



NIH PUBLIC ACCESS

Author Manuscript

Alzheimer Dis Assoc Disord. Author manuscript; available in PMC 2012 April 1.

Published in final edited form as:

Alzheimer Dis Assoc Disord. 2011 ; 25(2): 128–137. doi:10.1097/WAD.0b013e3181ffa76d.

Factor Structure of the National Alzheimer's Coordinating Centers Uniform Dataset Neuropsychological Battery: An evaluation of invariance between and within groups over time

Kathleen M. Hayden^{1,2}, Richard N. Jones³, Catherine Zimmer⁴, Brenda L. Plassman², Jeffrey N. Browndyke^{1,2,5}, Carl Pieper^{1,6}, Lauren H. Warren^{1,2}, and Kathleen A. Welsh-Bohmer^{1,2}

¹ Joseph and Kathleen Bryan Alzheimer's Disease Research Center, Duke University Medical Center, Durham, NC

² Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC

³ Aging Brain Center, Institute for Aging Research, Hebrew SeniorLife, Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston MA

⁴ The Odum Institute and Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁵ Duke institute for Brain Sciences, Duke University, Durham NC

⁶ Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC

Abstract

The neuropsychological battery from the National Alzheimer's Disease Coordinating Center (NACC) is designed to provide a sensitive assessment of mild cognitive disorders for multicenter investigations. Comprised of eight common neuropsychological tests (12 measures), the battery assesses cognitive domains affected early in the course of Alzheimer's disease (AD). We examined the factor structure of the battery across levels of cognition (normal, mild cognitive impairment (MCI), dementia) based on Clinical Dementia Rating (CDR) scores to determine cognitive domains tapped by the battery. Using data pooled from 29 NIA funded Alzheimer's Disease Centers, exploratory factor analysis was used to derive a general model using half of the sample; four factors representing memory, attention, executive function, and language were identified. Confirmatory factor analysis (CFA) was used on the second half of the sample to evaluate invariance between groups and within groups over one year. Factorial invariance testing included systematic addition of constraints and comparisons of nested models. The general CFA model had a good fit. As constraints were added, model fit deteriorated slightly. Comparisons within groups demonstrated stability over one year. In a range of cognition from normal to dementia, factor structures and factor loadings will vary little. Further work is needed to determine if domains become more or less distinct in severely cognitively compromised individuals.

Corresponding Author: Kathleen M. Hayden, Ph.D. DUMC, Joseph and Kathleen Bryan ADRC 2200 West Main Street, Suite A-200 Durham, NC 27705 Phone: 919-668-1584 Fax: 919-668-0828 khayden@duke.edu.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Factor Analysis; invariance; neuropsychological tests; dementia; Alzheimer's disease

Introduction

An emerging line of research seeks to identify biomarkers for individuals at-risk for Alzheimer's disease (AD) with imaging,¹ cerebrospinal fluid,² blood samples,³ and detailed neuropsychological testing.⁴ In dementia research, neuropsychological batteries were originally developed with the goal of facilitating diagnosis; more recently the focus has been on early diagnosis. In 1984, the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA),⁵ specified 7 domains of cognition that may be impaired in AD. Neuropsychological test batteries, designed to tap these areas of cognition, have been used to identify variations in cognitive performance such as declines in memory,⁶ that may represent early disease pathology⁷. Different cognitive domains presumably reflect regional brain functions that may be affected either as part of the normal aging process, through the course of a neurodegenerative disease, or some other underlying pathology.⁷ We do not know, however, whether or not cognitive domains change or shift over the course of disease, i.e., whether neuropsychological tests measure the same thing in normal individuals compared with those suffering from dementia, or if the variation is simply one of a general decline in performance reflecting a more global cognitive decline. In samples of individuals with mild impairment or AD, slightly different factor structures seem to emerge implying that there is variation between normals, mildly impaired, and AD patients' pattern of performance on tests.⁸⁻¹¹

Factor analysis is used to evaluate the construct validity of neuropsychological batteries and measures of invariance are used to gauge the stability of factor structures. Indications of stability or instability across groups have implications for researchers seeking to determine how cognitive domains are affected when comparing normal individuals to those suffering from neurodegenerative disease such as AD. The stability with which factor structures correspond to cognitive domains across groups and over time has not been studied extensively.

Understanding factor stability over the continuum of brain disease is important as we begin diagnosing neurodegenerative disease at earlier stages. Shifts in the factor structure between groups would suggest different relationships between the measures and may signal demise of particular underlying neural systems or cognitive processes that cut across multiple cognitive domains. Herein we evaluate the factorial invariance of the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS) neuropsychological battery to determine if the factor structure varies across groups or over time.

Methods

The NACC was established in 1999 in order to facilitate collection of standardized data from Alzheimer's Disease Centers (ADC's) across the United States. A neuropsychological battery, part of the UDS, comprising 8 tests (12 measures) was developed with the goal of tapping the following cognitive domains in mild cognitive impairment (MCI) and AD: attention, speed of processing, executive function, episodic memory, and language.¹² Tests in the battery (listed below) were selected to focus on markers of aging, MCI, and AD by building on tests already being administered by a majority of ADCs while keeping participant burden low.¹² All protocols and procedures were approved by the Institutional

Review Boards of each ADC. Informed consent was obtained from all participants. (For a more detailed description of the UDS battery and NACC methodology see Morris 2006,13 Beekley 2007,14 and Weintraub 2009.12)

Setting & Participants

Data were collected at 29 NIA funded ADC's using a standardized protocol. Participants are volunteers from the community; each ADC has its own recruitment protocol and participants are re-evaluated annually. The data consist of general demographic characteristics (age, sex, race, and education level in years), family history, health history, behavioral and functional assessments, and clinical information. As of May 5, 2008, the ADC's participating in NACC administered 14,428 UDS initial assessment batteries to participants across the United States.

