





Research Article

Factor structure of the Harmonized Cognitive Assessment Protocol neuropsychological battery in the Health and Retirement Study

Richard N. Jones^{1,2} , Jennifer J. Manly³ , Kenneth M. Langa^{4,5,6,8}, Lindsay H. Ryan⁴, Deborah A. Levine^{5,6,7}, Ryan McCammon⁴ and David Weir⁴

¹Department of Psychiatry and Human Behavior, Department of Neurology, Brown University Warren Alpert Medical School, Providence, Rhode Island, United States, ²Department of Neurology, Brown University Warren Alpert Medical School, Providence, Rhode Island, United States, ³Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, United States, ⁴Institute for Social Research, University of Michigan, Ann Arbor, Michigan, United States, ⁵Cognitive Health Sciences Research Program and Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, United States, ⁶Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, Michigan, United States, ⁷Department of Neurology and Stroke Program, University of Michigan, Ann Arbor, Michigan, United States and ⁸Veterans Affairs Center for Clinical Management Research, Ann Arbor, Michigan, United States

Abstract

Objective: The Harmonized Cognitive Assessment Protocol (HCAP) describes an assessment battery and a family of population-representative studies measuring neuropsychological performance. We describe the factorial structure of the HCAP battery in the US Health and Retirement Study (HRS). **Method:** The HCAP battery was compiled from existing measures by a cross-disciplinary and international panel of researchers. The HCAP battery was used in the 2016 wave of the HRS. We used factor analysis methods to assess and refine a theoretically driven single and multiple domain factor structure for tests included in the HCAP battery among 3,347 participants with evaluable performance data. **Results:** For the eight domains of cognitive functioning identified (orientation, memory [immediate, delayed, and recognition], set shifting, attention/speed, language/fluency, and visuospatial), all single factor models fit reasonably well, although four of these domains had either 2 or 3 indicators where fit must be perfect and is not informative. Multidimensional models suggested the eight-domain model was overly complex. A five-domain model (orientation, memory delayed and recognition, executive functioning, language/fluency, visuospatial) was identified as a reasonable model for summarizing performance in this sample (standardized root mean square residual = 0.05, root mean square error of approximation = 0.05, confirmatory fit index = 0.94). **Conclusions:** The HCAP battery conforms adequately to a multidimensional structure of neuropsychological performance. The derived measurement models can be used to operationalize notions of neurocognitive impairment, and as a starting point for prioritizing pre-statistical harmonization and evaluating configural invariance in cross-national research.

Key words: factor analysis; statistical; dementia; aged; 80 and over; surveys and questionnaires

(Received 22 February 2022; final revision 24 January 2023; accepted 31 January 2023; First Published online 14 July 2023)

Introduction

In 2015, Alzheimer disease and dementia were health outcomes of topmost concern for older adults living in the US and Europe (Cutler, 2015). The relatively high prevalence of Alzheimer's disease and related dementias (ADRD), together with shifting demographics and an absence of disease-modifying treatments conspire to make ADRD a costly public health concern (Alzheimer's Association, 2019; Hurd et al., 2013). Research to support public health planning and knowledge accumulation regarding the economic, epidemiologic, sociologic, and psychological impact of ADRD requires high quality information on cognitive functioning of older and aging adults.

The Health and Retirement Study (HRS) has been collecting information on cognition using brief measures from more than

43,000 community dwelling older adults since 1992 (Sonnega et al., 2014). HRS investigators previously conducted a focused dementia substudy known as the Aging, Demographics and Memory Study (Langa et al., 2005; Plassman et al., 2008). A new dementia-focused substudy launched in 2016 (Langa et al., 2020). The new cognitive assessment, named the Harmonized Cognitive Assessment Protocol (HCAP), was designed to create a new data resource to better assess the prevalence, determinants, and costs and consequences of mild cognitive impairment (MCI) and dementia in the US and to facilitate cross-national multidisciplinary studies of cognitive aging. The instrumentation was collaboratively assembled with leaders of international partner studies (IPS) implementing similar studies around the world and field leaders in cross-national cognitive epidemiology (Langa et al., 2020; Meng et al., 2019).

