2 test ($X^2(55)=80.531$, $p=.014$; CFI=0.966; SRMR=0.042; RMSEA=0.045, 90% C.I.=0.021–0.065). However, there was also evidence of measurement bias for WAIS-III Similarities and Trail Making Part B-A (see Figure 2). That is, there was a significant improvement in model fit ($X^2(54)=73.366$, $p=.041$; CFI=0.974; SRMR=0.039; RMSEA=0.040, 90% C.I.=0.009–0.061) with a direct effect of family history on the intercept for Similarities (standardized beta=0.155, $p=.007$) based on the Oort-adjusted critical chi-square value of 5.166 (X^2 difference (1)=7.229, $p<.05$). In addition, there was a significant improvement in model fit ($X^2(55)=67.310$, $p=.089$; CFI=0.981; SRMR=0.037; RMSEA=0.034, 90% C.I.=0.000–0.057) with an additional direct effect of family history on the intercept for Trail Making Part B-A (standardized beta=–0.147, $p=.008$) based on the Oort-adjusted critical chi-square value of 4.787 (X^2 difference (1)=8.099, $p<.05$). This finding indicated that those with a family history of AD had better scores on Similarities and lower scores on Trail Making Part B-A controlling for any group differences in semantic memory and executive function, respectively. Importantly, with these direct effects included, all fit indices were within acceptable ranges. No significant group differences between those with and without family history were observed on any latent factors in either the base model or the Direct Effects model (see Table 4, and Figure 2). Thus, family history does not appear to significantly influence mean levels of the latent factors.

APOE Status—Fit indices for the base MIMIC model with APOE status added as a covariate for latent factors were within acceptable ranges ($X^2(55)=71.193$, $p=.070$; CFI=0.978; SRMR=0.039; RMSEA=0.036, 90% C.I.=0.000–0.058). There was also evidence of measurement bias for the SRT Free Recall task. That is, with a direct effect of APOE status on the intercept of this test (standardized beta=–0.132, $p=.030$; see Figure 3), there was a significant improvement in model fit ($X^2(54)=66.517$, $p=.118$; CFI=0.983; SRMR=0.038; RMSEA=0.032, 90% C.I.=0.000–0.055) based on the Oort-adjusted chi-square value of 4.576 (X^2 difference (1)=4.996 $p<.05$). This finding suggests that those with an APOE4 allele had lower scores on this test controlling for any group differences in episodic memory. There was a significant effect of APOE status on the Executive Function latent factor for both the base model and the Direct Effects model (standardized beta=–0.159, $p=.045$ for both). No other significant APOE group differences were observed on the other latent factors in either the base model or the Direct Effects model (see Table 4, and Figure 3). Thus, APOE status appears to significantly influence mean levels of only the Executive Function latent factor.

Additional post-hoc analyses

Full factorial invariance with multi-group CFA was not conducted for the primary *a priori* analyses in consideration that there were relatively small sample sizes for some cells in analyses. The small sample sizes could lead to concerns about biased parameter estimates and associated standard errors, which can affect the power to detect a difference. However, considering the potential relevance of group differences in factor loadings in interpreting

differences in composite scores, a set of post-hoc analyses was conducted in an attempt to obtain some limited information about possible factor loading differences across groups.

Thus, we assessed for configural, metric, and scalar invariance using multi-group CFA, and followed this with Monte Carlo simulations of the models to estimate bias. This was only done for Family History as the sample size for the APOE4+ group was too small for modeling. The configural model examined whether the factor structure was the same across groups (i.e., the same pattern of fixed and free parameters). Next, factor loadings were constrained to be equal across groups to test for metric invariance. Lastly, we also examined a model with both factor loadings and intercepts constrained to be equal across groups (i.e., scalar invariance). This last model permitted examining whether the results from the primary MIMIC model were observed for intercepts in the multi-group CFA.

For the Monte Carlo simulations, there were 10 thousand replications with the parameter estimates derived from the above analyses used as starting values. The simulations were conducted with both the current sample sizes per group (i.e., $n=140$ for positive family history and $n=89$ for negative family history) and with sample sizes four times as large per group (i.e., $n=560$ for positive family history and $n=356$ for negative family history). Parameter bias was estimated by subtracting the obtained factor loading estimate from the CFA model from the average parameter value from the Monte Carlo replications, and dividing this number by the obtained factor loading from CFA (Muthén & Muthén, 1998–2009). A similar formula was used to estimate standard error bias.