The UDS Neuropsychological Battery

The UDS Neuropsychological Battery¹² consists of 8 neuropsychological tests which are focused on characterization of non-demented aging, mild cognitive impairment (MCI), and mild AD. The battery is fixed, i.e., administered in a standardized, uniform fashion at all ADCs. Designed to be brief and to cover major cognitive domains, the tests included in the battery are: Logical Memory Story A (Immediate and Delayed recall),¹⁵ the Boston Naming Test (30 item),¹⁶ 17 WAIS Digit Symbol,¹⁸ Trail Making Test Parts A & B,¹⁹ Digits Forward,¹⁵ Digits Backward,¹⁵ and semantic fluency (Animals,²⁰ and Vegetables) (Table 1). The Mini Mental State Examination (MMSE)²¹ was administered as a global indicator of dementia severity along with the Clinical Dementia Rating (CDR).²² 23

Clinical Dementia Rating

The CDR,²³ 24 administered to all participants as a part of the diagnostic process, takes into consideration decline from a prior level of function by rating 6 cognitive domains. Domains assessed include memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Although the CDR is used in the diagnostic process, the results of the neuropsychological battery are not specifically used to determine CDR scores.

Statistical methods

Factor Analysis—Factor analysis is a statistical technique that attempts to explain covariation among a set of observed variables by introducing unobserved (latent) variables that are presumed to be causes of the observed variables. Conditional on the unobserved variables, or factors, the observed variables are assumed to be uncorrelated. Observed variables “load” on factors with regression parameter estimates that are referred to as factor loadings. Because latent variables are unobserved, their scale is arbitrary so it is common to presume a unit normal distribution. There are a variety of different algorithms for estimating parameters of factor analysis models. In our study we used maximum likelihood parameter estimates as implemented in *MPlus* software (version 5.21, Muthén & Muthén, Los Angeles CA). Our analytic approach included both exploratory factor analysis (EFA; no *a priori* specification of latent factors and loading matrix is unconstrained) and confirmatory factor analysis (CFA; models are estimated given a fixed factor structure and the factor loading matrix contains many constraints).

Measures of Factorial Invariance—Construct validity can be assessed by evaluating the variability of the factor structure, or factorial invariance.²⁵ Each form of invariance described below is part of a hierarchy.²⁶ 27 Beginning with simpler forms of invariance, restrictions are added and the level of similarity or dissimilarity between groups is evaluated

by checking model fit¹¹ (see description of fit statistics below). The most basic form of invariance, dimensional invariance, is an indication of the general structure specified by each group and is present when the same number of common factors is identified in each group. Configural invariance criteria are met when the same items are associated with the same common factors in each group. Metric invariance, or weak factorial invariance, is achieved when adequate model fit is demonstrated while the factor loadings are held constant across groups. For scalar (strong) invariance, factor loadings and intercepts are held constant across groups. Finally, strict factorial invariance is met when factor loadings, intercepts, and residual variances are held constant across groups.²⁸ We performed tests of invariance using a multiple group analysis approach, where factor analytic models are estimated in separate groups simultaneously, and hypotheses about invariance are tested by adding constraints across groups and assessing differences in global indices of model fit.

Analytic Approach—The sample was subdivided according to CDR score rather than diagnoses because the results of the UDS neuropsychological battery are used in the diagnostic process. While the clinicians who assign CDR scores are not blinded to the results of the UDS neuropsychological battery, it is not a required source of information for staging disease severity. Individuals with CDR scores of 0.0 were considered cognitively normal. Those with scores of 0.5 were considered to have mild cognitive impairment (MCI), and those with scores greater than 0.5 were considered to have dementia. Because some of the tests in the battery typically yield skewed data (ceiling and floor effects), a Blom transformation²⁹ was applied to normalize the data.

Initial EFA were performed using one half of the sample in order to develop an empirical model. A simple structure CFA model was then derived from the pattern of results found in the EFA model. CFA was used to evaluate invariance between groups. We similarly evaluated the stability of the factors over one year by comparing two time points within each group. We did not consider individuals' group membership from one time point to the next because our focus was on the factor structure rather than individuals' progress over time. Descriptive data analyses were performed using SAS statistical software.³⁰ All factor analyses were completed using *MPlus* statistical software.³¹

Model comparisons and fit statistics—A hierarchical set of models were produced in CFA with successively more restrictive criteria to assess levels of invariance. These nested models were tested using Satorra-Bentler Scaled χ^2 test³² to determine if each successive model was significantly different from the prior one. This test corrects for the sensitivity of the χ^2 to large sample sizes by applying a scaling factor to correct for multivariate normality. A second test, the Bayesian Information Criterion (BIC) was also applied to differentiate nested models.³³ The BIC corrects for the χ^2 sensitivity to sample size by multiplying the degrees of freedom by the natural log of the sample size and subtracting the product from the χ^2 ; values less than zero represent models that have a better fit than a fully saturated model. Finally, Modification Indices (MI) provide estimates of the change in χ^2 value that would result if a given parameter was left unconstrained. MI can be inspected for sources of model misfit.

The configural invariance model, the least constrained model, was used as the starting point. Individual model fit was evaluated by examining the following fit statistics. The Comparative Fit Index (CFI)³⁴ captures the relative goodness of fit by comparing the model to the data while adjusting for complexity or parsimony³⁵ (better fit >0.90; range 0.0 -1.0). The Tucker-Lewis Index (TLI)³⁶ is an incremental fit index that makes a comparison between a null model and an incrementally more complex model³⁷ (better fit > 0.90; range 0.0 -1.0). These two indices are measures of relative fit, indicating improvement relative to a null model.³⁸ The Root Mean Square Error of Approximation (RMSEA)³⁹ is not very

sensitive to sample size and distribution and is a good measure of practical fit (better fit <0.05). The Standardized Root Mean Square Residual (SRMR)⁴⁰ is the standardized difference between variance and covariance and it is less sensitive to distribution and sample size (better fit <0.05). The RMSEA and the SRMR reflect the size of the residuals associated with the model.³⁸

Results

A total of 14,428 participants completed initial assessment batteries as of 5 May 2008. Individuals under the age of 55 ($n=517$), those who reported a primary language other than English (1,080), and individuals who were missing a significant portion (10 or more) of the test scores ($n=811$) were excluded. A total of 12,020 individuals provided sufficient information for inclusion in the analysis. Individuals who were not included in the sample tended to be younger (mean 70.3(14.2) vs. 75.6(8.9), $p<0.001$), female (59.8% vs. 57.2%, $p=0.02$), with lower levels of education (13.0(4.9) vs. 15.0(3.1), $p<0.001$), lower MMSE scores (20.0(9.8) vs. 25.5(5.4), $p<0.001$), and a greater proportion were non-white participants (25.2% vs. 16.7%, $p<0.001$).

