



Published in final edited form as:

*Brain Imaging Behav.* 2012 December ; 6(4): 528–539. doi:10.1007/s11682-012-9190-3.

## Confirmatory Factor Analysis of the ADNI Neuropsychological Battery

Lovingly Quitania Park<sup>1</sup>, Alden L. Gross<sup>2</sup>, Donald McLaren<sup>4,5</sup>, Judy Pa<sup>3</sup>, Julene K. Johnson<sup>3,6</sup>, Meghan Mitchell<sup>4</sup>, Jennifer J. Manly<sup>7</sup>, and for the Alzheimer's Disease Neuroimaging Initiative

<sup>1</sup>Alzheimer's Disease Center, Department of Neurology, University of California, Davis, CA

<sup>2</sup>Institute for Aging Research, Harvard Medical School and Division of Gerontology, Beth Israel Deaconess Medical Center, Boston, MA

<sup>3</sup>Department of Neurology, University of California, San Francisco, CA

<sup>4</sup>Geriatric Research Education and Clinical Center, ENRM Veterans Hospital; Department of Neurology, Massachusetts General Hospital; and Harvard Medical School

<sup>5</sup>Athinoula A. Martinos Center for Biomedical Imaging and Department of Neurology, Massachusetts General Hospital

<sup>6</sup>Institute for Health and Aging; Department of Social and Behavioral Sciences, University of California, San Francisco, San Francisco, CA

<sup>7</sup>Taub Institute for Research on Alzheimer's Disease and the Aging Brain; Department of Neurology, Columbia University Medical Center, New York, NY

### Abstract

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a large multi-center study designed to develop optimized methods for acquiring longitudinal neuroimaging, cognitive, and biomarker measures of AD progression in a large cohort of patients with Alzheimer's disease (AD), patients with mild cognitive impairment, and healthy controls. Detailed neuropsychological testing was conducted on all participants. We examined the factor structure of the ADNI Neuropsychological Battery across older adults with differing levels of clinical AD severity based on the Clinical Dementia Rating Scale (CDR). Confirmatory factor analysis (CFA) of 23 variables from 10 neuropsychological tests resulted in five factors (memory, language, visuospatial functioning, attention, and executive function/processing speed) that were invariant across levels of cognitive impairment. Thus, these five factors can be used as valid indicators of cognitive function in older adults who are participants in ADNI.

### Keywords

ADNI; neuropsychology; cognition; cognitive change; confirmatory factor analysis

---

Send Correspondence to: Lovingly Quitania Park, Alzheimer's Disease Center, UCDCM Department of Neurology, 4860 Y Street, Suite 3900, Sacramento, CA 95817, lovingly.quitania@ucdmc.ucdavis.edu.

\*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators with the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.ucla.edu/wpcontent/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

## INTRODUCTION

Alzheimer's disease (AD) affects approximately 2.5 million people in the United States and is the sixth leading cause of death (Brookmeyer et al., 2011). Given the detrimental public health implications of AD, identification of neuropsychological markers of early decline are of tremendous benefit since cognitive changes are common in the preclinical stages or those with Mild Cognitive Impairment (MCI). Impairment in episodic memory is a hallmark characteristic of AD, and thus memory functions have received most of the attention (Desgranges, Eustache, Rioux, de La Sayette, & Lechevalier, 1996; Gallagher & Koh, 2011; Grossman et al., 2003; MacDonald, Almor, Henderson, Kempler, & Andersen, 2001; Nestor, Fryer, & Hodges, 2006; Spaan, Raaijmakers, & Jonker, 2003). Generally, older adults with MCI have worse cognitive function than what is typically observed in healthy aging, and the greatest levels of impairment occur in AD. Subsequently, those with AD typically perform more poorly on tests of memory, language, executive function, and visuospatial ability as part of disease progression (DeCarli et al., 2004; Pike & Savage, 2008; Ready, Ott, & Grace, 2004; Petersen et al., 2000; Stern, Albert, Tang, & Tsai, 1999; Welsh, Butters, Hughes, Mohs, & Heyman, 1991). These cognitive domains eventually become differentially affected as part of the neurodegenerative disease process, subsequently resulting in progressive functional decline and an eventual diagnosis of dementia.

Neuropsychological tests quantify cognition in terms of achieved scores, however it is unknown if there are qualitative differences in performance that can emerge across the spectrum of cognitive aging (Siedlecki, Honig, & Stern, 2008). Cognitive scores may represent something different in those with MCI and AD. For example, in young children mathematical ability in first grade is measured by addition or subtraction problems, but by older adulthood or even high school such abilities are so ingrained, that more complicated tests such as serial 7's are used as measures of concentration and attention (Binet, 1905). One way to address the issue of qualitative differences in the meaning of tests over time or level of cognitive impairment is through the use of confirmatory factor analysis (CFA). CFA quantifies unobserved, or latent, constructs that describe the common variability among sets of indicators, or cognitive tests, and allows tests from a neuropsychological battery to be summarized into cognitive domains (Reise, Widaman, & Pugh, 1993; Widaman, Ferrer, & Conger, 2010). In turn, measurement invariance is used to test the factors generated by the CFA to determine if the correlations among neuropsychological tests (e.g. list-learning and recognition) vary across different groups (e.g. MCI versus AD), and whether there are qualitative differences in a construct (e.g., memory) across group (Siedlecki, Honig, & Stern, 2008). Measurement invariance analyses using CFA can provide information about the construct validity of a neuropsychological battery across subgroups of older adults with varying levels of cognitive ability (Mungas, Widaman, Reed, & Tomaszewski Farias, 2011).

CFA has been used to test measurement invariance of neuropsychological batteries (Dowling et al., 2010; Hayden et al., 2011; Tuokko et al., 2009) to characterize how well these compilations of tests measure cognition in older adults with and without cognitive impairment. For example, Hayden and colleagues (2011) conducted a factor analysis of the National Alzheimer's Coordinating Centers' Neuropsychological Battery to assess invariance over time and between groups of a large sample of older adults who were cognitively normal or diagnosed with varying levels of dementia including AD. They identified four factors representing attention, executive, memory, and language. The factors were stable across the different diagnostic groups, suggesting that variability in cognitive performance across these groups represents underlying neurodegenerative process rather than a fundamental change in the meaning of the tests. Similarly, Siedlecki and colleagues (2008) used CFA to establish that memory, the predominant cognitive impairment in AD,