Both the configural model ($X^2(94)=95.503$, $p=.437$; CFI=0.998; SRMR=0.047; RMSEA=0.012, 90% C.I.=0.000–0.052) and the metric model ($X^2(102)=105.445$, $p=.388$; CFI=0.995; SRMR=0.075; RMSEA=0.017, 90% C.I.=0.000–0.052) had fit indices in acceptable ranges. Furthermore, the additional constraints of the metric model did not result in a significant worsening in fit compared to the configural model (X^2 difference (8)=9.640, $p=.291$). Fit indices for the scalar invariance model were within acceptable ranges except for the SRMR ($X^2(110)=124.046$, $p=.170$; CFI=0.980; SRMR=0.083; RMSEA=0.033, 90% C.I.=0.000–0.060). This model represented a significant reduction in fit compared to the metric model (X^2 difference (8)=19.229, $p=.014$). Examination of modification indices indicated that the source of bias was in the intercepts for the Similarities and Trail Marking B-A variables. In a subsequent model in which the intercepts for these scales were freed to differ across groups, the fit indices were within acceptable ranges ($X^2(108)=110.018$, $p=.428$, CFI=0.997; SRMR=0.080; RMSEA=0.013, 90% C.I.=0.000–0.050). This partial scalar invariance model did not have a significantly worse fit than the metric model (X^2 difference (6)=4.486, $p=.611$). Notably, this represents convergent findings for the multi-group CFA and the MIMIC modeling in terms of intercept differences.

The factor loading estimates from the configural CFA on the current data set (see Table 5) were similar to the average factor loadings across the replications of the Monte Carlo simulation with $n=229$ modeled. Thus, there was minimal evidence of notable parameter bias for either family history group (Family History-Positive: average parameter bias=–.006; range=–.008 to –.004; Family History-Negative: average parameter bias=–.009; range=–.014 to –.007). These values were somewhat smaller based on the average factor loading

estimates across the replications of the Monte Carlo simulation with $n=916$ modeled (Family History-Positive: average parameter bias $=-.001$; Family History-Negative: average parameter bias $=-.003$). A similar pattern was observed for the standard errors of the factor loading estimates with small bias observed for the $n=229$ Monte Carlo simulations (Family History-Positive: average standard error bias $=-.015$; range $=-.024$ to $-.006$; Family History-Negative: average standard error bias $=-.027$; range $=-.034$ to $-.020$). Again, these values were somewhat smaller based on the Monte Carlo simulation with $n=916$ modeled (Family History-Positive: average parameter bias $=-.005$; Family History-Negative: average parameter bias $=-.005$). Very similar values (out to the third decimal point) were present when examining the partial scalar invariance CFA model and associated Monte Carlo simulations.

Discussion

The current study tested the validity of an *a priori* latent factor structure of the ACS psychometric battery in a sample of cognitively normal adults. With the minor addition of correlated errors for two tasks in one domain, CFA supported the 4-factor structure of the battery with episodic memory, semantic memory, visuospatial function, and executive function latent factors extracted. These factors correspond to commonly examined neuropsychological domains, and domains demonstrated to be relevant to the study of aging and AD. Use of cognitive domain scores derived from multiple tests has the advantages of increased reliability and decreased redundancy in testing. In addition, use of these reliable composite scores can reduce the number of comparisons between cognitive performance and a wide array of biomarkers, thus reducing Type 1 error.

Importantly, post-hoc analyses indicated that the hypothesized four-factor model evidenced a significantly improved fit compared to a three-factor model. In particular, this provided support for the separability of the episodic and semantic components of memory. In addition, there was evidence that the four-factor model was preferable to the five-factor models. The fit for the model with separate attention and working memory factors (Model 5b) was inadmissible due to an above 1 correlation between these two factors. This is consistent with the strong conceptual overlap between these two constructs. The five-factor model with separate working memory and speeded processing factors (Model 5a) had adequate fit indices. However, this model evidenced higher AIC and BIC values than the four-factor model suggesting that the four-factor model was preferable. Thus, the comparisons of alternative models generally lent some level of additional support for our proposed model. Nonetheless, the adequate fit of Model 5a suggests that it would also be reasonable to derive reliable composite scores for five domains in the case of interest in having separate working memory and speeded processing factors.

The lack of task cross-loadings across factors in our hypothesized model is consistent with several past CFA studies in older adult and AD samples (Loewenstein et al., 2001; Moleiro et al., 2013; Park et al., 2012; Robbins et al., 1994). However, other investigations have either dropped tasks with cross-loadings during an initial EFA procedure (Dowling et al., 2010), or added cross-loadings based on modification indices and theoretical considerations to improve model fit (e.g., Hayden et al., 2011; Mungas et al., 2011; Siedlecki et al., 2008). It

is recognized that the observed lack of cross-loadings in the current study does not necessarily imply that any individual task does not require multiple cognitive processes.

Intercorrelations among latent factors ranged from 0.49 to 0.80 and are generally consistent with findings from previous research. More specifically, the current study demonstrated moderate to strong associations amongst the episodic memory, semantic memory, and executive function factors (0.71 to 0.80), which is consistent with past neuropsychological studies demonstrating strong associations amongst these domains (e.g., Duff, Schoenberg, Scott, & Adams, 2005). This pattern is also consistent with prior studies demonstrating common prefrontal activations during tasks designed to assess these functions (e.g., Nyberg et al., 2003; for a review, see: Cabeza & Nyberg, 2000). In addition, the relatively strong correlation between the executive function and visuospatial ability domains is consistent with previously observed associations between working memory and perceptual reasoning for the WAIS scales (e.g., Davis et al., 2003; Holdnack, Zhou, Larrabee, Millis, & Salthouse, 2011).