Corresponding author: Richard N. Jones, Department of Psychiatry and Human Behavior, Department of Neurology, Brown University Warren Alpert Medical School, Providence. E-mail: Richard_Jones@Brown.edu

Cite this article: Jones R.N., Manly J.J., Langa K.M., Ryan L.H., Levine D.A., McCammon R., & Weir D. (2024) Factor structure of the Harmonized Cognitive Assessment Protocol neuropsychological battery in the Health and Retirement Study. *Journal of the International Neuropsychological Society*, 30: 47–55, <https://doi.org/10.1017/S135561772300019X>

Copyright © INS. Published by Cambridge University Press, 2023. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Currently, the HCAP has been fielded by 9 IPS, and 3 IPS are planning to put the HCAP in the field in the near future.

There are three anticipated uses of the HCAP in the HRS. First, the HCAP can be used to generate summaries of multiple domains of cognitive performance among older adults in a nationally representative population sample. Second, the HCAP can be used to evaluate measurement invariance in support of cross-national studies of adult cognition. Third, the HCAP can be used together with other relevant information to approximate clinically relevant states such as MCI and probable dementia. Each of these intended uses places different pressures on the decision-making process regarding the measurement of cognitive performance. We use latent variable models to describe how several cognitive performance tests measure a smaller number of cognitive domains. The results provide a framework for the three intended uses of the HCAP (summarizing multiple domains of cognitive performance, supporting cross-national studies of adult cognition, approximating clinically relevant states such as MCI and probable dementia). Latent variable modeling, and specifically factor analysis, invokes the following assumptions: (1) persons possess certain related but to some degree distinct cognitive capacities that are not directly measurable; (2) observed cognitive performance reflects the causal effect of these underlying cognitive capacities; (3) the selection of cognitive tasks in the battery and in our analysis is sufficient to reveal the hypothesized underlying domains; and (4) our modeling decisions regarding co-dependence of test performance data on underlying domains is a reasonable reflection of the true causal dependence. Even if these assumptions are false, our modeling results may be useful. Simplifying cognitive performance data on many tests that similarly reflect underlying domains provides a framework for the efficient description of cognitive performance. Cross-national studies must be concerned with measurement equivalence, and latent variable models provide a natural framework for exploring and characterizing differences due to language and culture. Finally, approximating cognitive states, for example those described by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013) require the demonstration of deficits among core neuropsychological processes, including learning and memory, executive function, language, perceptual-motor, and complex attention. Basing judgments on the presence or absence of deficits at the latent variable level, rather than the individual test level, is an approach that can be generalized to different studies with different batteries and generate more comparable results.

Methods

Participants

The HRS began in 1992 and is ongoing in 2022. The HRS was initially conceived as a study of health and wealth as adults approached retirement age. A stratified national probability sample of 12,652 adults aged 51–61 (birth years 1931–1941) was recruited in 1992. The content areas of the interview included health, cognition, income, wealth, family relationships, and factors relevant to retirement decisions. The sampling frame has been expanded by adding additional birth years every 6 years. Currently, the HRS represents cohorts born 1965 and before, and includes nursing home residents and has oversampled African-Americans and Hispanics. In 2006, the sample was randomly and evenly split to a telephone mode and an in-person, face-to-face mode. Thereafter, participants alternate mode at each subsequent biennial interview. The in-person face-to-face mode

allows for expanded data collection. Following the 2016 wave a separate in-person interview was conducted with a more in-depth neuropsychological performance evaluation component: the HCAP.

As described more completely in Langa *et al.* (2020), the HRS sample provided the sampling frame for the HCAP field study. Half of the HRS sample was randomly selected for participation in the HCAP, subject to some restrictions. Inclusion/exclusion criteria were only based on age and completion of the 2016 HRS core interview. Participation was restricted to those who completed the 2016 core HRS interview, and who did so when at least age 65 (birth year 1952 or earlier) ($n = 5,001$). Some of those identified for recruitment were never invited to HCAP ($n = 576$) as accrual goals were met prior to their invitation date. About 21% ($n = 929/(5001-576)$) of eligible and invited participants did not participate in the HCAP. For about 4% of those participating in HCAP (149/3496) only informant interviews were available. For these 149 participants, there were no non-missing observations on the cognitive indicators used in the present analysis, and these participants were not included. The analytic sample size for this study is therefore 3,347 persons (see Figure S3). Factors associated with non-participation are discussed in Langa *et al.*'s review of the HCAP (2020). Briefly, non-participants and participants were similar in terms of age, sex, race/ethnicity, educational attainment, and equally distributed across US Census regions. However, non-participants were more likely to be living in a nursing home (6% vs 3% among participants) and more likely to have participated in the HRS 2016 core via proxy (14% vs 6%). Participants provided written informed consent using procedures approved by the University of Michigan Institutional Review Board. The HRS and HCAP studies are conducted in accordance with the Helsinki Declaration.