may actually represent something different in those with MCI or dementia as compared to those who are cognitively normal. However, cognitive performance can be affected by issues unrelated to neurodegeneration such as language and level of education. As such, it is unclear if neuropsychological tests can measure cognition accurately in older adults from different ethnic groups. To address this issue, researchers have used CFA to test five factors from the Spanish English Neuropsychological Scales (SENAS) (i.e. episodic memory, semantic memory, spatial ability, attention, and fluency) in a sample of diverse older adults with varying levels of education and linguistic backgrounds. They found that these measures were invariant across groups of monolingual and bilingual Hispanics, African Americans, and Caucasians (Mungas, Widaman, Reed, & Farias, 2011). These findings demonstrate that using CFA to test measurement invariance can identify differences in constructs between different diagnostic groups, and it also has cross-cultural applicability.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a multi-site longitudinal study designed to identify biological and clinical markers of AD (Mueller et al., 2005). The neuropsychological battery used in ADNI is comprehensive and includes measures of memory, executive function, attention, visuospatial ability, and language. Many studies in ADNI use summary scores from brief global cognitive screening measures (e.g., Mini Mental State Examination, Alzheimer's Disease Assessment Scale-Cognitive Subscale) (Evans et al., 2010; Leow et al., 2009), rather than the more comprehensive tests that are collected as part of the battery. However, these scales have limited sensitivity in high functioning and well-educated groups, and the summary scores from these measures do not capture varying levels of change that can occur across different cognitive domains (Crane et al., 2008; Landau et al., 2011; Leow et al., 2009). There are multiple neuropsychological variables available from each test in the ADNI Neuropsychological Battery and despite the large volume of studies using ADNI data, the underlying factor structure of this battery has not been investigated. As such, there is no systematic way to take advantage of its resources without possibly increasing the probability of committing a Type I error. If the factor structure differs as the disease progresses, this may indicate alterations in underlying neural systems that affect multiple cognitive domains (Hayden et al., 2011). Applying a CFA to the ADNI Neuropsychological Test battery and testing invariance over a range of cognitive function can consolidate the number of variables into composites that represent distinct cognitive domains. Sensitivity of the test scores to detect cognitive change would also increase over the entire course of the disease, from preclinical to clinical, and among normal older adults with different demographic backgrounds (Mungas, Widaman, Reed, & Farias, 2011).

Although using a CFA to test measurement invariance has been done in numerous samples of older adults, many studies use groups that are pre-defined by their cognitive test scores. This approach may be biased if the test scores used to define the groups also serve as an outcome measure. Neuropsychological test performance was a key feature of assigning participants into diagnostic groups in ADNI. The present study divided the participants into two groups based on functional capacity, as an alternative way to characterize the participants' level of severity. This was done to prevent circularity in defining groups based on their neuropsychological test performance.

The primary aim of the present study was to characterize the factor structure of the ADNI Neuropsychological Battery, and to test its measurement invariance across a range of subjects with varying degrees of disease severity. On the basis of expert opinion, we hypothesized a priori that five distinct factors would emerge from a factor analysis of the ADNI neuropsychological battery representing language, attention, memory, executive function/processing speed, and visuospatial functioning. Further, we hypothesized that the

factors would be invariant across two groups with differing levels of clinical dementia severity.

## METHODS

### The Alzheimer's Disease Neuroimaging Initiative

The National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organizations launched ADNI in 2003 (Weiner et al., 2011, see Disclosure Statement in acknowledgements). The primary goal of ADNI was to test neuroimaging and other biological markers in the progression of MCI and early AD to improve methods in clinical trials for evaluating pathological progression. More than 800 participants, aged 55–90 years, were recruited from 59 sites across the United States and Canada. The data used for the present study, which are freely available to subscribers and continually updated, were downloaded on December 9, 2011.

### Participants in Current Study

The sample for this current study was comprised of 229 normal controls, 337 patients with MCI, and 193 patients with probable AD. Detailed inclusion and exclusion criteria for ADNI can be found at: (<http://www.adniinfo.org/Scientists/ADNIGrant/ProtocolSummary.aspx>). The dataset includes participants between 55–90 years of age who had a study partner able to provide an independent evaluation of functioning and spoke either English or Spanish. Those who were taking specific psychoactive medications were excluded. Additionally, participants were willing and able to undergo all test procedures and agreed to longitudinal follow up. Further basis for study inclusion and exclusion are as follows: healthy controls had to have an MMSE score between 24–30, CDR of 0.0, were not depressed, and cognitively intact. Mild Cognitive Impairment patients had to have an MMSE score between 24–30, CDR of 0.5, subjective memory complaint, objective memory loss as measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, absence of significant levels of impairment in other cognitive domains, and no dementia. Mild Alzheimer's disease: had to have an MMSE score between 20–26, CDR of 0.5 or 1.0; and met National Institute of Neurological and Communicative Diseases and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD (McKhann et al., 1984).

### Group Assignment Based on Clinical Dementia Rating Scale - Sum of Boxes (CDR-SOB)

The CDR is a semi-structured interview developed to provide a global rating dementia severity and it is useful for staging and tracking decline in AD (Fillenbaum, Peterson, & Morris, 1996; Morris, 1997; Morris et al., 1997). The CDR assesses six domains of cognitive and daily functioning (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care). Each domain in the CDR is scored on an ordinal scale (0 = no impairment, 0.5 = questionable impairment, 1 = mild impairment, 2 = moderate impairment, 3 = severe impairment), and the ratings are summed to create a global estimate of dementia severity (CDR-sum of box score, theoretical range = 0–18). We performed a median split across all participants, based on the global Clinical Dementia Rating Scale – sum of boxes score (CDR-SOB). This created two groups: those who were less clinically impaired and more clinically impaired. Based on the CDR scoring criteria, we chose the global CDR-SOB score of 1.5 as the cut off for inclusion into the “more clinically impaired” group.

We did not test measurement invariance among normal, MCI, and AD diagnostic groups. Correlations among many cognitive test scores in the overall sample were absent in

diagnostic- specific groups due to the problem of subsetting on the outcome, which was described earlier. Dividing the sample of older adults with varying degrees of cognitive impairment into groups defined by the CDR is an approach that has been used in other studies (e.g., Jones & Ayers, 2006; Hayden et al., 2011) and provided just enough overlap to minimize this problem.

## Neuropsychological Assessment

The ADNI baseline neuropsychological assessment procedures have been previously described elsewhere (Mueller et al., 2005). Preliminary analysis of the selected neuropsychological tests used in this study, included examination of variable distributions and their intercorrelations. For the confirmatory factor analysis, 23 variables of interest were included and are described below. Attempts to include all available neuropsychological test scores were made and were a priori categorized into five cognitive domains by consensus among a team of neuropsychologists (Figure 1).

## Memory

**ADAS-Cog Word Recall and Recognition Test**—In the word recall task, participants read a list of 10 high-frequency nouns over three trials and were asked to recall as many words as possible after each trial. These immediate recall trials are followed by brief and delayed recall trials and a recognition test. The variables used for this project were delayed recall and recognition memory total number correct. The ADAS-Cog trial sum recall was considered for the present study but not included because ADAS-Cog word recall was poorly correlated with other memory measures, including AVLT recall.

**Rey Auditory Verbal Learning Test (AVLT)**—The AVLT is a test of episodic verbal memory that assesses the ability to acquire a list of 15 words over the course of five trials (Rey, 1964). The test includes a short-delay recall trial presented after a distracter list, a 30-minute long delay recall trial, and a yes/no recognition trial following the delayed recall trial. A learning score can be calculated from the AVLT using the difference between the last and first immediate recall trials. The variables used for this project were the learning score (trial 5 minus trial 1), short and long delay recall, and recognition.

## Language

**ADAS-Cog Naming Test**—In the Naming test, participants are asked to name twelve objects and the fingers on their dominant hand. The total number of items spontaneously named served as the variable of interest, which was dichotomized into 0 (difficulty) and 1 (no difficulty) due to its non-normal distribution.