As both family history and APOE genotype are documented risk factors for cognitive decline and developing AD, we further tested aspects of measurement invariance of the latent structure across groups using MIMIC modeling. The four-factor structure had adequate fit in these analyses. However, there was some indication of measurement bias after adjustment for possible Type 1 error inflation. For the family history analyses, performance on WAIS-III Similarities and Trail Making Part B-A differed across groups controlling for latent factor levels of semantic memory and executive function, respectively. For the APOE status analyses, performance on the SRT Free Recall task differed across groups controlling for latent factor levels of episodic memory. However, these findings should be interpreted with caution as the X^2 difference test can also be dependent on sample size. In fact, the other model fit indices were quite similar across models.

The magnitude of the standardized parameter estimates suggests that these effects were in the small range. In addition, it has been suggested that having at least two scales per latent construct without evidence of invariance can be considered adequate for supporting comparisons of the construct across groups (Dowling et al., 2010; Steenkamp & Baumgartner, 2010). Furthermore, there were no significant effects of family history or APOE status on the relevant latent cognitive constructs in either the base models or in the Direct Effects models that accounted for potential measurement bias effects. This suggests that any measurement bias that may have been present did not substantially influence interpretations of the effect of family history or APOE status on the latent construct of interest. That is, the concern with measurement bias is that it may lead to biased interpretations of the effects of the grouping variable of interest on the latent constructs, and there was not strong evidence of this in the current analyses. Overall, current results are suggestive that estimates of the cognitive constructs including these tasks are appropriate to use across groups. Notably, measurement bias with respect to family history status was not observed for WAIS-III Similarities and Trail Making tasks in a similar, but much larger sample of middle-aged adults (Dowling et al., 2010).

The MIMIC approach employed here in the primary *a priori* analyses does not allow assessment of whether the factor loadings differ across groups, whether the unexplained variance of the indicators differs across groups, or whether the associations between factors and their variabilities differ across groups. Thus, for example, it is not possible to know whether there are group differences in the relationships of the indicators to the latent constructs (i.e., in factor loadings) based on the MIMIC approach. Post-hoc analyses using multi-group CFA and Monte Carlo simulations provided some limited evidence that there were not notable family history group differences in factor loading estimates or the associated standard errors. The magnitude of the factor loading differences across the family history groups was in the small range (average difference=0.023; range=-.122 to .147). However, it is still possible that there was insufficient power to detect group differences. The factor loading differences observed here are somewhat similar to those noted in the work by Dowling and colleagues (2010) (i.e., average difference=.022; range=-.08 to .13). This study did not observe any significant measurement invariance across family history groups for a five-factor model (verbal ability, visuospatial ability, processing speed/executive function, working memory, and verbal learning/memory) in a much larger sample. Notably, there is also some simulation evidence that while the MIMIC approach is not sensitive to factor loading invariance, there is still minimal bias in the estimate of group differences in the latent factor means (Kim, Yoon, & Lee, 2012). Thus, there is some support of the appropriateness of comparing groups on our cognitive constructs, particularly in terms of family history.

Overall, results indicate that rather than examining a large set of inter-related psychometric tasks, composite scores representing the domains of episodic memory, semantic memory, visuospatial function, and executive function can appropriately be used. Furthermore, the current results indicate that the observed factor structure is consistent across family history and APOE status groups. This demonstration that family history does not appear to significantly alter latent cognitive structure is important as more research groups begin examining children of AD individuals who are at increased risk of developing AD. Our results provides support for the use of the same cognitive composite scores to assess for group differences and for comparisons with other biomarkers in the ACS sample as well as by other investigators examining aging and AD-related research questions in middle-aged and older adult samples. As noted, the use of the reliable composite scores would be particularly beneficial in minimizing Type 1 error inflation when examining associations between cognitive performance and other variables. The cognitive battery consists of readily available tasks with evidence of reliability. A particularly advantageous feature of the current test battery is that each domain is assessed by separate tasks rather than relying on multiple variables from a task as done for certain domains in some prior CFA work in middle-aged and older adult samples (e.g., Dowling et al., 2010; Hayden et al., 2011; Park et al., 2012; Siedlecki et al., 2008). Thus, each domain is represented by a broader range subdomains and methods. For example, the episodic memory domain incorporates list learning, paired associate learning and prose recall rather than multiple trials from a list-learning task. However, it should be noted that the current sample is a convenience sample of generally highly educated individuals. Thus, the current results may not generalize to less educated populations. Nonetheless, a significant strength of the ACS sample is the longitudinal

collection of a wide array of cognitive, clinical, and biomarker data. Future follow-up studies will model longitudinal changes in these biomarkers and relate them to the validated cognitive composites to further characterize the preclinical progression of AD.