Measures

The HCAP project and neuropsychological assessment battery have been described previously (Langa *et al.*, 2020; Weir *et al.*, 2014). The HCAP assessment battery includes items from several different tests. In Table 1, we detail the tests included in HCAP and the domain to which they were assigned. The HCAP includes items from brief mental status tests including the Mini-Mental State Examination (MMSE) (Folstein *et al.* 1975), the Telephone Interview for Cognitive Status (TICS) (Brandt *et al.* 1988), and the Community Screening Instrument for Dementia (CSI-D) (Hall *et al.*, 2000; Hall *et al.*, 1993). Elements from neuropsychological tests and batteries are also included. Additionally:

- from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery (Morris *et al.*, 1989), a 10 item word list learning, delayed recall, and recognition of a word list, and constructional praxis, the immediate and delayed copying of figures (four geometric forms: circle, overlapping rectangles, diamond and cube);
- Timed backwards counting from the Brief Test of Adult Cognition by Telephone (BTACT), and used in the MIDUS (Midlife in the US) study (Lachman *et al.*, 2013), which assesses how low a participant can count starting from 100 in 30 seconds;
- Animal naming, adapted from the Woodcock Johnson-III (WJ-III) Tests of Achievement (Woodcock *et al.*, 2001), which asks participants to name as many animals as possible in 1 minute (Schrank & Flanagan, 2003; Weir *et al.*, 2014);
- Symbol Digit Modalities Test (SDMT) (Smith, 1982), that asks the respondent to substitute a number for geometric figures using a key under a 90 second time limit;

Table 1. Health and Retirement Study (HRS) Harmonized Cognitive Assessment Protocol (HCAP) cognitive performance tasks, initial domain assignment, observed score ranges, and analytic variable definitions

Cognitive performance item(s) or task(s) and initial domain assignment	Source	Raw score range	Analytic variable range [†]
Orientation			
Orientation to time and place (10 items)	MMSE	0-10	0-10
Name President	TICS	0/1	0/1
Memory: immediate			
10 word list immediate recall	CERAD	0-30	0-30
3 word immediate recall	MMSE	0-3	0-3 [†]
Logical Memory I	WMS	0-25	0-25
Story recall, immediate (Brave man)	EBMT	0-12	0-12
Memory: delayed			
10 word list delayed recall	CERAD	0-10	0-10
Logical Memory II	WMS	0-25	0-25
3 word delayed recall	MMSE	0-3	0-3
Constructional praxis, delayed	CERAD	0-11	0-11
Story recall, delayed (Brave man)	EBMT	0-12	0-12
Memory: recognition			
10 Word list recognition	CERAD	0-20	0-20
Logical Memory recognition	WMS	0-15	0-15
Set shifting			
Standard Progressive Matrices	Raven	0-17	0-17
Number series	HRS	409-584	409-504
Trail making part B	Army Alpha	T = [0-300]	1-log(T)/log(300)
Attention, speed of processing			
Substitution number correct, minus mistakes and skips	SDMT	0-71	0-71
Trail making part A	Army Alpha	T = [0-300]	1-log(T)/log(300)
Backwards spelling	MMSE	0-5	0-5 [†]
Backwards counting	BTACT	0-80	0-80
Letter cancellation, number correct	NHSD	0-37	0-37
Language, fluency			
Animal naming	WJ-III	0-43	0-43
Naming two objects (cactus, scissors)	TICS	0-2	0-2 [†]
Naming two objects (pencil, watch)	MMSE	0-2	0-1/2 [‡]
Write a sentence	MMSE	0/1	0/1
Read and follow command	MMSE	0/1	0/1
Object naming	CSI-D	0-4	0-2/3/4 [‡]
Visuospatial			
Constructional praxis, immediate	CERAD	0-11	0-11
Copy polygons	MMSE	0/1	0/1

Note. The order of tests and indicators in Table 1 is according to initial domain assignment, rather than order of administration. Abbreviations: BTACT, Brief Test of Adult Cognition by Telephone; CERAD, Consortium to Establish a Registry for Alzheimer's Disease; CSI-D, Community Screening Instrument for Dementia; EBMT, East Boston Memory Test; MMSE, Mini-Mental State Examination; NHSD, National Survey of Health and Development; SDMT, Symbol Digit Modalities Test; T, time in seconds; TICS, Telephone Interview for Cognitive Status; WJ-III, Woodcock Johnson III; WMS, Wechsler Memory Scale.