**Boston Naming Test**—The 30-item (odd numbered items) version of the Boston Naming Test was administered to assess confrontation naming ability (Kaplan, Goodglass, & Weintraub, 1983). Participants are presented with a series of line drawings ranging from high to low frequency items and are given 20 seconds to spontaneously generate the name of the picture. If the participant has perceptual difficulties that preclude them from coming up with the correct answer, they are given a stimulus cue. A phonemic cue is provided when the participant can recognize the purpose of the object, but cannot retrieve the correct name. The variable used in our analysis was the sum of correct spontaneous responses and correct responses following stimulus cues.

**Category Fluency**—This is a test of one's ability to spontaneously generate a set of semantically related words in one minute (Harrison, Buxton, Husain, et al., 2000). In two

separate one-minute trials, participants were asked to name as many different animals in the first trial, and then as many different fruits as possible in the second.

## Attention

**WMS-R Digit Span**—The first component of this test is digit span forward, which is a test of verbal attention. Participants read a string of numbers in increasing length and are then asked to repeat them. The second component of this test is digit span backward, which is a test of working memory. During this test, the participants read a string of numbers with increasing length, and are asked to recite the numbers in the reverse order. The variables used were digit span forward and backward length.

## Executive Function/Processing Speed

**Trail Making Test**—The Trail Making Test is comprised of two parts, A and B (Reitan & Wolfson, 1985). Part A is a test of psychomotor processing speed and visual scanning. Participants are presented with an array of numbers on a page and asked to draw lines connecting the numbers in sequential order within the allotted time limit. Part B is a test of psychomotor processing speed, visual scanning, and attentional set-shifting (i.e., executive function). Participants are presented an array of numbers and letters and they are instructed to draw connecting lines while alternating between numbers and letters in sequential order. Time to completion from Parts A and Part B minus Part A were used in analyses.

**ADAS-Cog Number Cancellation Test**—The number cancellation task is a test of visual attention and psychomotor processing speed. Participants are given 45 seconds to cross out specific numbers mixed in with other numbers on several lines. Total number of correct targets was the variable used for this study.

## Visuospatial Functioning

**ADAS-Cog Construction Praxis Test**—The construction praxis test is a test of visuospatial functioning that assesses participants' ability to copy four geometric figures that include a circle, a rhombus, a diamond and a 3-dimensional cube. Total constructional praxis score was used for this study.

**Clock Drawing Test**—Participants are asked to draw the face of a clock and to “set the hands to ten after eleven.” They are scored on the symmetry of number placement, correctness of numbers, the presence of two hands, and hand placement. For the purposes of our study, we used the total clock drawing score.

## Statistical Analyses

A series of four two-group CFAs were used to evaluate the stability of the factor structure in the ADNI neuropsychological test battery across the less clinically impaired and more clinically impaired participants. Measurement invariance is tested through a series of hierarchical models with progressively stricter constraints placed on the factor structure (Bontempo & Hofer, 2006; Boorsbom et al., 2008). Configural invariance tests if the two groups have the same set of factors, and is met by having good fits from a model in which the same common factors exist in each group. In metric invariance, factor loadings, or the slopes relating test scores to their underlying factors, are constrained to be equal across groups to test whether constructs are measured in the same way across group. The next step is to test scalar or strong invariance, which entails constraining factor loadings and intercepts to be the same across both groups, but allows means of latent factors to vary over group. The intercepts are fixed to be equal across groups because item means, and subsequently factor means, are allowed to vary across the less and more impaired groups.

Finally, strict invariance restricts factor loadings, intercepts, and further constrains the residual variances that are constrained to test whether the model explains each indicator equally well across group.

We did not test structural invariance. Structural invariance tests whether the means of the latent variables are similar across group and whether the correlations amongst latent variables are the same across groups. As participants at different levels of cognitive function will most certainly show different levels of a construct such as memory or visuospatial function (e.g., mean differences), and because each cognitive domain is differentially affected by disease processes (e.g., correlations between constructs), we did not expect the cognitive domains to be structurally invariant across levels of clinical severity.

Models were estimated by minimizing the weighted least-squares with mean and variance adjustment (WLSMV) in MPlus software (version 6.11, Muthen & Muthen, 2010). Model fit was evaluated using the root mean square error of approximation (RMSEA; Steiger, 1989) and comparative fit index (CFI; Hu & Bentler, 1999) which captures the relative goodness of fit. We considered RMSEA below 0.05 and CFI above 0.90 to be indicators of good model fit (Hu & Bentler, 1999), based on previous work studying the measurement invariance of neuropsychological batteries with older adults (Mungas, Widaman, Reed, & Farias, 2011; Pedrazza et al., 2005). Nevertheless, we acknowledge that they are somewhat arbitrary markers along a continuum and that there are other ways to assess model fit. Figure 1 provides a model diagram for the hypothesized factor structure. This factor structure was derived through consensus agreement by a panel of neuropsychological experts in consultation with empirical data distributions. Memory was represented by immediate trial learning, short delay recall, long delay recall, and recognition from the AVLT and by delayed recall and recognition from the ADAS-Cog list-learning task. Correlations among AVLT items were accommodated with methods correlations, which improved model fit by relaxing conditional independence assumptions. Methods correlations are correlations among items from the same test that may not represent meaningful variance of a construct (Garner, Hake, & Eriksen, 1956; Strauss, Thompson, Adams, Redline, & Burant, 2000) and across all invariance testing steps for the current project, methods correlations were constrained to be the same across groups. Language ability was measured using animal and vegetable recall from the Verbal Fluency Test, spontaneous recall from the Boston Naming Test, and ADAS-Cog Naming. Similar to memory, a methods correlation accommodated intercorrelations among measures in the Verbal Fluency Test. The construct representing executive function and processing speed was measured using Trails A, Trails B-A, digit symbol coding, and ADAS number cancellation. Scores at or above the ceiling level (2.5 minutes for Trails A, 5 minutes for Trails B) were categorized as missing because a respondent's true ability was unobserved (Mueller et al., 2005). Additionally, ceiling effects can induce spurious correlations (Austin, 2003). The top and bottom 1 percent of Trails A scores were winsorized to prevent outliers and voids, which can lead to Heywood cases in factor analysis (Barnett & Lewis, 1994; Heywood, 1931). Timed tests (e.g., Trails B-A, Trails A) were converted to z-scores. A summation of errors committed was initially considered as part of the executive function composite, however the resulting variable displayed poor psychometric properties. Attention was indicated by the digit span forward and backward tasks. Visuospatial functioning was indicated by clock copy scores, clock score, and ADAS-Cog construction score. All variables were coded so that higher numbers indicate better performance.

## RESULTS

### Baseline Demographics and Neuropsychological Performance

The CDR-SOB median split (CDR-SOB = 1.5) produced 389 participants who were included in the less clinically impaired (CDR-SOB<1.5) and 430 participants in the more clinically impaired group (CDR-SOB>=1.5). Demographic information is presented in Table 1. No cognitively normal individuals were classified as more clinically impaired, conversely, only 1 mild AD patient was included in the less clinically impaired group. Among the MCI participants, 40% were classified into the less clinically impaired group. Participants were mostly male (58%) and white (93%), and the mean age was approximately 75 years. Using Cohen's (1988) criteria, there were trivial to small differences between groups with respect to sex, race, education, and age (Table 1). As expected, those who were less impaired performed better across all cognitive measures in the CFA model than those who were more functionally impaired (Table 1).