Among those retained in the sample, individuals classified as normal were significantly younger than individuals classified as mildly cognitively impaired ($p<0.001$) and those suffering from dementia ($p<0.001$) (Table 2). Dementia cases were in turn older than the mildly impaired individuals ($p<0.001$). The three groups were significantly different from each other on all other demographic characteristics. A listing of the sample means and standard deviations for all the test scores can be found in Table 3. The mean scores for normal individuals listed here are in line with means reported in a prior evaluation of the battery;¹² differences may be attributed to exclusionary criteria imposed on our sample. The sample was randomly split into two subsamples of $n=6010$ each. The two samples were not statistically significantly different from each other in their demographic characteristics or test scores.

Initial exploratory factor analysis (EFA) was performed using half of the sample. Because the analysis was focused on factor structure and not data reduction, we considered model fit statistics rather than factor extraction procedures such as examination of scree plots⁴¹ or the application of the “Eigenvalue greater than one” rule.⁴² After reviewing the fit statistics (CFI = 0.999, TLI = 0.997, RMSEA = 0.019, SRMR = 0.003) we concluded that a 5-factor solution fit the data best.⁴³ EFA was repeated for each of the cognitive subgroups to determine if the general model was the same in each group. All three groups’ factor structures were in agreement based on visual inspection, thus meeting criteria for dimensional invariance.

Using the EFA factor structure, a confirmatory factor analysis (CFA) model was built (Figure 1). We modified the original model in CFA to improve our fit by collapsing the four digits tests into one factor and allowing for a correlation between Digits Forward/Digits Forward Length, and Digits Backward/Digits Backward Length. Factor loadings, intercepts, and residual variances were free to vary across group (CFI = 0.989; TLI = 0.985; RMSEA = 0.045; SRMR = 0.019). The sample was then stratified into three groups based on CDR scores and a multiple group CFA analysis was performed. The normal group consisted of $n=2409$ individuals defined as having a CDR global score ≥ 0.0 . A second group consisted of $n=2037$ individuals with MCI defined as having a CDR score ≥ 0.5 . The dementia group ($n=1564$) was defined as having a CDR score >0.5 . The initial multiple group CFA had only two constraints: variances were fixed at 1.0 and means were fixed at 0.0; this model had a good model fit, meeting criteria for configural invariance. There was a Heywood case (factor loading >1.0 ; negative variance)⁴⁴ in the dementia group Logical Memory

Immediate Recall factor loading. Evaluation of observed data suggests this phenomenon could be explained by a lack of variability in Immediate and Delayed recall responses due to floor effects in the dementia group (data not shown). Because the value in excess of 1.0 was small, as was the negative variance, and the model converged, we chose to proceed with the analysis⁴⁵ acknowledging that, as one would expect, performance on memory measures does not follow a linear pattern among dementia cases.

Factor loadings for each group and fit statistics are shown in Table 4, Model 1. This table lists each factor and corresponding tests in the left column. For each model, three sets of standardized factor loadings are presented, one for each group. The type of invariance measured and the parameter constraints of each model are listed across the top; fit statistics for each model are listed on the bottom. The fit statistics for Model 1 suggest configural invariance. Considering only the CFI and TLI, Models 1-4 met demonstrate configural, metric, strong (scalar), and strict factorial invariance. The values for the RMSEA and SRMR, however, were >0.05 for models 2, 3, and 4. Although model fit deteriorated with the addition of constraints, Model 3 (strong invariance) still fit better than a fully saturated model ($BIC = -70.33$) implying some level of invariance. The non-negative BIC value for Model 4, however, indicated that a fully saturated model would fit better. Inspection of modification indices (MI) suggested that cross-loadings such as Attention by Trails B, Attention by Logical Memory Immediate Recall, and Logical Memory Delayed Recall would improve model fit. When these modifications were added, the resulting Model 4 fit statistics were as follows: $CFI=0.968$, $TLI=0.968$, $RMSEA=0.056$, $SRMR=0.052$ and $BIC=-267.165$.

We next considered the stability of the factors over time at the group level. For this analysis, we compared each group at baseline to the first follow-up evaluation (about a 1 year lag, on average). This analysis was performed without regard for participants that may have progressed from one group to another because our focus was on the stability of the factor structure rather than individual performance. A total of 2,449 participants who were evaluated at the first annual follow-up were classified as normal. For the MCI group, a total of 1,514 were evaluated one year later as were 1,428 in the dementia group. Using the CFA model described above, we evaluated invariance over time for each group. In the three groups, all four levels of invariance criteria were met indicating remarkable stability in the factors over time within group (Tables 5a, 5b, and 5c).

Discussion

Our empirically derived EFA and CFA models were closely aligned with the theoretical models originally proposed in the development of the NACC UDS Battery (Table 1).¹² When tested for invariance, the factor structure met criteria for dimensional and configural invariance. Findings suggest the presence of metric, strong, and strict factorial invariance although some fit statistics were not ideal (i.e., RMSEA, SRMR). When model modifications allowed for cross loadings on the Attention factor by Trails B, LM Immediate, and LM Delayed, the result suggested a level of strict invariance. The implication is that acknowledging some cross loadings of test scores; the battery is relatively invariant across groups. That is, normal, MCI, and dementia subgroups, as defined by the CDR, show slight differences in patterns of performance across cognitive domains but the factor structure of the NACC battery remains generally stable for each group. Factors (or cognitive domains) were also stable over time *within* each group as criteria were met for all four classifications of invariance in an evaluation over a one year time interval.

These findings provide evidence of the stability of cognitive domains as measured by the NACC UDS battery. For researchers using data derived from a nationwide sample of

participants covering a range of cognitive function, it is important to know that the same domains are being measured at different levels of cognitive function. Variation in test performance then represents quantitative change rather than qualitative. This allows for comparison across groups and over time.