[†]Variables with fewer than 10 response levels treated as categorical in analytic models, all other indicators treated as continuous and normalized to a 0-1 scale given sample observed minimum and maximum prior to factor analysis; [‡]categories with sparsely populated cells are collapsed as indicated.

- Trail Making parts A & B, which was included in the Army Individual Test (Armitage, 1946), involves tracing lines among sequentially numbered nodes (part A) or alternating numbered and letter-labeled nodes (part B) within a 300 second time limit for each part;
- Raven's Standard Progressive Matrices (Raven, 1981) as adapted for studies of cognitive aging based at Rush University (Wilson et al., 2002), wherein respondents are shown a series of 17 illustrations of geometric shapes and patterns and are asked to identify the missing picture from among six to eight choices;
- The HRS Number Series task (Fisher et al., 2014). The Number Series task used in the HRS was adapted from measures used in the WJ-III Cognitive Abilities Assessment (Fisher et al., 2014).

The test measures the ability to reason with concepts that depend upon numerical relationships (quantitative reasoning). Respondents are presented with a series of numbers and a blank indicating a number missing from the series (e.g., "6 7 [] 9"). The respondent must determine the numerical pattern and then provide the missing number (e.g., "8") in the series. In HRS, the number series task is administered in an adaptive fashion, as described in Fisher et al (2014). The resulting score is based on a Rasch (1960) measurement model using the items administered and item parameters derived from a separate calibrations sample. The score is standardized to a typical metric used in the WJ-III test battery, called a W score. On this score, a value of 500 is the expected average score for the calibration sample, and a 10 point difference on the W-score metric represents a halving or doubling of the probability of a correct response to a given item (Fisher et al., 2014);

- A letter cancellation task developed by the National (UK) Survey of Health and Development (NSHD) used in the British 1946 Birth Cohort Study (Richards et al., 1999) and the English Longitudinal Study of Aging (Stephens et al., 2013);
- Immediate and delayed verbal memory was assessed with two story recall tasks. The East Boston Memory Test (EBMT) (Scherr et al., 1988), also referred to as the Brave Man story, is a short story (three sentences) that is read to participants. Immediately after, and again after a delay, respondents are asked to recall twelve details from the story;
- The HCAP also included Logical Memory I & II from the Wechsler Memory Scale (Wechsler, 1987). Note that sections from the WMS-R are not the newer versions to maintain comparability with other field studies of cognitive aging. Like the EBMT, participants are read a story, but respondents are asked to recall the story verbatim (LM-I) and after a 20 to 30-minute delay (LM-II). Respondents are also presented with a recognition task regarding the story.

The HCAP was offered in both English (95% of completed HCAP) and Spanish (5% of completed HCAP), but no other languages. Some instruments or tasks from instruments had Spanish language versions available from the test publisher (MMSE, WMS, CERAD, SDMT, Raven's Standard Progressive Matrices, Trail Making) while others were translated. Translations were completed collaboratively by HRS/HCAP investigators and investigators affiliated with the Mexican Cognitive Aging Ancillary Study (Mex-Cog) study, a sub-study of the Mexican Health and Aging Study, and discussed by Mejia-Arango et al., (2020), especially the EBMT. Other translations were performed by HRS investigators, completed for prior waves of the HRS or for the current HCAP field study, and included tasks from the TICS, Woodcock-Johnson III tasks, HRS Number Series, letter cancellation, backwards counting task from the BTACT, and items from the CSI-D. Interviewers were trained to skip a test or to be flexible in administration to accommodate sensory impairments or physical limitations.

For our modeling we selected 29 indicators (items) capturing responses to tasks contained in 12 distinct instruments in the HCAP. Specific details are presented in Table 1, organized according to eight domains. Initial assignments of items to domains proposed by an expert committee of HRS and ISP investigators (Weir et al., 2014) and revised as a starting domain structure for the present analyses. In addition to the analytic variables summarized in Table 1, we also considered error scores on letter cancellation, symbol digit, and animal fluency. However, after initial descriptive

analyses revealed these indicators were extremely skewed and demonstrated very low absolute correlation with other indicators (maximum of $r = .26$), these error scores were removed from the analytic model.