### Confirmatory Factor Analysis

Results of invariance testing are summarized in Table 2. The clinical group and type of invariance tested is provided at the top of each column. Fit statistics as well as unstandardized parameter estimates (standardized estimates are in parentheses) for the models are provided in Table 2. Prior to multiple-group invariance testing, we first estimated a single-group model in the full sample; model fit was excellent (CFI: 0.955; RMSEA: 0.039). Multiple group CFA was used in a series of analyses to evaluate the measurement invariance across both groups. The first model tested configural invariance and the model fit was excellent (RMSEA=.042; CFI=.941;  $\chi^2 = 477.8$ ,  $df=280$ ), suggesting that the five factor model was shared across groups.

The next model tested the equivalence of factor loadings, which were constrained across group to test metric invariance. Although the relative fit significantly worsened with metric invariance ( $\chi^2$ difference=163.1;  $df=19$ ;  $p<.001$ ), the absolute fit was still excellent (RMSEA=.053; CFI=.898;  $\chi^2 = 640.8$ ,  $df=299$ ) (Table 2). Model modification indices revealed the largest source of misfit arose from constraining memory factor loadings, particularly for AVLT short and long delay recall, to be equal across group. When we tested partial metric invariance by allowing short and long delay AVLT recall to vary by group, model fit was similar to the model for configural invariance (RMSEA=.042; CFI=.937;  $\chi^2=507.4$ ,  $df=297$ ; versus configural invariance:  $\chi^2$ difference=29.7,  $df=17$ ,  $p=0.03$ ). In addition, correlations among the factors did not change considerably from the model with configural invariance. Analysis of correlations among cognitive domains suggests that memory and language are the most highly correlated domains (0.70) in the less clinically severe group while the visuospatial and executive/speed were most highly correlated domains (0.90) in the more clinically severe group (Table 3). Attention was least correlated with other constructs.

Fit statistics for the model testing scalar invariance were good (RMSEA=.051; CFI=.900;  $\chi^2 =645.8$ ,  $df=312$ ;  $\chi^2$ difference=5.2;  $df=13$ ;  $p=.98$ ). Although the RMSEA was slightly above the cut off of .05 however, there was no significant change between metric to scalar invariance and we considered the fit acceptably close to being excellent. Additionally, rejecting the model of scalar invariance solely on the basis the RMSEA value would lead to the conclusion that the two models are significantly different, which would be contrary to the actual test comparing the two models. All factor loadings were significantly correlated with their respective factor scores. Additionally, the intercepts as well as factor loadings were constrained to be equal over groups, and the difference in level apparent in cognitive tests in Table 1 could be explained by differences in means of latent factors. The model of

scalar invariance produces estimated means for memory ( $z=-1.71$ ), visuospatial function ( $z=-0.88$ ), language ( $z=-1.12$ ), executive function ( $z=-1.10$ ), and attention ( $z=-0.71$ ) that are each significantly lower in the group with more clinical impairment than those with less clinical impairment, for which means were constrained to be 0. Despite clear differences in mean levels of observed indicators across groups evident in Table 1, we observed scalar invariance because allowing for differences in means of the latent constructs was able to account for differences in level. In other words, although all test scores were lower in the more impaired group, no test was differentially lower relative to other tests, so that a difference in the latent variable mean was able to account for test differences. Thus, scalar invariance was upheld.

The test of strict invariance assessed whether the same factor loadings, intercepts, and residual variances were equivalent across the less and more clinically significantly impaired groups, and the overall model fit for strict invariance remained moderate (RMSEA=.059; CFI=.859;  $\chi^2 = 802.632$ ,  $df=330$ ). However, compared to the model testing scalar invariance, the RMSEA and CFI for the fourth model fell outside of the predefined limits as indicated by a nested  $\chi^2$  test with significantly worse relative fit ( $\chi^2$  difference=156.8,  $df=18$ ,  $p<0.001$ ) and suggests that residual variation in cognitive tests differs across levels of impairment. Taken together, these results indicate that configural, metric, and scalar invariance, but not strict invariance were met for the CFA across levels of impairment.

## DISCUSSION

The present study tested the measurement invariance of the ADNI Neuropsychological Battery variables across two discrete levels of clinical dementia severity in older adults with cognitive impairment. In summary, the resulting CFA in this study confirmed the emergence of five dimensions in the ADNI battery (i.e. memory, visuospatial functioning, attention, executive function/processing speed, and language), and the findings are consistent with other CFA studies of neuropsychological test batteries in older adults (Dowling et al., 2010; Hayden et al., 2011; Pedrazza et al., 2005; Tuokko et al., 2009 and Siedlecki 2008). These factors and the latent structure of the five cognitive constructs were invariant across the spectrum of cognitive ability which ranges from normal aging to dementia. In our study, there were significant differences in the average levels of performance among cognitively normal individuals, MCI, and AD; however, the results from the CFA indicates that cognition is organized in the same way across the range impairment present in these groups. As such, the summary scores generated from these factors can be used as descriptive markers of cognitive decline and subsequently, disease progression in ADNI participants.

The results of our models indicated that configural, metric, scalar, but not strict measurement invariance were met. The lack of strict measurement invariance suggests that the amount of unexplained residual variance, or error variance, in some cognitive tests varies as a function of functional severity. Thus, some tests may describe performance in one population better than in another. Post-hoc inspections of residual variances in the different groups revealed that, in general, levels of unexplained variability in memory, language, and attention items were greater in the less functionally impaired group than in the more functionally impaired group. The opposite pattern was found for visuospatial and executive functioning/processing speed items. This pattern of findings suggests some cognitive tests may provide more useful information about their respective domains in certain subgroups than in other groups.

The focus of ADNI is to identify early clinical and biological markers of AD and as such, the neuropsychological battery emphasizes a comprehensive memory assessment by utilizing several different measures of learning and recall. Although memory performance is

often used as an outcome measure in studies of neurodegenerative disease, other domains such as executive function are also useful markers of progression to dementia (Drijgers, et al., 2011; Pereira, Yassuda, Oliveira, & Forlenza, 2008; Rozzini et al., 2007). Executive functioning is a common term that encompasses a wide variety of higher-order abilities that include problem-solving, set-shifting, generation, and rule-monitoring (Alvarez & Emory, 2006), that can be measured by a variety of different paradigms. Indicators from the Trail Making Test were included in the final CFA model to serve as measures of executive functioning. Given the limited number of executive function measures in the ADNI battery, we also considered including verbal fluency in this domain since other studies have found this test to load with other executive function measures (Mungas, Widaman, Reed, & Farias, 2011). However, the correlation matrix in the current study indicates that verbal fluency was most highly correlated with language and memory, which has also been shown in other studies (Siedlecki et al., 2008).

Several limitations to the current results are worth noting. First, as Peterson and colleagues (2010) acknowledge, ADNI participants were selected to resemble a clinical trial sample, and study recruitment efforts focused on persons at risk for memory decline and prodromal AD. Thus, results may not generalize to impairment in non-Alzheimer's dementias and more severely impaired AD samples. Similarly, this study's findings may not generalize to community-living populations of older adults with more diverse ethnic, language, and educational backgrounds. A further limitation of our study is that the measurement invariance of constructs across levels of clinical impairment may not generalize to constructs or cognitive tests that were not part of the ADNI neuropsychological battery. It is possible that other test variables would not be invariant. Third, we report in this study measurement invariance of one factor structure that we hypothesized a priori. Consideration of competing models in future work is needed to provide stronger support for the findings. However, we note that our findings with respect to measurement invariance are consistent with findings from other studies (Dowling et al., 2010; Hayden et al., 2011; Pedrazza et al., 2005; Tuokko et al., 2009 and Siedlecki 2008).