Others have performed similar analyses using different neuropsychological batteries on a range of diagnostic groups with varying results.^{8, 11, 46, 47} In samples of individuals with mild impairment or AD, slightly different factor structures than those seen in normal samples seem to emerge implying variation in the cognitive domains being measured. There are also differences in the number of factors extracted from various batteries. There are a couple of reasons for these differences. First, the psychometric characteristics of each test and each battery are important to consider as some may be better at cleanly measuring individual cognitive domains than others. As noted previously,⁷ normals' performance on neurocognitive tests tends to be uniform, hence the one-factor solution typically found in normal subjects. Individuals suffering from neurodegenerative disease such as AD, tend to perform heterogeneously.

Second, the number of factors extracted can vary depending on the objective of the analysis. If data reduction is the goal, factor extraction based on the use of Eigenvalues⁴² or scree plots⁴¹ may be appropriate and can yield a smaller number of factors. When more factors are extracted, the fit statistics tend to improve. It is up to the researchers to determine the appropriate number of factors that hang together and make theoretical sense. Parsimony and simplicity play important roles.

The strengths of the current study are the large sample size and the fact that the NACC battery was administered in a standardized fashion across sites. There are a number of limitations to be noted. The sample is not population-based and therefore results may not be generalizable. Centers may follow different protocols for volunteer recruitment and various centers have different areas of specialty, which may tend to attract volunteers with a particular interest or family history. Participants were classified into three groups based on the CDR. While not required, it is possible that test scores were considered in the assignment of CDR scores. The choice of the CDR as a grouping variable over actual diagnoses was an attempt to avoid any tautology. Because we did not rely on diagnoses, there is likely a large degree of heterogeneity within each subgroup of participants. We did not exclude participants with comorbid conditions, nor did we adjust for demographic characteristics as our focus was on the general factor structure of the battery. The battery itself, while broad in scope may have limitations in its ability to tap cognitive domains. Some of the test results tend to be skewed, having floor and ceiling effects. To remedy this possibility, a Blom transformation was applied to normalize the data and we used an MPLUS estimator that is robust to non-normality.

We found that the CFA factor structure fit all three CDR groups (i.e. dimensional and configural invariance held) and the ability of individual tests to adequately represent each factor varied only slightly between different groups defined by CDR scores. Our findings provide strong support for continued efforts to elucidate a biomarker of early dementia based on detailed neuropsychological assessment. Insofar as factor structures are consistent, differences in performance within each group suggest impairments in those cognitive domains rather than the inability of the battery to adequately measure performance. It is important to note the stability of the battery because impairments in specific cognitive domains have been correlated with neuropathology in prior studies.^{7, 48} In general, our findings are in contrast to others who have found singular factor structures or factor structures that vary in different patient populations.^{7, 49} Although this can be largely attributed to differences in batteries used and sample characteristics, our results suggest that

there are quantitatively, not qualitatively, different cognitive patterns in the three diagnostic categories. In conclusion, we have found that the theoretical constructs behind the development of the UDS neuropsychological battery seem to hold. Within groups, the battery demonstrates stability over a one year period. Longer term follow-up will be useful in determining whether factor stability is maintained longitudinally at an individual level.

Acknowledgments

We would like to thank Drs. Walter Kukull and Leslie Phillips for their help in acquiring the NACC data, and Dr. Weintraub for permission to reproduce table 1. We would also like to thank the study participants who contributed their time and effort for the advancement of Alzheimer's disease research.

Funding Sources: This work was supported by grants from the National Alzheimer's Coordinating Center (NACC Junior Investigator grant: 2008-JI-02) and from the National Institute on Aging (P30- AG 028377, K01 AG029336).

References

1. Thal LJ, Kantarci K, Reiman EM, et al. The role of biomarkers in clinical trials for Alzheimer disease. *Alzheimer Dis Assoc Disord.* Jan-Mar; 2006 20(1):6–15. [PubMed: 16493230]
2. Andreasen N, Blennow K. CSF biomarkers for mild cognitive impairment and early Alzheimer's disease. *Clinical Neurology and Neurosurgery.* 2005; 107(3):165–173. [PubMed: 15823670]
3. Irizarry MC. Biomarkers of Alzheimer disease in plasma. *NeuroRx.* Apr; 2004 1(2):226–234. [PubMed: 15717023]
4. Nestor PJ, Scheltens P, Hodges JR. Advances in the early detection of Alzheimer's disease. *Nat Med.* Jul; 2004 10(Suppl):S34–41. [PubMed: 15298007]
5. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* Jul; 1984 34(7):939–944. [PubMed: 6610841]
6. Welsh K, Butters N, Hughes J, et al. Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Arch Neurol.* Mar; 1991 48(3):278–281. [PubMed: 2001185]
7. Kanne SM, Balota DA, Storandt M, et al. Relating anatomy to function in Alzheimer's disease: neuropsychological profiles predict regional neuropathology 5 years later. *Neurology.* Apr; 1998 50(4):979–985. [PubMed: 9566382]
8. 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. *Psychol Aging.* Dec; 2003 18(4):836–843. [PubMed: 14692868]
9. Delis DC, Jacobson M, Bondi MW, et al. 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.* Sep; 2003 9(6):936–946. [PubMed: 14632252]
10. Woodard JL, Salthouse TA, Godsall RE, et al. Confirmatory factor analysis of the Mattis Dementia Rating Scale in Patients with Alzheimer's Disease. *Psychol Assessment.* 1996; 8(1):85–91.
11. 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.* May; 2008 22(3):400–411. [PubMed: 18444718]
12. Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. *Alzheimer Dis Assoc Disord.* Apr-Jun; 2009 23(2):91–101. [PubMed: 19474567]
13. Morris JC, Weintraub S, Chui HC, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Dis Assoc Disord.* Oct-Dec; 2006 20(4):210–216. [PubMed: 17132964]
14. Beekly DL, Ramos EM, Lee WW, et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. *Alzheimer Dis Assoc Disord.* Jul-Sep; 2007 21(3):249–258. [PubMed: 17804958]