Analytic approach

The goal of the analyses was to develop and evaluate latent variable models that inform the measurement of broad domains of cognitive performance measured with the HCAP battery. Our process included (1) descriptive analyses to inform analytic sample inclusion and variable definitions (Table 1), (2) single factor models, approached as confirmatory factor analysis (CFA) within *a priori* defined domains, (3) correlated factors models, approached as a CFA attempting to fit a multidimensional factor model, (4) second-order factor models, approached as a CFA with multiple specific and one general latent variable, and (5) estimation of factor score estimates for HRS/HCAP participants. The general sequence of model steps is illustrated in Supplementary Figure S1.

The purpose of the single factor models is to assess whether domains specified *a priori* form reasonable latent variable models as hypothesized. The results of these models are not expected to be unambiguously successful, as some factors are measured with only 2 or 3 indicators, and are just identified (have a unique solution) or identified only with the imposition of modeling constraints (Brown, 2006). The correlated factors models address whether the *a priori* specific domains are measured distinctly by the tests included in HCAP. Because our goal was to define a relatively smaller number of *a priori* defined cognitive domains – relative to the number of included neuropsychological tests – it would have been sufficient for our purposes to stop our modeling with the description of adequately fitting single factor models. Using the HCAP battery to define a small number of domains to operationalize neurocognitive impairment criteria was one of the main goals of this analysis. The second-order factor model, a model nested within the correlated factors model, specifies a single common general cognitive performance trait underlies the specific domains. This model, if adequately fitting in the HCAP data, could be useful for descriptive analyses and for cross-national linking and comparison (Nichols *et al.*, 2022).

Latent variable measurement models, or CFA models, were estimated using Mplus software (version 8.2, Muthén & Muthén, Los Angeles CA). We used three different estimators: least squares, maximum likelihood, and Bayesian. We used the weighted least squares estimator with mean and variance standardization (WLSMV) and theta parameterization to evaluate proposed and alternative factor models. Because our indicators included categorical variables, this is the only of the three estimators for which conventional model fit statistics are available. Because the least squares estimator relies upon a pairwise complete missing data handling strategy that invokes unreasonable assumptions about missing data, we use robust maximum likelihood (MLR) estimators to obtain final parameter estimates and expected *a posteriori* (EAP) factor score estimates, which involves the less restrictive conditionally random missingness assumption. We use Bayesian estimation to obtain factor score estimates as plausible values (PVs), which are draws from the model-implied posterior distribution of factor scores for each individual. Factor indicators were treated as continuous or categorical as indicated in Table 1. We only moved to multidimensional models after a satisfactorily fitting unidimensional model within a pre-specified domain was

achieved. Specific methods factors, uncorrelated with domain or general factors, were specified *a priori* to capture item-level residual covariance when indicators derived from the same stimulus (e.g., immediate and delayed recall of a common word list, short story, or design set) (Reise, 2012). The correlated factors models were specified following a simple structure implied by the single factor models. Factor models were identified by freely estimating the factor variance and all but one of the factor loadings. Model fit was evaluated with the standardized root mean square residual (SRMR), confirmatory fit index (CFI), and root mean squared error of approximation. Greatest weight was afforded the SRMR and an examination of model residuals. The SRMR is an absolute measure of model fit and suggests the value of the average residual for the model-implied correlation matrix given the observed correlation matrix. We considered a model to fit satisfactorily if the SRMR was less than or equal to 0.08 (Asparouhov & Muthén, 2018; Hu & Bentler, 1999). We also considered two model fit statistics derived from the model χ^2 : the CFI (a relative or incremental fit index, meaning the target model is compared to a null or baseline model) and the RMSEA (an absolute fit index, meaning no comparison to a null or baseline model). Both the CFI and RMSEA reward more parsimonious models. We used CFI values of at least 0.95 and RMSEA values of less than 0.05 (Hu & Bentler, 1999) to indicate satisfactory model fit. Because the CFI and RMSEA are known to be sensitive to the number of parameters estimated (Kenny *et al.*, 2015), and our model designs are largely determined *a priori*, we gave the greatest emphasis to the SRMR in adjudging model fit. When misfit was identified, we explored the inclusion of omitted paths and dropped indicators with weak loadings or other identified problems (e.g., high residual correlation with other indicators, suggesting lack of conditional independence) in attempts to resolve identified issues.