In summary, the findings from this current study demonstrate that the ADNI Neuropsychological Battery can be summarized into distinct cognitive domains that are stable across levels of functional impairment in older adults who are cognitively normal or diagnosed with MCI or AD. Although the current study divided the sample into two groups based on the participants' functional impairment, there is utility in studying the measurement invariance of these factors in diagnostic groups created by neuropsychological tests that are different from those used in the CFA. Investigating the relationships among these factor scores and other biomarkers of AD in ADNI could yield more meaningful relationships than individual cognitive tests or summary scores. In addition, studying the measurement invariance of the ADNI Neuropsychological Battery across time and disease progression has yet to be explored. Testing measurement invariance of the ADNI Neuropsychological Battery over time might enable summary scores to be used as surrogate markers of AD pathology in future studies with ADNI participants.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We gratefully acknowledge a conference grant from the National Institute on Aging (R13AG030995, PI: Dan Mungas) that facilitated data analysis for this project.

Dr. Park was supported by a grant from the National Institute of Aging (R01 AG031252 PI: Sarah Farias). Dr. Gross was supported by a National Institutes of Health Translational Research in Aging fellowship (AG023480) and a grant from the National Institute on Aging (AG031720, PI: Sharon Inouye). Dr. McLaren was supported by National Institute on Aging grants AG036694 (PI: Reisa Sperling) and AG027171 (PI: Alireza Atri). Dr. Pa was supported by the National Institute on Aging (K01 AG034175, PI: Dr. Pa). Dr. Johnson was supported by a grant from the National Institute of Aging grant AG022538 (PI: Johnson). Dr. Manly was supported by National Institute on Aging grants AG028786 (PI: Manly) and AG037212 (PI: Mayeux).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfex Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

The contents do not represent the views of the Dept. of Veterans Affairs, the United States Government, or any other funding entities.