15. Wechsler, D. Wechsler Memory Scale-Revised Manual. Psychological Corp; San Antonio: 1987.
16. Goodglass, H.; Kaplan, E.; Barresi, B. Boston Diagnostic Aphasia Examination. Third Edition. Pro-Ed; Austin, TX: 2001.
17. Kaplan, EF.; Goodglass, H.; Weintraub, S. The Boston Naming Test. Lea & Febiger; Philadelphia: 1983.
18. Wechsler, D. The Wechsler Adult Intelligence Scale Revised. Psychological Corporation; San Antonio, TX: 1987.
19. Reitan, R.; Wolfson, D. The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation. 2nd. ed. Neuropsychology Press; Tucson, AZ: 1993.
20. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. Sep; 1989 39(9):1159–1165. [PubMed: 2771064]
21. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975; 12(3):189–198. [PubMed: 1202204]
22. Hughes CP, Berg L, Danziger WL, et al. A new clinical scale for the staging of dementia. *Br J Psychiatry*. Jun.1982 140:566–572. [PubMed: 7104545]
23. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. Nov; 1993 43(11):2412–2414. [PubMed: 8232972]
24. Williams MM, Roe CM, Morris JC. Stability of the Clinical Dementia Rating, 1979–2007. *Arch Neurol*. Jun; 2009 66(6):773–777. [PubMed: 19506139]
25. Bontempo, DE.; Hofer, SM. Assessing Factorial Invariance in Cross-Sectional and Longitudinal Studies. In: Ong, AD.; Van Dulmen, MHM., editors. *Handbook of Methods in Positive Psychology*. Vol. Vol 13. Oxford: 2007.
26. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care*. Nov; 2006 44(11 Suppl 3):S78–94. [PubMed: 17060839]
27. Meredith W, Teresi JA. An essay on measurement and factorial invariance. *Med Care*. Nov; 2006 44(11 Suppl 3):S69–77. [PubMed: 17060838]
28. Teresi JA. Overview of quantitative measurement methods. Equivalence, invariance, and differential item functioning in health applications. *Med Care*. Nov; 2006 44(11 Suppl 3):S39–49. [PubMed: 17060834]
29. Blom, G. *Statistical Estimates and Transformed Beta Variables*. John Wiley & Sons, Inc.; New York: 1958.
30. SAS/STAT [computer program]. Version 9.1. Cary, NC: 2003.
31. Mplus: Statistical Analysis with Latent Variables [computer program]. Version. Muthen & Muthen; Los Angeles, CA: 1998.
32. Satorra A, Bentler P. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001; 66(4):507–514.
33. Rafferty AE. Bayesian Model Selection in Social Research. *Sociological Methodology*. 1995; 25:111–163.
34. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull*. Mar; 1990 107(2):238–246. [PubMed: 2320703]
35. Iacobucci D. Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*. 2010; 20:90–98.
36. Tucker LR, Lewis C. A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika*. 1973; 38(1):1–10.
37. Widaman KF, Thompson JS. On specifying the null model for incremental fit indices in structural equation modeling. *Psychol Methods*. Mar; 2003 8(1):16–37. [PubMed: 12741671]
38. Bryant, FB.; King, SP.; Smart, CM. Multivariate Statistical Strategies for Construct Validation. In: Ong, AD.; Van Dulmen, MHM., editors. *Handbook of Methods in Positive Psychology*. Vol. Vol 13. Oxford: 2007.

39. Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180.
40. Joreskog, KG.; Sorbom, D. LISREL 8: User's reference guide. Scientific Software International; Chicago: 1996.
41. Cattell RB. The Scree Test for the Number of Factors. *Multivariate Behavioral Research*. 1966; 1(2):245–276.
42. Kaiser HF. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*. 1960; 20:141–151.
43. Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychol Bull*. Nov; 1993 114(3):552–566. [PubMed: 8272470]
44. Heywood HB. On finite sequences of real numbers. *Proc R Soc Lond A Math Phys Char*. 1931; 134(824):486–501.
45. Bollen, KA. Structural equations with latent variables. Wiley; New York: 1989.
46. Collie A, Shafiq-Antonacci R, Maruff P, et al. Norms and the effects of demographic variables on a neuropsychological battery for use in healthy ageing Australian populations. *Aust N Z J Psychiatry*. Aug; 1999 33(4):568–575. [PubMed: 10483853]
47. Loewenstein DA, Ownby R, Schram L, et al. An evaluation of the NINCDS-ADRDA neuropsychological criteria for the assessment of Alzheimer's disease: a confirmatory factor analysis of single versus multi-factor models. *J Clin Exp Neuropsychol*. Jun; 2001 23(3):274–284. [PubMed: 11404806]
48. Price JL, McKeel DW Jr, Buckles VD, et al. Neuropathology of nondemented aging: presumptive evidence for preclinical Alzheimer disease. *Neurobiol Aging*. Jul; 2009 30(7):1026–1036. [PubMed: 19376612]
49. Jones SN, Ayers CR. Psychometric properties and factor structure of an expanded CERAD neuropsychological battery in an elderly VA sample. *Arch Clin Neuropsychol*. May; 2006 21(4): 359–365. [PubMed: 16766158]



Figure 1. Confirmatory factor analysis model derived from exploratory factor analysis results.