Related to model fit, we report an internal consistency reliability coefficient for our latent variable factor models: MacDonald's omega (ω_t). Omega is interpreted as is Cronbach's internal consistency coefficient alpha. We computed omega as the sum of factor loadings, squared, divided by the sum of factor loadings, squared, plus the sum of the indicator residual variances (Peters, 2014). Reliability of 0.70 is typical for early measurement development work, but should be at least 0.80 for group differences research, and 0.90–to 0.95 is the preferred standard for individual-level inference (Nunnally & Bernstein, 1994).

We estimated factor scores using Mplus and the MLR estimator. The MLR estimator has more advantageous missing data handling than the least squares estimator and will generate estimates of the latent trait for respondents with at least 1 non-missing indicator. MLR factor score estimates are EAP factor score estimates. EAP factor score estimates are similar to regression method factor scores when all indicators are continuous. Each response pattern (i.e., the vector of item responses for an individual participant) has a distinct EAP factor score associated with it. The EAP estimate is the average of the posterior distribution of ability given the observed response pattern and estimated measurement model parameters. We compare these EAP factor score estimates to factor scores estimated as Bayesian PV. Bayesian PVs are draws from the posterior distribution of the latent trait determined for a participant given their response pattern and the estimated measurement model. PV are centered on the EAP estimate but have a level of variability that reflects the level of precision of the model. The comparison of EAP and PV can give insight into the level of precision of the measurement across the range of the latent trait.

Table 2. Characteristics of sample who participated in HCAP interview and who were evaluated in this study

Characteristic	Sampled	Included	Not Included
Total (N)	3,496	3,347	149
Age (years, M (SD))	76.9 (7.6)	76.7 (7.5)	81.6 (8.9)
Sex (N, %)			
Male	1,401 (40)	1,327 (40)	74 (50)
Female	2,095 (60)	2,020 (60)	75 (50)
Race/ethnicity (N, %)			
White, not Latinx or Hispanic	2,483 (71)	2,383 (71)	100 (67)
Black, not Latinx or Hispanic	551 (16)	527 (16)	24 (16)
Latinx or Hispanic, any race	383 (11)	363 (11)	20 (13)
All other race/ethnicity groups	79 (2)	74 (2)	5 (3)
Mini-Mental State Examination (MMSE) score (M (SD) (N=3,347))	26.6 (3.9)	26.6 (3.9)	NA

Table 3. Summary of single factor domain-specific models

Model	Domain	p	ω	χ^2	df	SRMR	RMSEA	CFI	NRV	Notes
I	Orientation	2	NA ^a	0	0	0.000	0.000	1.000	+	Redundant indicators; retain MMSE orientation as single indicator
IIA	Memory, immediate	4	0.76	36	2	0.021	0.071	0.981		Poor RMSEA, retain as final model
IIB	Memory, delayed	5	0.85	44	5	0.018	0.048	0.985		Retain as final model
IIC	Memory, recognition	2	0.65	0	0	0.000	0.000	1.000		No identified issues other than only 2 indicators
III	Set shifting	3	0.81	0	0	0.000	0.000	1.000		No identified issues other than only 3 indicators
IV	Attention, speed	5	0.87	126	5	0.020	0.085	0.974		Poor RMSEA, retain as final model
V	Language, fluency	6	0.82	44	9	0.051	0.034	0.978		Retain as final model
VI	Visuospatial	2	NA ^a	0	0	0.000	0.000	1.000	+	Redundant indicators; retain CERAD constructional praxis as single indicator
IID	Memory, delayed & recognition	7	0.89	147	12	0.023	0.058	0.979		Adopted after fitting multidimensional models
IIIA	Executive functioning (Set shifting & Attention, speed)	8	0.94	766	20	0.038	0.106	0.916		Considered after fitting multidimensional models

Note. Abbreviations: p, number of indicators; ω , MacDonald's omega; χ^2 model chi-square, df, degrees of freedom; SRMR, standardized root mean square residual (≤ 0.08 satisfactory for); RMSEA, root mean square error of approximation (< 0.05 satisfactory fit); CFI, confirmatory fit index (> 0.95 satisfactory fit); NRV, negative residual variance; MMSE, Mini-Mental State Examination; CERAD, Consortium to Establish a Registry for Alzheimer's Disease.

^aOmega estimate greater than 1.0, an out-of-bounds estimate due to negative residual variance.