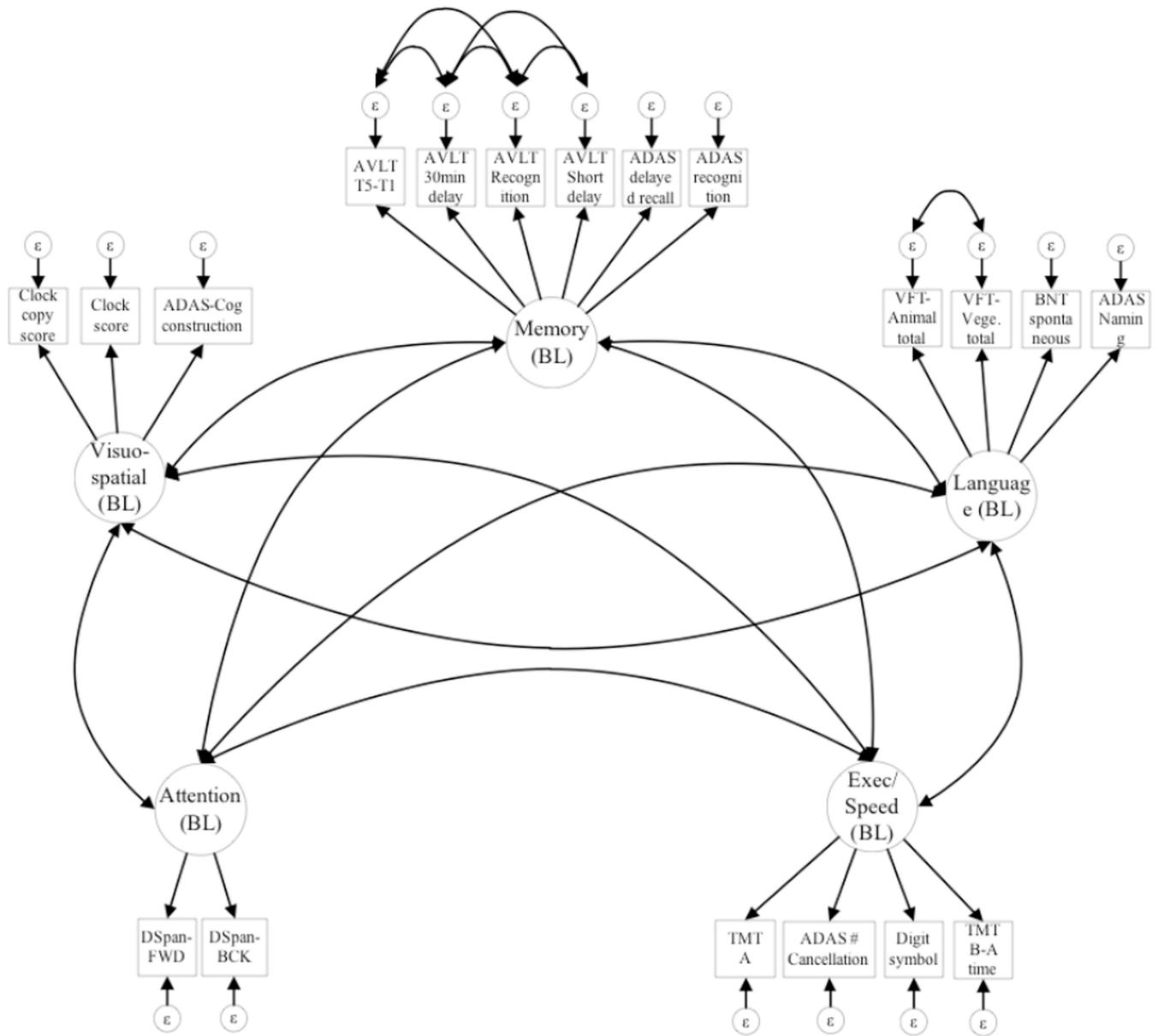
## REFERENCES

- Albert, M. The neuropsychology of the development of Alzheimer's disease. In: Craik, FIM.; Salthouse, TA., editors. *The Handbook of Aging and Cognition*. 4th ed. London: Academic Press; 2008. p. 97-132.
- Alvarez JA, Emory E. Executive function and the frontal lobes: a meta-analytic review. *Neuropsychol Rev*. 2006; 16(1):17-42. [PubMed: 16794878]
- Austin PC. Type I Error Inflation in the Presence of a Ceiling Effect. *The American Statistician*. 2003; 57:97-104.
- Barnett; Lewis. *Outliers in Statistical Data*. 3rd. Ed.. John Wiley and Sons; 1994.
- Binet, Alfred. *L'Annee Psychologique*. 1905; 12:191-244.
- Bontempo, DE.; Hofer, SM. Assessing factorial invariance in cross-sectional and longitudinal studies. In: Ong, AD.; van Dulmen, M., editors. *Handbook of methods in positive psychology*. New York, NY: Oxford University Press; 2007. p. 153-175.
- Borsboom D, Romeijn JW, Wicherts JM. Measurement invariance versus selection invariance: Is fair selection possible? *Psychological Methods*. 2008; 13:75-98. [PubMed: 18557679]
- Brookmeyer R, Evans DA, Hebert L, Langa KM, Heeringa SG, Plassman BL, et al. National estimates of the prevalence of Alzheimer's disease in the United States. *Alzheimers Dement*. 2011; 7(1):61-73. [PubMed: 21255744]
- Buckner RL. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. *Neuron*. 2004; 44(1):195-208. [PubMed: 15450170]
- Cargin JW, Maruff P, Collie A, Shafiq-Antonacci R, Masters C. Decline in verbal memory in non-demented older adults. *J Clin Exp Neuropsychol*. 2007; 29(7):706-718. [PubMed: 17891680]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd. Hillsdale, NJ: Erlbaum; 1988.
- Collette F, Van der Linden M, Bechet S, Salmon E. Phonological loop and central executive functioning in Alzheimer's disease. *Neuropsychologia*. 1999; 37(8):905-918. [PubMed: 10426516]
- Crane PK, Narasimhalu K, Gibbons LE, Mungas DM, Haneuse S, van Belle G. Item response theory facilitated calibrating cognitive tests and reduced bias in estimated rates of decline. *Journal of Clinical Epidemiology*. 2008; 61:1018-1027. e9. [PubMed: 18455909]

- DeCarli C, Mungas D, Harvey D, Reed B, Weiner M, Chui H, et al. Memory impairment, but not cerebrovascular disease, predicts progression of MCI to dementia. *Neurology*. 2004; 63(2):220–227. [PubMed: 15277612]
- 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. *J Int Neuropsychol Soc*. 2003; 9(6):936–946. [PubMed: 14632252]
- Desgranges B, Eustache F, Rioux P, de La Sayette V, Lechevalier B. Memory disorders in Alzheimer's disease and the organization of human memory. *Cortex*. 1996; 32(3):387–412. [PubMed: 8886519]
- 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(6):742–756. [PubMed: 21038965]
- Drijgers RL, Verhey FR, Leentjens AF, Kohler S, Aalten P. Neuropsychological correlates of apathy in mild cognitive impairment and Alzheimer's disease: the role of executive functioning. *Int Psychogeriatr*. 2011; 23(8):1327–1333. [PubMed: 21729413]
- Evans MC, Barnes J, Nielsen C, Kim LG, Clegg SL, Blair M, et al. Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. *Eur Radiol*. 2010; 20(3):674–682. [PubMed: 19760240]
- Fillenbaum GG, Peterson B, Morris JC. Estimating the validity of the clinical Dementia Rating Scale: the CERAD experience. Consortium to Establish a Registry for Alzheimer's Disease. *Aging (Milano)*. 1996; 8(6):379–385. [PubMed: 9061124]
- Gallagher M, Koh MT. Episodic memory on the path to Alzheimer's disease. *Current Opinion in Neurobiology*. 2011; 21(6) 929-3.
- Garner WR, Hake HW, Eriksen CW. Operationism and the concept of perception. *Psychol Rev*. 1956; 63(3):149–159. [PubMed: 13323170]
- Grossman M, Koenig P, Glosser G, DeVita C, Moore P, Rhee J, et al. Neural basis for semantic memory difficulty in Alzheimer's disease: an fMRI study. *Brain*. 2003; 126(Pt 2):292–311. [PubMed: 12538399]
- Hayden KM, Jones RN, Zimmer C, Plassman BL, Browndyke JN, Pieper C, et al. Factor structure of the National Alzheimer's Coordinating Centers uniform dataset neuropsychological battery: an evaluation of invariance between and within groups over time. *Alzheimer Dis Assoc Disord*. 2011; 25(2):128–137. [PubMed: 21606904]
- Heywood HB. 'On finite sequences of real numbers', *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1931; 134(824):486–501.
- Harrison JE, Buxton P, Husain M, Wise R. Short test of semantic and phonological fluency: normal performance, validity and test-retest reliability. *Br J Clin Psychol*. 2000; 39(Pt 2):181–191. [PubMed: 10895361]
- Hu L, Bentler PM. Fit indices in covariance structure modeling: sensitivity to under parameterized model misspecification. *Psychological Methods*. 1998; 3:424–453.
- Jalbert JJ, Daiello LA, Lapane KL. Dementia of the Alzheimer type. *Epidemiol Rev*. 2008; 30:15–34. [PubMed: 18635578]
- Jones SN, Ayers CR. Psychometric properties and factor structure of an expanded CERAD neuropsychological battery in an elderly VA sample. *Arch Clin Neuropsychol*. 2006; 21(4):359–365. [PubMed: 16766158]
- Kaplan, E.; Goodglass, H.; Weintraub, S. *The Boston Naming Test*. Philadelphia: Lea and Febiger; 1983.
- Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging*. 2011; 32(7):1207–1218. [PubMed: 19660834]
- Larrabee GJ, Kane RL, Schuck JR. Factor analysis of the WAIS and Wechsler Memory Scale: an analysis of the construct validity of the Wechsler Memory Scale. *J Clin Neuropsychol*. 1983; 5(2): 159–168. [PubMed: 6863562]
- Larrabee GJ, Curtiss G. Factor structure and construct validity of the Denman Neuropsychology Memory Scale. *Int J Neurosci*. 1985; 26(3–4):269–276. [PubMed: 4019053]