Acknowledgments

This work was supported by grant P01 AG026276 (PI: JCM) from the National Institute on Aging, and grant P50 AG05681 (PI: JCM) from the Charles F. and Joanne Knight Alzheimer's Research Initiative of the Washington University Alzheimer's Disease Research Center. S. Allison and N. Lucena were supported by National Institute on Aging 5T32AG00030. We thank the Genetics Core (Alison Goate, DPhil, Core Leader) of the Alzheimer's Disease Research Center for the APOE data. The authors have no conflicts of interest to report.

References

- Akaike H. Factor analysis and AIC. *Psychometrika*. 1987; 512(3):317–332.
- Aschenbrenner AJ, Balota DA, Tse CS, Fagan AM, Holtzman DM, Benzinger TL, Morris JC. Alzheimer disease biomarkers, attentional control, and semantic memory retrieval: Synergistic and mediational effects of biomarkers on a sensitive cognitive measure in non-demented older adults. *Neuropsychology*. 2015; 29(3):368–381. [PubMed: 25222200]
- Bäckman L, Jones S, Berger A, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*. 2005; 19(4):520–531. [PubMed: 16060827]
- Bangert AS, Balota DA. Keep up the pace: Declines in simple repetitive timing differentiate healthy aging from the earliest stages of Alzheimer's disease. *Journal of the International Neuropsychological Society*. 2012; 18(6):1052–1063. [PubMed: 22929329]
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]
- Bentler, PM. EQS: Structural equations program manual, Version 5.0. Los Angeles: BMDP Statistical Software; 1995.
- Benton, AL., Hamser, K., Varney, NR., Spreen, O. Contributions to neuropsychological assessment: A clinical manual. New York: Oxford University Press; 1983.
- Brown, TA. Confirmatory factor analysis for applied research. New York: The Guilford Press; 2006.
- Browne, MW., Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA., Long, JS., editors. Testing structural equation models. Beverly Hills: Sage; 1993.
- Burke WJ, Miller JP, Rubin EH, Morris JC, Coben LA, Ducheck J, Wittels IG, Berg L. Reliability of the Washington University Clinical Dementia Rating. *Archives of Neurology*. 1988; 45:31–32. [PubMed: 3337672]
- Cabeza R, Nyberg L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*. 2000; 12(1):1–47. Review.
- Chapman RM, Mapstone M, Porsteinsson AP, Gardner MN, McCrary JW, DeGrush E, Reilly LA, Sandoval TC, Guillily MD. Diagnosis of Alzheimer's disease using neuropsychological testing improved by multivariate analyses. *Journal of Clinical and Experimental Neuropsychology*. 2010; 32:793–808. [PubMed: 20358452]
- Cronin-Golomb, A., Amick, M. Spatial abilities in aging, Alzheimer's disease, and Parkinson's disease. In: Boller, F., Cappa, S., editors. Handbook of neuropsychology, vol. 6: Aging and dementia. 2. Amsterdam: Elsevier; 2001.
- Davis RN, Massman PJ, Doody RS. WAIS-R factor structure in Alzheimer's disease patients: A comparison of alternative models and an assessment of their generalizability. *Psychology and Aging*. 2003; 18(4):836–843. [PubMed: 14692868]
- Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP. The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*. 2003; 9(6):936–946. [PubMed: 14632252]