Table 1Uniform Data Set Neuropsychologic Test Battery^{12*}

Domain/Function	Test/Measure	Maximum Scores
Dementia Severity	Mini-Mental State Examination (MMSE)	Total MMSE (30) Total Orientation items (10)
Attention	Digit Span Forward	Longest sequence (9) Total correct Trials (14)
	Digit Span Backward	Longest sequence (8) Total correct Trials (12)
Processing Speed	Digit Symbol	Total no. items completed in 90 s (99)
	Part A, Trail Making Test	Total time (150 s)
Executive Function	Part B Making Test	Total time (300 s)
Memory	Logical Memory Story A Immediate Recall	Total items recalled (25)
	Logical Memory Story A Delayed Recall	Total items recalled (25)
Language		
Verbal Fluency	Animal List generation	Total items in 1 minute
	Vegetable List generation	Total items in 1 minute
Naming	Boston Naming Test (30 odd items)	Total correct (30)

* Used with permission from NACC (Dr. Weintraub).¹²

Table 2

Demographic characteristics of 12,020 NACC participants at initial visit

Characteristic	Normal n=4,780	MCI n=4,081	Dementia n=3,159	Total n=12,020
Baseline Age (SD)	74.7 (8.9)	75.5 (8.7)	76.9 (9.0)	75.6 (8.9)
Female Sex (%)	3,142 (65.7)	2,072 (50.8)	1,657 (52.5)	6,871 (57.2)
Race (White)(%)	3,925 (82.3)	3,450 (84.8)	2,618 (83.0)	9,993 (83.3)
Education (SD)	15.4 (2.8)	15.0 (3.2)	14.2 (3.3)	15.0 (3.1)
Baseline MMSE (SD)	28.8 (1.5)	26.4 (3.5)	19.3 (6.1)	25.5 (5.4)

Abbreviations: MCI=Mild Cognitive Impairment; SD=Standard Deviation; MMSE =Mini Mental State Examination.

Table 3

Neuropsychological Battery Test Scores, Means (SD)

Test	Normal (CDR=0.0)	MCI (CDR=0.5)	Dementia (CDR>0.5)	Total Sample
LM Immediate Recall	13.5 (4.0)	8.7 (4.7)	3.6 (3.4)	9.4(5.7)
LM Delayed Recall	12.15 (4.4)	6.35 (5.1)	1.51 (2.8)	7.50 (5.1)
Digits Forward	8.52 (2.0)	7.80 (2.1)	6.65 (2.4)	7.80 (2.3)
Forward Length	6.69 (1.1)	6.34 (1.2)	5.68 (1.4)	6.31 (1.3)
Digits Backward	6.74 (2.2)	5.68 (2.2)	4.06 (2.1)	5.70 (2.4)
Backward Length	4.87 (1.2)	4.31 (1.3)	3.33 (1.4)	4.28 (1.4)
Trail Making Part A	36.45 (17.2)	50.40 (28.4)	84.71 (43.9)	52.64 (37.8)
Trail Making Part B	97.21 (55.8)	152.05 (83.0)	227.80 (85.1)	140.94 (86.7)
Digit Symbol	45.37 (12.5)	35.15 (13.4)	21.23 (13.7)	36.28 (16.1)
Animals	19.45 (5.7)	14.96 (5.6)	8.86 (4.8)	15.21 (6.9)
Vegetables	14.35 (4.4)	10.53 (4.0)	5.87 (3.7)	10.89 (5.3)
Boston Naming	26.82 (3.6)	24.13 (5.5)	17.58 (7.8)	23.56 (6.6)

Abbreviations: LM=Logical Memory; SD=Standard Deviation; CDR=Clinical Dementia Rating

Table 4

Evaluation of Invariance Across Groups: Standardized Factor Loadings

Test	Model 1 Configural Invariance			Model 2 Metric Invariance			Model 3 Strong (Scalar) Factorial Invariance			Model 4 Strict Invariance		
	Norm N=2409	MCI N=2037	Dem N=1564	Norm N=2409	MCI N=2037	Dem N=1564	Norm N=2409	MCI N=2037	Dem N=1564	Norm N=2409	MCI N=2037	Dem N=1564
	Fix variances=1 and means=0; free intercepts factor loadings, and residual variances											
	Fix factor loadings, variances=1, and means=0; free intercepts and residual variances											
	Fix intercepts, factor loadings, and variances; free means=0 in reference group, free residual variances											
	Fix intercepts, factor loadings, variances, and residual variances; fix means=0 in reference group											
Memory												
LM Immediate Recall	0.92	0.98	1.04	0.999	0.98	0.92	0.92	0.90	0.85	0.90	0.90	0.90
LM Delayed Recall	0.94	0.84	0.63	0.85	0.81	0.83	0.93	0.88	0.89	0.89	0.89	0.89
Attention												
Digits Forward	0.65	0.59	0.66	0.65	0.64	0.61	0.64	0.63	0.60	0.62	0.62	0.62
Digits Fwd Length	0.59	0.56	0.63	0.61	0.58	0.57	0.60	0.58	0.57	0.58	0.58	0.58
Digits Backward	0.88	0.91	0.89	0.89	0.89	0.90	0.90	0.89	0.91	0.90	0.90	0.90
Digits Bkwd Length	0.83	0.86	0.85	0.85	0.84	0.85	0.85	0.85	0.86	0.86	0.86	0.86
Executive												
Trails A	0.78	0.78	0.85	0.80	0.80	0.81	0.78	0.78	0.78	0.78	0.78	0.78
Trails B	0.86	0.86	0.80	0.86	0.84	0.84	0.86	0.84	0.84	0.85	0.85	0.85
WAIS (Digit Symbol)	0.79	0.82	0.90	0.82	0.83	0.86	0.83	0.84	0.87	0.84	0.84	0.84
Language												
Animals	0.74	0.83	0.90	0.78	0.82	0.89	0.76	0.80	0.86	0.79	0.79	0.79
Vegetables	0.58	0.69	0.79	0.64	0.69	0.75	0.68	0.73	0.79	0.72	0.72	0.72
Boston Naming	0.63	0.64	0.67	0.66	0.63	0.64	0.64	0.61	0.63	0.62	0.62	0.62
Fit Statistics												
CFI	0.985			0.975			0.967			0.957		
TLI	0.978			0.969			0.963			0.958		
RMSEA	0.047			0.055			0.060			0.064		
SRMR	0.029			0.065			0.072			0.071		
χ^2 BIC, p -value, df	-460.77, -, df =138			-254.30, p <.0001, df =162			-70.33, p <.0001, df =178			112.90, p <.0001, df =202		