- Leow AD, Yanovsky I, Parikshak N, Hua X, Lee S, Toga AW, et al. Alzheimer's disease neuroimaging initiative: a one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. *Neuroimage*. 2009; 45(3):645–655. [PubMed: 19280686]
- MacDonald MC, Almor A, Henderson VW, Kempler D, Andersen ES. Assessing working memory and language comprehension in Alzheimer's disease. *Brain Lang*. 2001; 78(1):17–42. [PubMed: 11412013]
- MacPherson SE, Della Sala S, Logie RH, Wilcock GK. Specific AD impairment in concurrent performance of two memory tasks. *Cortex*. 2007; 43(7):858–865. [PubMed: 17941344]
- McGuinness B, Barrett SL, Craig D, Lawson J, Passmore AP. Executive functioning in Alzheimer's disease and vascular dementia. *Int J Geriatr Psychiatry*. 2010; 25(6):562–568. [PubMed: 19810010]
- 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. [Guideline Practice Guideline]. *Neurology*. 1984; 34(7):939–944. [PubMed: 6610841]
- Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int Psychogeriatr*. 1997; 9(Suppl 1):173–176. discussion 177–178. [PubMed: 9447441]
- Morris JC, Ernesto C, Schafer K, Coats M, Leon S, Sano M, et al. Clinical dementia rating training and reliability in multicenter studies: the Alzheimer's Disease Cooperative Study experience. *Neurology*. 1997; 48(6):1508–1510. [PubMed: 9191756]
- Muthen, LK.; Muthen, BO. *MPlus: User's Guide*. sixth. Los Angeles: Muthen & Muthen; 2010.
- Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement*. 2005; 1(1):55–66. [PubMed: 17476317]
- 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]
- Nestor PJ, Fryer TD, Hodges JR. Declarative memory impairments in Alzheimer's disease and semantic dementia. *Neuroimage*. 2006; 30(3):1010–1020. [PubMed: 16300967]
- Pedraza O, Lucas JA, Smith GE, Willis FB, Graff-Radford NR, Ferman TJ, et al. Mayo's older African American normative studies: confirmatory factor analysis of a core battery. *Journal of the International Neuropsychological Society*. 2005; 11(2):184–191. [PubMed: 15962706]
- Pereira FS, Yassuda MS, Oliveira AM, Forlenza OV. Executive dysfunction correlates with impaired functional status in older adults with varying degrees of cognitive impairment. *Int Psychogeriatr*. 2008; 20(6):1104–1115. [PubMed: 18752698]
- Petersen RC, Jack CR Jr, Xu YC, Waring SC, O'Brien PC, Smith GE, et al. Memory and MRI-based hippocampal volumes in aging and AD. *Neurology*. 2000; 54(3):581–587. [PubMed: 10680786]
- Pike KE, Rowe CC, Moss SA, Savage G. Memory profiling with paired associate learning in Alzheimer's disease, mild cognitive impairment, and healthy aging. *Neuropsychology*. 2008; 22(6):718–728. [PubMed: 18999345]
- Pike KE, Savage G. Memory profiling in mild cognitive impairment: can we determine risk for Alzheimer's disease? *J Neuropsychol*. 2008; 2(Pt 2):361–372. [PubMed: 19824171]
- Ready RE, Ott BR, Grace J. Validity of informant reports about AD and MCI patients' memory. *Alzheimer Dis Assoc Disord*. 2004; 18(1):11–16. [PubMed: 15195458]
- Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychol Bull*. 1993; 114(3):552–566. [PubMed: 8272470]
- Reitan, RM.; Wolfson, D. *The Halstead-Reitan neuropsychological test battery*. 2 ed.. Tucson, AZ: Neuropsychology Press; 1985.
- Rey, A. *L'examen clinique en psychologie [clinical examination in psychology]*. Paris France: Presses Univeritaires de France; 1964.

- Rozzini L, Chilovi BV, Conti M, Bertolotti E, Delrio I, Trabucchi M, et al. Conversion of amnesic Mild Cognitive Impairment to dementia of Alzheimer type is independent to memory deterioration. *Int J Geriatr Psychiatry*. 2007; 22(12):1217–1222. [PubMed: 17562522]
- 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]
- Spaan PE, Raaijmakers JG, Jonker C. Alzheimer's disease versus normal ageing: a review of the efficiency of clinical and experimental memory measures. *J Clin Exp Neuropsychol*. 2003; 25(2): 216–233. [PubMed: 12754679]
- Steiger, JH. *EzPath: A supplementary module for SYSTAT and SYGRAPH*. Evanston IL: SYSTAT; 1989.
- Stern Y, Albert S, Tang MX, Tsai WY. Rate of memory decline in AD is related to education and occupation: cognitive reserve? *Neurology*. 1999; 53(9):1942–1947. [PubMed: 10599762]
- Strauss ME, Thompson P, Adams NL, Redline S, Burant C. Evaluation of a model of attention with confirmatory factor analysis. *Neuropsychology*. 2000; 14(2):201–208. [PubMed: 10791860]
- Thompson CL, Henry JD, Withall A, Rendell PG, Brodaty H. A naturalistic study of prospective memory function in MCI and dementia. *Br J Clin Psychol*. 2011; 50(4):425–434. [PubMed: 22003951]
- Tuokko HA, Chou PH, Bowden SC, Simard M, Ska B, Crossley M. Partial measurement equivalence of French and English versions of the Canadian Study of Health and Aging neuropsychological battery. *J Int Neuropsychol Soc*. 2009; 15(3):416–425. [PubMed: 19402928]
- Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation; 1981.
- Wechsler, D. *Manual for the Wechsler Memory Scale-Revised*. San Antonio, TX: The Psychological Corporation; 1987.
- Welsh K, Butters N, Hughes J, Mohs R, Heyman A. Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Arch Neurol*. 1991; 48(3):278–281. [PubMed: 2001185]
- Widaman KF, Ferrer E, Conger RD. Factorial Invariance within Longitudinal Structural Equation Models: Measuring the Same Construct across Time. *Child Dev Perspect*. 2010; 4(1):10–18. [PubMed: 20369028]
- Yuspeh RL, Vanderploeg RD, Crowell TA, Mullan M. Differences in executive functioning between Alzheimer's disease and subcortical ischemic vascular dementia. *J Clin Exp Neuropsychol*. 2002; 24(6):745. [PubMed: 12424649]



**Figure 1. Confirmatory Factor Analysis Model of the Baseline ADNI Neuropsychological Battery** Latent variables (shown in circles) were formed from sets of observed neuropsychological tests (shown in boxes), all of which have unique error terms  $\epsilon$ . Correlations between latent variables are shown with curved double-headed arrows.

**Table 1**

Baseline demographic characteristics and cognitive performance: Results from ADNI (n=819)

	Less impaired (n=430)	More impaired (n=389)	Cohen's d effect size	P-value for group differences
Age, Mean (SD)	75.8 (6)	75 (7.6)	0.1	0.099
Years of education, <i>Mean (SD)</i>	15.9 (3)	15.2 (3.1)	0.2	<0.001
Sex, Female, <i>n (%)</i>	163 (41.9)	179 (41.6)	0.0	0.937
Race, White, <i>n (%)</i>	358 (92)	404 (94)	0.1	0.199
Mini-Mental State Exam, <i>Mean (SD)</i>	28.4 (1.6)	25.3 (2.6)	1.4	<0.001
Baseline diagnostic status, <i>n (%)</i>				<0.001
Health control	229 (58.9)	0 (0)		
Mild cognitive impairment	159 (40.9)	238 (55.3)		
Alzheimer's disease	1 (0.3)	192 (44.7)		
AVLT Learning (Trial 5 - Trial 1), <i>Mean (SD)</i>	4.9 (2.6)	2.5 (2.1)	1.0	<0.001
AVLT Long Delay (30min), <i>Mean (SD)</i>	5.9 (4.1)	1.6 (2.6)	1.3	<0.001
AVLT Recognition, <i>Mean (SD)</i>	11.8 (3.3)	8.4 (4)	0.9	<0.001
AVLT Short Delay, <i>Mean (SD)</i>	6.6 (3.8)	2.6 (2.6)	1.2	<0.001
ADAS Delayed Recall, <i>Mean (SD)</i>	6 (2.4)	2.6 (2.2)	1.5	<0.001
ADAS Recognition, <i>Mean (SD)</i>	8.7 (2.6)	6.4 (2.9)	0.8	<0.001
Clock Score, <i>Mean (SD)</i>	4.5 (0.8)	3.8 (1.2)	0.7	<0.001
Clock Copy Score, <i>Mean (SD)</i>	4.8 (0.5)	4.6 (0.7)	0.3	<0.001
ADAS Construction, <i>Mean (SD)</i>	4.6 (0.5)	4.3 (0.6)	0.5	<0.001
ADAS Naming, Correct, <i>n (%)</i>	335 (86.1)	290 (67.4)	0.5	<0.001
Verbal Fluency Test-Animal total, <i>Mean (SD)</i>	18.3 (5.6)	14.2 (5.2)	0.8	<0.001
Verbal Fluency Test-Vegetables, <i>Mean (SD)</i>	13.2 (4.2)	9.3 (3.6)	1.0	<0.001
Boston Naming Test, spontaneous recall, <i>Mean (SD)</i>	26.5 (3.4)	24.3 (3.8)	0.6	<0.001
Trails A, Mean (SD)	5.1 (0.6)	4.7 (0.8)	0.6	<0.001
ADAS Number Cancellation, <i>Mean (SD)</i>	23.8 (5.6)	19.7 (6.6)	0.7	<0.001
Digit Symbol, <i>Mean (SD)</i>	42.9 (11)	33.1 (11.1)	0.9	<0.001
Trails B-A time, <i>Mean (SD)</i>	4.6 (0.8)	4 (1.1)	0.6	<0.001
Digit Span Backward, <i>Mean (SD)</i>	6.9 (2.2)	5.5 (2)	0.7	<0.001
Digit Span Forward, <i>Mean (SD)</i>	8.6 (2)	7.9 (2)	0.3	<0.001

Legend. SD=standard deviation.