Data and code availability

The source variables used in this analysis are publicly available at the HRS website (<http://hrsonline.isr.umich.edu/index.php?p=shoavail&iyear=ZU>). Full details on distributions, data handling, and analytic models are available on GitHub (<https://github.com/rnj0nes/HCAP>).

Results

Characteristics of the HCAP sample used in this analysis

Characteristics of the HRS respondents randomly selected for the HCAP sample and included in this analysis (not missing on at least one of the cognitive performance tasks) are shown in Table 2. The range of MMSE scores was 0–30 with a mean of about 27 and standard deviation of 3.9; about 11% ($n = 384/3347$) had a total score less than 23.

Single factor models

Single factor models are summarized in Table 3. Three hypothesized domains had only two indicators (*orientation*, *memory: recognition*, and *visuospatial*; models I, IIC, and VI, Table 3). These models were identified by forcing the two factor loadings to be equal to 1, and a variance parameter for the latent variable was estimated (as were residual variances for continuous factor indicators and indicator means/thresholds). As specified, these models have 0 degrees of freedom and will generate fit statistics that indicate perfect fit. However, this perfect fit is misleading, as we have merely

estimated a covariance parameter rather than measured a latent dimension. Similarly, factor models with 3 indicators (*set shifting*, model III, Table 3) were identified estimating 2 factor loadings and one latent variable variance (and residual variances for continuous indicators and indicator means/thresholds) and this specification produces a just-identified model with 0 degrees of freedom.

The *orientation* model (model I, Table 3) had only two indicators, and was specified as described above to produce a just-identified model. This model fit well, as it must, but the results revealed a problem. As can be seen in Table 3, one of the items was estimated with a negative residual variance, a standardized measurement slope (i.e., factor loading) of greater than 1, and the computed ω coefficient was greater than 1. These inadmissible values suggest that the linearity assumptions imposed by the model are unreasonable. We determined that the most plausible explanation for this was that beyond the MMSE *orientation to time and place* questions, the ability to *name the president* offered no new information relevant to the underlying trait. Based on these results we opted to drop *name president* item.

The *memory: immediate episodic* single factor model (model IIA, Table 3) included 4 indicators and had 2 degrees of freedom. The model had reasonable reliability (judged by ω) for group level inference (Nunnally & Bernstein, 1994) and fit well by SRMR and CFI, but not RMSEA. The RMSEA is known to be a poor indicator of model fit in low degree of freedom models. We retained this model as a satisfactorily fitting model.

The *memory: delayed episodic* single factor model (model IIB, Table 3) included 5 indicators and had 5 degrees of freedom.

Table 4. Summary of multidimensional models

Model	Description	p	m	χ^2	df	SRMR	RMSEA	CFI	NPD	Notes
VII	Correlated factors	27	6	6,312	299	0.055	0.078	0.858	+	Poor fit
VIII	Correlated factors, with methods factors	27	6	3,138	294	0.043	0.054	0.933		Acceptable fit by SRMR only
IX	Model VIII, drop immediate memory, delayed and recognition as 1 factor	23	4	1,672	215	0.047	0.045	0.948		Acceptable fit. Not nested within above models
X	IX, combine Set shifting and Attention, speed	23	3	1,777	220	0.047	0.046	0.945		Reasonably acceptable fit. CFI is lower than ideal.
XI	Second-order factor model implied from X	23	4	2,053	225	0.049	0.049	0.935		Degraded fit relative to model X

Note. Abbreviations: p, number of indicators; m, number of common latent variables (excluding methods factors and single indicator factors); χ^2 model chi-square, df, degrees of freedom; SRMR, standardized root mean square residual (≤ 0.08 satisfactory fit); RMSEA, root mean square error of approximation (< 0.05 satisfactory fit); CFI, confirmatory fit index (> 0.95 satisfactory fit); NPD, non-positive definite covariance matrix for latent variables.

The model had relatively high reliability (judged by ω), close to the level suitable for individual-level inference (Nunnally & Bernstein, 1994) and fit well by all fit criteria (SRMR, CFI, and RMSEA). We retained this model as a satisfactorily fitting model.

The *memory: recognition* single factor model (model IIC, Table 3) had only 2 indicators, and must fit perfectly. We identified no evidence of inadmissible parameter estimates. The factor reliability (ω) was relatively poor, which cannot be surprising for a two-indicator model. We decided to retain this model as a satisfactorily fitting model as there was no way to improve it.