- Dikmen SS, Heaton RK, Grant I, Temkin NR. Test-retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological test battery. *Journal of the International Neuropsychological Society*. 1999; 5:346–356. [PubMed: 10349297]
- Dowling NM, Hermann B, La Rue A, Sager MA. Latent structure and factorial invariance of a neuropsychological test battery for the study of preclinical Alzheimer's disease. *Neuropsychology*. 2010; 24:742–756. [PubMed: 21038965]
- Duff K, Schoenberg MR, Scott JG, Adams RL. The relationship between executive functioning and verbal and visual learning and memory. *Archives of Clinical Neuropsychology*. 2005; 20:111–122. [PubMed: 15620817]
- Franzen, M. Reliability and validity in neuropsychological assessment. New York: Kluwer Academic/Plenum Publishers; 2000.
- Goodglass, H., Kaplan, E. Boston Diagnostic Aphasia Examination booklet, Oral expression, J. Animal naming (fluency in controlled association). Philadelphia: Lea & Febiger; 1983.
- Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. *Neurology*. 1988; 38:900–903. [PubMed: 3368071]
- Grober E, Ocepek-Welickson K, Teresi JA. The Free and Cued Selective Reminding Test: Evidence of psychometric adequacy. *Psychology Science Quarterly*. 2009; 51(3):266–282.
- Hayden KM, Jones RN, Zimmer C, Plassman BL, Browndyke JN, Pieper C, Warren LH, Welsh-Bohmer KA. Factor structure of the National Alzheimer's Coordinating Centers Uniform Dataset Neuropsychological Battery: An evaluation of invariance between and within groups over time. *Alzheimer Disease and Associated Disorders*. 2011; 25(2):128–137. [PubMed: 21606904]
- Hodges JR, Salmon DP, Butters N. The nature of the naming deficit in Alzheimer's and Huntington's disease. *Brain*. 1991; 114(4):1547–1558. [PubMed: 1832072]
- Holdnack JA, Zhou X, Larrabee GJ, Millis SR, Salthouse TA. Confirmatory factor analysis of the WAIS-IV/WMS-IV. Assessment. 2011; 18(2):178–191. [PubMed: 21208975]
- Hoyle, RH. Structural Equation Modeling for Social and Personality Psychology. Los Angeles: Sage; 2011.
- Hu L, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Hutchison KA, Balota DA, Duchek JM. The utility of Stroop task switching as a marker for early-stage Alzheimer's disease. *Psychology and Aging*. 2010; 25(3):545–559. [PubMed: 20853964]
- Johnson DK, Storandt M, Morris JC, Langford ZD, Galvin JE. Cognitive profiles in dementia: Alzheimer disease vs healthy brain aging. *Neurology*. 2008; 71(22):1783–1789. [PubMed: 19029518]
- Kim ES, Yoon M, Taehun L. Testing measurement invariance using MIMIC: Likelihood ratio test with a critical value adjustment. *Educational and Psychological Measurement*. 2012; 72(3):469–492.
- Jöreskog, KG., Sörbom, D. LISREL V: Analysis of linear structural relationships by the method of maximum likelihood. Chicago: National Educational Resources; 1981.
- Loewenstein DA, Ownby R, Schram L, Acevedo A, Rubert M, Argüelles 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. *Journal of Clinical and Experimental Neuropsychology*. 2001; 23(3):274–284. [PubMed: 11404806]
- McCulla MM, Coats M, Van Fleet N, Duchek J, Grant E, Morris JC. Reliability of clinical nurse specialists in the staging of dementia. *Archives of Neurology*. 1989; 46:1210–1211. [PubMed: 2818255]
- Mertens VB, Gagnon M, Coulombe D, Messier C. Exploratory factor analysis of neuropsychological tests and their relationship to the Brown-Peterson Task. *Archives of Clinical Neuropsychology*. 21(7):733–739.
- Mitchell MB, Shaughnessy LW, Shirk SD, Yang FM, Atri A. Neuropsychological test performance and cognitive reserve in healthy aging and the Alzheimer's disease spectrum: A theoretically driven factor analysis. *Journal of the International Neuropsychological Society*. 2012; 18(6):1071–1080. [PubMed: 23039909]
- Mitrushina, M., Boone, KB., Razani, J., D'Elia, LF. Handbook of normative data for neuropsychological assessment. 2. New York: Oxford University Press; 2005.

- Moleiro C, Madureira S, Verdelho A, Ferro JM, Poggesi A, Chabriat H, Erkinjuntti T, Fazekas F, Hennerici M, O'Brien J, Pantoni L, Salvadori E, Scheltens P, Visser MC, Wahlund LO, Waldemar G, Wallin A, Inzitari D. LADIS Study. Confirmatory factor analysis of the Neuropsychological Assessment Battery of the LADIS Study: A longitudinal analysis. *Journal of Clinical and Experimental Neuropsychology*. 2013; 35(3):269–278. [PubMed: 23398381]
- Morris JC, McKeel DW Jr, Storandt M, Rubin EE, Price JL, Grant EA, Ball JM, Berg L. Very mild Alzheimer's disease: Informant-based clinical, psychometric, and pathologic distinction from normal aging. *Neurology*. 1991; 41:469–478. [PubMed: 2011242]
- Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*. 1993; 43:2412–2414.
- Morris JC, Ernesto C, Schafer K, Coats M, Leon S, Sano M, Thal LJ, Woodbury P. Clinical Dementia Rating training and reliability in multicenter studies: The Alzheimer's Disease Cooperative Study experience. *Neurology*. 1997; 48:1508–1510. [PubMed: 9191756]
- Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology*. 2001; 58:397–405. [PubMed: 11255443]
- Mungas D, Widaman KF, Reed BR, Tomaszewski Farias S. Measurement invariance of neuropsychological tests in diverse older persons. *Neuropsychology*. 2011; 25(2):260–269. [PubMed: 21381830]
- Muthén, BO. Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In: Wainer, H., Braun, H., editors. *Test Validity*. Hillsdale: Lawrence Erlbaum; 1988.
- Muthén, LK., Muthén, BO. *Mplus user's guide*. 5. Los Angeles: Muthén & Muthén; 1998–2009.
- Nowinski CJ, Victorson D, Debb SM, Gershon RC. Input on NIH Toolbox inclusion criteria. *Neurology*. 2013; 80(Suppl 3):S7–S12. [PubMed: 23479548]
- Nyberg L, Marklund P, Persson J, Cabeza R, Forkstam C, Petersson KM, Ingvar M. Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia*. 2003; 41(3):371–377. [PubMed: 12457761]
- Owby RL, Loewenstein DA, Schram L, Acevedo A. Assessing the cognitive abilities that differentiate patients with Alzheimer's disease from normals: Single and multiple factor models. *International Journal of Geriatric Psychiatry*. 2004; 19:232–242. [PubMed: 15027038]
- Park LQ, Gross AL, McLaren DG, Pa J, Johnson JK, Mitchell M, Manly JJ. Alzheimer's Disease Neuroimaging Initiative. Confirmatory factor analysis of the ADNI Neuropsychological Battery. *Brain Imaging and Behavior*. 2012; 6(4):528–539. [PubMed: 22777078]
- Perry RJ, Hodges JR. Attention and executive deficits in Alzheimer's disease. A critical review. *Brain*. 1999; 122(3):383–404. Review. [PubMed: 10094249]
- Peterson LR, Peterson MJ. Short-term retention of individual verbal items. *Journal of Experimental Psychology*. 1959; 58:193–198. [PubMed: 14432252]
- Robbins TW, James M, Owen AM, Sahakian BJ, McInnes L, Rabbitt P. Cambridge Neuropsychological Test Automated Battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia*. 1994; 5:266–281. [PubMed: 7951684]
- Salmon, DP. Disorders of memory in Alzheimer's disease. In: Cermak, LS., editor. *Handbook of neuropsychology*, vol. 2: Memory and its disorders. 2. Amsterdam: Elsevier; 2000.
- Satorra A, Bentler P. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001; 66(4):507–514.
- Schwarz G. Estimating the dimension of a model. *Annals of Statistics*. 1978; 6(2):461–464.
- Schumacker, RE., Lomax, RG. *A beginner's guide to structural equation modeling*. Vol. 2. Hillsdale: 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(3):400–411. [PubMed: 18444718]
- Small BJ, Rosnick CB, Fratiglioni L, Bäckman L. Apolipoprotein E and cognitive performance: A meta-analysis. *Psychology and Aging*. 2004; 19:592–600. [PubMed: 15584785]