**Table 2**

Measurement invariance across diagnostic group: Results from ADNI (n=819)

Indicator	Configural invariance		Metric invariance		Scalar invariance		Strict invariance	
	Less impaired (n=389)	More impaired (n=430)						
Memory								
AVLT trial learning (5-1)	1.7 (0.7)	1.3 (0.6)	1.6 (0.6)	1.6 (0.7)	1.5 (0.6)	1.5 (0.7)	1.5 (0.7)	1.5 (0.7)
AVLT 30min delay	3.4 (0.8)	1.8 (0.7)	2.4 (0.6)	2.4 (0.9)	2.4 (0.6)	2.4 (0.9)	2.4 (0.9)	2.4 (0.9)
AVLT Recognition	2.2 (0.7)	2.4 (0.6)	2.3 (0.7)	2.3 (0.6)	2.2 (0.7)	2.2 (0.6)	2.2 (0.6)	2.2 (0.6)
AVLT Short delay	3.2 (0.9)	1.8 (0.7)	2.4 (0.6)	2.4 (0.9)	2.4 (0.6)	2.4 (0.9)	2.4 (0.9)	2.4 (0.9)
ADAS delayed recall	1.9 (0.8)	1.9 (0.8)	1.8 (0.7)	1.8 (0.8)	1.8 (0.8)	1.8 (0.8)	1.8 (0.8)	1.8 (0.8)
ADAS recognition	1.3 (0.5)	1.6 (0.6)	1.4 (0.5)	1.4 (0.5)	1.4 (0.5)	1.4 (0.5)	1.4 (0.5)	1.4 (0.5)
Visuospatial functioning								
Clock copy score	0.3 (0.5)	0.4 (0.5)	0.3 (0.6)	0.3 (0.4)	0.3 (0.6)	0.3 (0.4)	0.3 (0.5)	0.3 (0.5)
Clock score	0.6 (0.7)	0.9 (0.8)	0.7 (0.8)	0.7 (0.6)	0.7 (0.8)	0.7 (0.6)	0.7 (0.8)	0.7 (0.8)
ADAS construction	0.3 (0.6)	0.3 (0.5)	0.3 (0.5)	0.3 (0.5)	0.3 (0.5)	0.3 (0.5)	0.3 (0.5)	0.3 (0.5)
Language								
VFT-Animal total	3.9 (0.7)	4.1 (0.8)	4.0 (0.7)	4.0 (0.8)	4.0 (0.7)	4.0 (0.8)	4.0 (0.7)	4.0 (0.7)
VFT-Vegetables total	2.9 (0.7)	2.9 (0.8)	2.9 (0.7)	2.9 (0.8)	3.0 (0.7)	3.0 (0.8)	3.0 (0.8)	3.0 (0.8)
BNT spontaneous	2.2 (0.6)	2.2 (0.6)	2.2 (0.6)	2.2 (0.6)	2.2 (0.6)	2.2 (0.6)	2.2 (0.6)	2.2 (0.6)
ADAS Naming	0.7 (0.7)	1.3 (0.6)	0.8 (0.8)	0.8 (0.5)	0.7 (0.7)	0.7 (0.6)	0.7 (0.7)	0.7 (0.6)
Processing speed, Executive function								
TMT B-A time	0.5 (0.6)	0.5 (0.5)	0.5 (0.7)	0.5 (0.5)	0.5 (0.7)	0.5 (0.5)	0.5 (0.6)	0.5 (0.6)
TMT-A	0.4 (0.6)	0.5 (0.6)	0.4 (0.6)	0.4 (0.5)	0.4 (0.6)	0.4 (0.5)	0.4 (0.6)	0.4 (0.6)
ADAS number cancellation	2.8 (0.5)	4.5 (0.7)	3.4 (0.6)	3.4 (0.5)	3.4 (0.6)	3.4 (0.5)	3.4 (0.6)	3.4 (0.6)
Digit symbol	9.7 (0.9)	9.1 (0.8)	9.1 (0.8)	9.1 (0.8)	9.1 (0.8)	9.1 (0.8)	9.1 (0.8)	9.1 (0.8)
Attention								
Digit span-FWD	1.2 (0.6)	1.0 (0.5)	1.0 (0.5)	1.0 (0.5)	1.0 (0.5)	1.0 (0.5)	1.0 (0.5)	1.0 (0.5)
Digit span-BCK	2.1 (1.0)	1.9 (1.0)	2.0 (0.9)	2.0 (1.0)	1.9 (0.9)	1.9 (1.0)	1.9 (1.0)	1.9 (1.0)
Model fit statistics								
RMSEA	0.042		0.053		0.051		0.059	
CFI	0.941		0.898		0.9		0.859	
$\chi^2$ , df	477.8, 280		640.8, 299		645.8, 312		802.6, 330	

Legend. Raw (standardized) factor loadings are shown. Standardized loadings are correlations between the observed indicator and latent factor. Raw loadings are more appropriate for inter-diagnostic group comparisons, while standardized loadings are most appropriate for intra-diagnostic group comparisons. The difference in degrees of freedom between configural and metric invariance is 19 because 19 factor loadings are constrained to be the same between configural and metric invariance. The difference in degrees of freedom between metric and scalar invariance is 13 because although indicator intercepts are fixed across class (19), we also allowed latent variable means to differ across group (minus 5). The degree of freedom difference between scalar and strict invariance is 18 because we constrained 18 residual variances to be equal across group (there are only 18 because ADAS naming was declared categorical and has no estimated residual variance).

**Table 3**

Correlations Among Constructs Assuming Scalar Invariance: Results from ADNI (n=819)

	(1)	(2)	(3)	(4)	(5)
Less impaired group (n=389)					
(1) Memory	1.0				
(2) Visuospatial functioning	0.5	1.0			
(3) Language	0.7	0.5	1.0		
(4) Speed, Executive	0.4	0.5	0.6	1.0	
(5) Attention	0.3	0.2	0.4	0.3	1.0
More impaired group (n=430)					
(1) Memory	1.0				
(2) Visuospatial functioning	0.4	1.0			
(3) Language	0.5	0.8	1.0		
(4) Speed, Executive	0.3	0.9	0.7	1.0	
(5) Attention	0.1	0.5	0.3	0.5	1.0

Legend. Correlations between latent factors for each diagnostic group.