Test	Model 1 Configural Invariance			Model 2 Metric Invariance			Model 3 Strong (Scalar) Factorial Invariance			Model 4 Strict Invariance		
	Norm N=2409	MCI N=2037	Dem N=1564	Norm N=2409	MCI N=2037	Dem N=1564	Norm N=2409	MCI N=2037	Dem N=1564	Norm N=2409	MCI N=2037	Dem N=1564
	Fix variances=1 and means=0; free intercepts, factor loadings, and residual variances			Fix factor loadings, variances=1, and means=0; free intercepts and residual variances			Fix intercepts, factor loadings, and variances; fix means=0 in reference group; free residual variances			Fix intercepts, factor loadings, variances, and residual variances; fix means=0 in reference group		
χ^2 S-B p -value, df	$df = 138$			$p < .0001, df = 162$			$p < .0001, df = 178$			$p < .0001, df = 202$		

†Nested models tested using BIC33 and Satorra-Bentler Scaled χ^2 test.³² Model 1 was used as the starting point as the least constrained model. Each model was found to be statistically significantly different from the prior model based on both measures

CFI=Comparative fit index, better fit >0.90; range 0.0 -1.0 34

TLL= Tucker-Lewis index: better fit > 0.90; range 0.0 -1.0 36

RMSEA= Root mean square error of approximation: better fit <0.05 39

SRMR = Standardized root mean square residual: better fit <0.0539

BIC = Bayesian Information Criterion: better fit <0.0 33

S-B=Satorra-Bentler Scaled χ^2 test. 32

Table 5a

Evaluation of Invariance Over Time: Normal (CDR=0.0)

Test	Model 1 Configural Invariance		Model 2 Metric Invariance		Model 3 Strong (Scalar) Factorial Invariance		Model 4 Strict Invariance	
	Baseline N=2409	Follow-up N=2449	Baseline N=2409	Follow-up N=2449	Baseline N=2409	Follow-up N=2449	Baseline N=2409	Follow-up N=2449
	Means and variances held constant		Factor Loading/Variances/Mean held constant		Intercepts/Factor Loading/Variances held constant, means=0 in reference group		Intercepts/Factor Loading/Variances /Residual Variances held constant, means=0 in reference group	
Memory								
LM Immediate Recall	0.92	0.90	0.91	0.91	0.91	0.91	0.91	0.91
LM Delayed Recall	0.94	0.96	0.95	0.95	0.95	0.95	0.95	0.95
Attention								
Digits Forward	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
Digits Fwd Length	0.59	0.60	0.60	0.60	0.60	0.60	0.60	0.60
Digits Backward	0.88	0.85	0.87	0.86	0.87	0.86	0.86	0.86
Digits Bkwd Length	0.83	0.81	0.82	0.82	0.82	0.82	0.82	0.82
Executive								
Trails A	0.78	0.81	0.79	0.81	0.79	0.81	0.80	0.80
Trails B	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
WAIS (Digit Symbol)	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79
Language								
Animals	0.74	0.75	0.74	0.75	0.74	0.75	0.75	0.75
Vegetables	0.58	0.62	0.59	0.61	0.59	0.61	0.60	0.60
Boston Naming	0.63	0.61	0.62	0.62	0.62	0.62	0.62	0.62
Fit Statistics								
CFI	0.984		0.984		0.983		0.984	
TLI	0.976		0.979		0.980		0.983	
RMSEA	0.050		0.047		0.045		0.042	
SRMR	0.029		0.030		0.030		0.030	
χ^2 BIC, <i>p</i> -value, <i>df</i>	-136.26, -, 92		-228.65, <i>p</i> <.0001, 104		-282.61, <i>p</i> <.0001, 112		-390.92, <i>p</i> <.0001, 124	

Test	Model 1 Configural Invariance		Model 2 Metric Invariance		Model 3 Strong (Scalar) Factorial Invariance		Model 4 Strict Invariance	
	Baseline N=2409	Follow-up N=2449	Baseline N=2409	Follow-up N=2449	Baseline N=2409	Follow-up N=2449	Baseline N=2409	Follow-up N=2449
	Means and variances held constant		Factor Loading/Variances/Means held constant		Intercepts/Factor Loading/Variances held constant, means=0 in reference group		Intercepts/Factor Loading/Variances / Residual Variances held constant, means=0 in reference group	
χ^2 p-value, df	df = 92		p = 0.002, df = 104		p < 0.001, df = 112		p = 0.03, df = 124	

†Nested models tested using BIC33 and Satorra-Bentler Scaled χ^2 test.32 Model 1 was used as the starting point as the least constrained model. Each model was found to be statistically significantly different from the prior model based on both measures

CFI=Comparative fit index, better fit >0.90; range 0.0 -1.0 34

TLL= Tucker-Lewis index: better fit > 0.90; range 0.0 -1.0 36

RMSEA= Root mean square error of approximation: better fit <0.0539

SRMR = Standardized root mean square residual: better fit <0.0540

BIC = Bayesian Information Criterion: better fit<0.033

S-B=Satorra-Bentler Scaled χ^2 test.32

Table 5b

Evaluation of Invariance Over Time: MCI (CDR=0.5)

Test	Model 1 Configural Invariance		Model 2 Metric Invariance		Model 3 Strong (Scalar) Factorial Invariance		Model 4 Strict Invariance	
	Baseline N=2037	Follow-up N=1514	Baseline N=2037	Follow-up N=1514	Baseline N=2037	Follow-up N=1514	Baseline N=2037	Follow-up N=1514
	Means and variances held constant		Factor Loading/Variances/Mean held constant		Intercepts/Factor Loading/Variances held constant, means=0 in reference group		Intercepts/Factor Loading/Variances /Residual Variances held constant, means=0 in reference group	
Memory								
LM Immediate Recall	0.98	0.99	0.97	1.01	0.97	1.00	0.98	0.98
LM Delayed Recall	0.84	0.85	0.84	0.85	0.84	0.85	0.85	0.85
Attention								
Digits Forward	0.59	.61	0.60	0.60	0.60	0.60	0.60	0.60
Digits Fwd Length	0.56	0.56	0.55	0.56	0.56	0.56	0.56	0.56
Digits Backward	0.91	0.88	0.89	0.90	0.89	0.90	0.90	0.90
Digits Bkwd Length	0.86	0.82	0.84	0.85	0.84	0.85	0.85	0.85
Executive								
Trails A	0.78	0.80	0.78	0.80	0.78	0.80	0.79	0.79
Trails B	0.86	0.87	0.85	0.87	0.85	0.87	0.86	0.86
WAIS (Digit Symbol)	0.82	0.83	0.81	0.84	0.81	0.84	0.82	0.82
Language								
Animals	0.83	0.84	0.83	0.84	0.83	0.84	0.83	0.83
Vegetables	0.69	0.73	0.69	0.72	0.69	0.72	0.70	0.70
Boston Naming	0.64	0.61	0.62	0.64	0.62	0.64	0.63	0.63
Fit Statistics								
CFI	0.987		0.986		0.986		0.984	
TLI	0.981		0.983		0.984		0.983	
RMSEA	0.043		0.041		0.039		0.040	
SRMR	0.026		0.030		0.030		0.033	
†BIC, <i>p</i> -value, <i>df</i>	-360.32, -, 92		-437.35, <i>p</i> <.0001, 104		-494.06, <i>p</i> <.0001, 112		-536.81, <i>p</i> <.0001, 124	