- Solomon PR, Hirschhoff A, Kelly B, Relin M, Brush M, DeVeaux RD, Pendlebury WW. A 7 minute neurocognitive screening battery highly sensitive to Alzheimer's disease. *Archives of Neurology*. 1998; 55(3):349–355. [PubMed: 9520009]
- Strauss, E., Sherman, EMS., Spreen, O. A compendium of neuropsychological tests. 3. New York: Oxford University Press; 2006.
- Steenkamp JBEM, Baumgartner H. Assessing measurement invariance in cross national consumer research. *Journal of Consumer Research*. 2010; 25:78–107.
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180. [PubMed: 26794479]
- Storandt M, Grant EA, Miller JP, Morris JC. Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. *Neurology*. 2006; 67:467–473. [PubMed: 16894109]
- Talbot C, Lendon C, Craddock N, Shears S, Morris JC, Goate A. Protection against Alzheimer's disease with apoE e2. *Lancet*. 1994; 343:1432–1433.
- Twamley EW, Ropacki SAL, Bondi MW. Neuropsychological and neuroimaging changes in preclinical Alzheimer's disease. *Journal of the International Neuropsychological Society*. 2006; 12:707–735. [PubMed: 16961952]
- Unsworth N, Spillers GJ, Brewer GA. Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *The Quarterly Journal of Experimental Psychology*. 2011; 64(3):447–466. [PubMed: 20839136]
- Watts AS, Loskutova N, Burns JM, Johnson DK. Metabolic syndrome and cognitive decline in early Alzheimer's disease and healthy older adults. *Journal of Alzheimer's Disease*. 2013; 35(2):253–265.
- Wechsler, D. Wechsler Memory Scale. 3. San Antonio: The Psychological Corporation; 1997a.
- Wechsler, D. Wechsler Adult Intelligence Scale. 3. San Antonio: The Psychological Corporation; 1997b.
- Woodcock, RW., McGrew, KS., Mather, N. Woodcock-Johnson III Tests of Cognitive Abilities. Itasca; Riverside: 2001.
- Xiong C, Roe CM, Buckles V, Fagan A, Holtzman D, Balota D, Duchek J, Storandt M, Mintun M, Grant E, Snyder AZ, Head D, Benzinger TL, Mettenberg J, Csernansky J, Morris JC. Role of family history for Alzheimer biomarker abnormalities in the Adult Children Study. *Archives of Neurology*. 2011; 68:1313–1319. [PubMed: 21987546]

Figure 1.

Confirmatory factor analysis for full sample. Values at unidirectional arrows represent factor loadings with the 95% confidence interval in parentheses. Values at bi-directional arrows represent correlations between factors. Degrees of freedom = 47. Free parameters = 43.

Figure 2.

Direct Effects MIMIC model for family history. Values at unidirectional arrows to individual tasks represent factor loadings with the 95% confidence interval in parentheses. Values (dark gray) at unidirectional arrows to the latent constructs represent standardized regression coefficients reflecting the association between the family history covariate and the latent factor; the standard errors are in parentheses. The dark gray dashed lines represent the direct effect of family history on the Similarities and Trail Making Part B-A tasks. Values at bi-directional arrows represent correlations between factors with the standard errors in parentheses. Degrees of freedom = 53. Free parameters = 49.

Figure 3.

Direct Effects MIMIC model for APOE status. Values at unidirectional arrows to individual tasks represent factor loadings with the 95% confidence interval in parentheses. Values (dark gray) at unidirectional arrows to the latent constructs represent standardized regression coefficients reflecting the association between the APOE status covariate and the latent factor; the standard errors are in parentheses. The dark gray dashed line represents the direct effect of APOE status on the SRT Free Recall test. Values at bi-directional arrows represent correlations between factors with the standard errors in parentheses. Degrees of freedom = 55. Free parameters = 48.

Table 1

Demographic data.

	Total	Family History Groups		APOE Groups	
		positive	negative	positive	negative
N	229	140	89	76	152
Gender (f/m)	154/75	96/44	58/31	52/24	102/50
Age (mean (SD))	58 (6)	58 (6)	58 (5)	58 (7)	58 (5)
Age range (yrs)	43–73	43–73	46–70	43–73	43–70
Education (mean (SD))	16 (2)	16 (2)	16 (2)	16 (2)	16 (2)
Education range (yrs)	6–23	6–20	12–23	6–20	11–23
MMSE (mean (SD))	29.4 (.96)	29.4 (1.02)	29.5 (.85)	29.4 (1.10)	29.5 (.88)
MMSE range	25–30	25–30	26–30	25–30	26–30

Note.

* p<.05;

APOE data was missing for one individual.

Table 2

Descriptive statistics for individual tasks for the total sample.

	Mean (SD)	Range	Skew	Kurtosis	Kolmogorov-Smirnov
Logical Memory	25.03 (8.07)	2–44	–.25	–.40	.06
Verbal Paired Associates	19.65 (7.74)	0–32	–.37	–.74	.08**
SRT Free Recall	32.77 (5.92)	16–48	–.24	–.02	.06*
Similarities	27.12 (4.00)	10–33	–1.28	2.41	.12***
Information	21.15 (4.14)	7–28	–.89	.91	.11***
Animal Naming	23.62 (5.81)	10–40	.10	–.35	.05
Letter-Number Sequencing	11.68 (2.83)	5–20	.43	.18	.13***
Auditory Consonant Trigrams	49.02 (7.47)	24–60	–.82	.38	.10***
Trailmaking B-A	33.91 (21.51)	5–130	2.03	4.85	.19***
Block Design	40.16 (11.16)	12–61	–.28	–.43	.06*
Benton Line Orientation	25.74 (3.69)	9–30	–1.17	1.59	.16***
Spatial Relations	71.90 (5.55)	53–81	–1.00	.77	.13***

Note.

* p<.05;

**

p<.01;

p<.001.

Table 3

Descriptive statistics for individual tasks for family history and APOE groups.

	Mean (SD)		Range		Skew		Kurtosis		Kolmogorov-Smirnov	
	FH+	FH-	FH+	FH-	FH+	FH-	FH+	FH-	FH+	FH-
Logical Memory	24.93 (8.25)	25.18 (7.83)	2-44	9-41	-.28	-.18	-.23	-.71	.06	.08
Verbal Paired Associates	19.88 (7.43)	19.30 (8.22)	4-32	0-32	-.31	-.41	-.72	-.83	.07	.13**
SRT Free Recall	32.79 (5.99)	32.73 (5.83)	18-47	16-48	-.24	-.25	-.07	.14	.08	.07
Similarities	27.37 (3.99)	26.72 (4.02)	10-33	12-32	-.136	-.122	2.66	2.36	.13***	.14***
Information	20.98 (4.21)	21.42 (4.03)	7-28	8-28	-.81	-.106	.93	1.03	.09**	.14***
Animal Naming	23.14 (5.62)	24.38 (6.07)	10-37	10-40	.08	.08	-.55	-.14	.08*	.06
Letter-Number Sequencing	11.74 (2.80)	11.58 (2.90)	6-20	5-19	.68	.08	.43	-.17	.15***	.10*
Auditor Consonant Trigrams	49.08 (7.56)	48.93 (7.36)	24-60	29-59	-.89	-.71	.75	-.19	.10**	.13**
Trailmaking B-A	36.32 (24.02)	30.09 (16.16)	7-130	5-90	1.87	1.87	3.78	4.31	.17***	.19***
Block Design	40.42 (10.77)	39.74 (11.80)	12-61	12-59	-.21	-.36	-.23	-.68	.05	.11*
Benton Line Orientation	26.05 (3.36)	25.26 (4.12)	14-30	9-30	-.99	-.123	.70	1.74	.14***	.17***
Spatial Relations	72.41 (5.07)	71.08 (6.17)	54-80	53-81	-.127	-.65	1.56	.08	.15***	.11*
Logical Memory	24.90 (8.70)	25.07 (7.79)	2-40	4-44	-.40	-.13	-.53	-.34	.10*	.06
Verbal Paired Associates	19.88 (8.08)	19.51 (7.60)	4-32	0-32	-.31	-.40	-.114	-.50	.14**	.09***
SRT Free Recall	31.66 (6.10)	33.31 (5.78)	18-43	16-48	-.20	-.24	-.41	.22	.07	.07
Similarities	26.76 (3.98)	27.33 (4.01)	12-33	10-33	-.125	-.136	2.21	2.79	.16***	.14***
Information	20.90 (4.36)	21.26 (4.04)	7-28	7-28	-.75	-.97	.62	1.14	.11*	.11***
Animal Naming	22.82 (5.37)	24.03 (6.02)	10-33	10-40	-.27	.19	-.52	-.44	.10	.07*
Letter-Number Sequencing	11.34 (2.62)	11.86 (2.94)	6-19	5-20	.64	.31	.35	.15	.16***	.12***
Auditory Consonant Trigrams	47.71 (7.86)	49.67 (7.22)	24-60	29-60	-.81	-.80	.57	.18	.11*	.11***
Trailmaking B-A	37.80 (25.69)	32.04 (18.87)	9-130	5-130	1.86	1.95	3.11	5.54	.21***	.16***
Block Design	39.39 (11.36)	40.60 (11.09)	12-60	12-61	-.16	-.36	-.54	-.31	.07	.07