Test	Model 1 Configural Invariance		Model 2 Metric Invariance		Model 3 Strong (Scalar) Factorial Invariance		Model 4 Strict Invariance	
	Baseline N=2037	Follow-up N=1514	Baseline N=2037	Follow-up N=1514	Baseline N=2037	Follow-up N=1514	Baseline N=2037	Follow-up N=1514
†S-B χ^2 p-value, df	Means and variances held constant		Factor Loading/Variations/Mean held constant		Intercepts/Factor Loading/Variations held constant, means=0 in reference group		Intercepts/Factor Loading/Variations / Residual Variations held constant, means=0 in reference group	
	df= 92		p <.0001, df= 104		p = 0.003, df= 112		p <.0001, df= 124	

†Nested models tested using the BIC33 and Satorra-Bentler Scaled χ^2 test.32 Model 1 was used as the starting point as the least constrained model. Each model was found to be statistically significantly different from the prior model based on both measures

CFI=Comparative fit index, better fit >0.90; range 0.0 -1.0 34

TLI= Tucker-Lewis index: better fit > 0.90; range 0.0 -1.0 36

RMSEA= Root mean square error of approximation: better fit <0.0539

SRMR= Standardized root mean square residual: better fit <0.0540

BIC= Bayesian Information Criterion: better fit<0.033

S-B=Satorra-Bentler Scaled χ^2 test.32

Table 5c

Evaluation of Invariance Over Time: Dementia (CDR>0.5)

Test	Model 1 Configural Invariance		Model 2 Metric Invariance		Model 3 Strong (Scalar) Factorial Invariance		Model 4 Strict Invariance	
	Baseline N=1564	Follow-up N=1428	Baseline N=1564	Follow-up N=1428	Baseline N=1564	Follow-up N=1428	Baseline N=1564	Follow-up N=1428
	Means and variances held constant		Factor Loading/Variances/Mean held constant		Intercepts/Factor Loading/Variances held constant, means=0 in reference group		Intercepts/Factor Loading/Variances /Residual Variances held constant, means=0 in reference group	
Memory								
LM Immediate Recall	1.04	1.06	1.03	1.07	1.03	1.06	1.04	1.04
LM Delayed Recall	0.63	0.59	0.61	0.61	0.62	0.61	0.61	0.61
Attention								
Digits Forward	0.66	0.66	0.66	0.67	0.66	0.67	0.66	0.66
Digits Fwd Length	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63
Digits Backward	0.89	0.92	0.90	0.91	0.90	0.91	0.91	0.91
Digits Bkwd Length	0.85	0.88	0.86	0.88	0.86	0.88	0.87	0.87
Executive								
Trails A	0.85	0.85	0.85	0.86	0.85	0.86	0.85	0.85
Trails B	0.80	0.77	0.78	0.79	0.78	0.79	0.79	0.79
WAIS (Digit Symbol)	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
Language								
Animals	0.90	0.90	0.89	0.90	0.89	0.91	0.90	0.90
Vegetables	0.79	0.83	0.80	0.83	0.80	0.83	0.81	0.81
Boston Naming	0.67	0.74	0.69	0.72	0.69	0.72	0.70	0.70
Fit Statistics								
CFI	0.988		0.988		0.988		0.988	
TLI	0.983		0.985		0.986		0.987	
RMSEA	0.044		0.041		0.040		0.038	
SRMR	0.030		0.033		0.034		0.036	
χ^2 /BIC, <i>p</i> -value, <i>df</i>	-381.56, -, 92		-464.77, <i>p</i> <.0001, 104		-522.01, <i>p</i> <.0001, 112		-601.56, <i>p</i> <.0001, 124	

Test	Model 1 Configural Invariance		Model 2 Metric Invariance		Model 3 Strong (Scalar) Factorial Invariance		Model 4 Strict Invariance	
	Baseline N=1564	Follow-up N=1428	Baseline N=1564	Follow-up N=1428	Baseline N=1564	Follow-up N=1428	Baseline N=1564	Follow-up N=1428
	Means and variances held constant		Factor Loading/Variances/Mean held constant		Intercepts/Factor Loading/Variances held constant, means=0 in reference group		Intercepts/Factor Loading/Variances / Residual Variances held constant, means=0 in reference group	
χ^2 -B-S	df = 92		p = 0.001, df = 104		p < 0.01, df = 112		p = 0.002, df = 124	

[†] Nested models tested using BIC33 and Satorra-Bentler Scaled χ^2 test.32 Model 1 was used as the starting point as the least constrained model. Each model was found to be statistically significantly different from the prior model based on both measures

CFI=Comparative fit index, better fit >0.90; range 0.0 -1.0 34

TLI= Tucker-Lewis index: better fit > 0.90; range 0.0 -1.0 36

RMSEA= Root mean square error of approximation: better fit <0.0539

SRMR= Standardized root mean square residual: better fit <0.0540

BIC= Bayesian Information Criterion: better fit <0.033

S-B=Satorra-Bentler Scaled χ^2 test.32