	Mean (SD)		Range		Skew		Kurtosis		Kolmogorov-Smirnov	
	FH+	FH-	FH+	FH-	FH+	FH-	FH+	FH-	FH+	FH-
Benton Line Orientation	25.43 (3.94)	25.93 (3.54)	9-30	14-30	-1.33	-1.08	2.72	.82	.16***	.15***
Spatial Relations	71.56 (5.23)	72.08 (5.72)	54-79	53-81	-1.04	-1.01	.87	.78	.19***	.12***

Note.
* p<.05;
** p<.01;
*** p<.001;
FH = family history.

Table 4

No-direct effects base MIMIC models for family history and APOE groups.

<i>Domain/Indicator</i>	Family History		APOE status	
	Parameter Estimate (SE)	Item Loading (95% CI)	Parameter Estimate (SE)	Item Loading (95% CI)
<u>Episodic Memory</u>	.01 (.08)		-.03 (.08)	
Logical Memory		.77 (.68-.87) **		.78 (.68-.87) **
Verbal Paired Associates		.67 (.57-.77) **		.68 (.57-.78) **
SRT Free Recall		.49 (.39-.59) **		.48 (.38-.58) **
<u>Semantic Memory</u>	-.04 (.10)		-.12 (.09)	
Similarities		.62 (.51-.72) **		.63 (.53-.72) **
Information		.59 (.49-.69) **		.59 (.48-.70) **
Animal Naming		.54 (.44-.64) **		.54 (.44-.63) **
<u>Executive Function</u>	-.03 (.08)		-.07 (.07)	
Letter-Number Sequencing		.64 (.56-.72) **		.64 (.56-.72) **
Auditory Consonant Trigrams		.72 (.64-.80) **		.72 (.64-.80) **
Trailmaking B-A		.53 (.42-.63) **		.53 (.42-.63) **
<u>Visuospatial Ability</u>	.10 (.08)		-.16 (.08) *	
Block Design		.79 (.73-.86) **		.80 (.73-.86) **
Benton Line Orientation		.65 (.57-.73) **		.65 (.57-.73) **
Spatial Relations		.77 (.69-.84) **		.76 (.69-.84) **

Note.

* p<.05;

** p<.0001;

Parameter estimates represent standardized regression coefficients reflecting the association between the covariate and the latent factor. SE = standard error. Degrees of freedom = 55. Free parameters = 47. Results for the Direct Effects Models are presented in Figure 2 (Family History) and Figure 3 (APOE status).

Table 5

Factor loadings for family history groups from multi-group CFA

Domain/Indicator	Family History - Positive	Family History - Negative
<u>Episodic Memory</u>		
Logical Memory	.84 (.73-.94) **	.72 (.53-.90) **
Verbal Paired Associates	.62 (.49-.76) **	.75 (.62-.88) **
SRT Free Recall	.42 (.30-.55) **	.57 (.40-.73) **
<u>Semantic Memory</u>		
Similarities	.61 (.46-.75) **	.65 (.52-.79) **
Information	.62 (.49-.74) **	.55 (.35-.76) **
Animal Naming	.54 (.41-.66) **	.55 (.39-.71) **
<u>Executive Function</u>		
Letter-Number Sequencing	.67 (.57-.77) **	.62 (.49-.76) **
Auditory Consonant Trigrams	.68 (.56-.79) **	.78 (.66-.91) **
Trailmaking B-A	.55 (.43-.68) **	.50 (.30-.71) **
<u>Visuospatial Ability</u>		
Block Design	.83 (.75-.91) **	.76 (.66-.86) **
Benton Line Orientation	.64 (.52-.76) **	.64 (.52-.77) **
Spatial Relations	.71 (.61-.80) **	.83 (.73-.93) **

Note.

** p<.0001