The *set shifting* factor model (model III, Table 3) had 3 indicators, and as we specified the model it was just identified with zero degrees of freedom and consequently fit perfectly. No modeling pathologies were identified. The model had reasonable reliability (judged by ω) for group level inference (Nunnally & Bernstein, 1994). We retained this model as a satisfactorily fitting model.

The *attention, speed* factor model (model IV, Table 3) included 5 indicators and had 5 degrees of freedom. The model had relatively high reliability (judged by ω), close to a level suitable for individual-level inference (Nunnally & Bernstein, 1994) and fit well by SRMR and CFI, but not by RMSEA. We retained this model as a satisfactorily fitting model.

The *language, fluency* single factor model (model V, Table 3) included 6 indicators and had 9 degrees of freedom. The model had reasonable reliability (judged by ω) for group level inference (Nunnally & Bernstein, 1994) and good fit was suggested by all fit criteria (SRMR, CFI, and RMSEA). We retained this model as a satisfactorily fitting model.

The *visuospatial* model (model VI, Table 3) had only two indicators and was specified as described above to produce a just-identified model. This model must fit well by standard fit criteria, but as with the *orientation* model we identified an item with a negative residual variance, a standardized measurement slope (i.e., factor loading) of greater than 1, and the computed ω coefficient was greater than 1. These inadmissible values suggest that the linearity assumptions imposed by the model are unreasonable. The most plausible explanation for this was that beyond the CERAD *constructional praxis* questions, which ask the respondent to copy 4 geometric figures, asking the respondent to copy 1 additional diagram from the MMSE provides no unique information. Based on these results we opted to drop the MMSE *copy polygons* task from the HRS/HCAP indicator set.

Multidimensional models

Multidimensional models are summarized in Table 4. The first set of models (VII and VIII) specified a correlated factors model (model VIII is depicted in Figure 1, panel A). These models included 27 indicators, omitting the TICS *name president* and

MMSE *copy polygons* as described above. Consequently, *orientation* and *visuospatial* factors were included with single indicators. The remaining factors retained the loading pattern implied in Table 1 and evaluated in the single factor models (Table 3). A comparison of model VII and VIII indicates the importance of including methods factors. The methods factors load only in the items reflecting shared content and are forced to be uncorrelated with each other and with the common factors, impart a significant improvement in model fit (change in model χ^2 (df) = 3,174 (5)).

While model fit for model VIII is satisfactory according to the SRMR, the CFI and RMSEA suggest the model could be improved. CFI and RMSEA reward parsimony, and our pattern of model fit results could signal the model is overly complex. Evidence for excess model complexity is revealed by examining the model-implied correlation matrix for the latent variables (Table S1). The factor correlations range from 0.33 (*visuospatial* with *orientation*) to 0.95 (*memory immediate episodic* with *memory delayed episodic*). The very high correlation of immediate and delayed episodic memory, even after accounting for the shared methods, implies that the two factors share 90% of their variance. Of 28 factor correlations among the 8 latent factors, 8 were estimated at .9 or higher. Included among this set were the correlations among the three memory factors (*immediate, delayed, recognition*, mean $r = .92$), the *set shifting and attention, speed* factors ($r = .94$). Interestingly, the *language* factor was also highly correlated with the memory domains (*immediate* $r = .94$; *delayed* $r = .88$; *recognition* $r = .84$).

Because in model VIII *immediate episodic memory* (MEI in Figure 1, panel A) was very highly correlated with *delayed episodic memory* and *language and fluency*, we made the decision to drop the items assessing immediate episodic memory from the factor model. As a consequence, our subsequent models (models IX and X in Table 4, and Panels B and C of Figure 1) are no longer nested within the previous multidimensional models and differences in model fit statistics (e.g., χ^2) are not comparable. Moreover, it is worth pointing out that we are not reducing the length of the HCAP battery because we must still ask participants the immediate recall questions in order to assess delayed recall and recognition. We also treated the indicators of *delayed episodic memory* and *recognition memory* as indicators of a single *memory* factor (model IX). This modification was motivated by the high correlation (.91) of these two specific factors in the multidimensional model and the presence of only two indicators for recognition memory. Additionally, we combined the *set shifting and attention, speed* factors into a single factor labeled *executive functioning*. This modified model (model X, Table 4) fit reasonably well against standard thresholds of model fit according to SRMR, RMSEA, and CFI (0.047, 0.046, and 0.945, respectively).

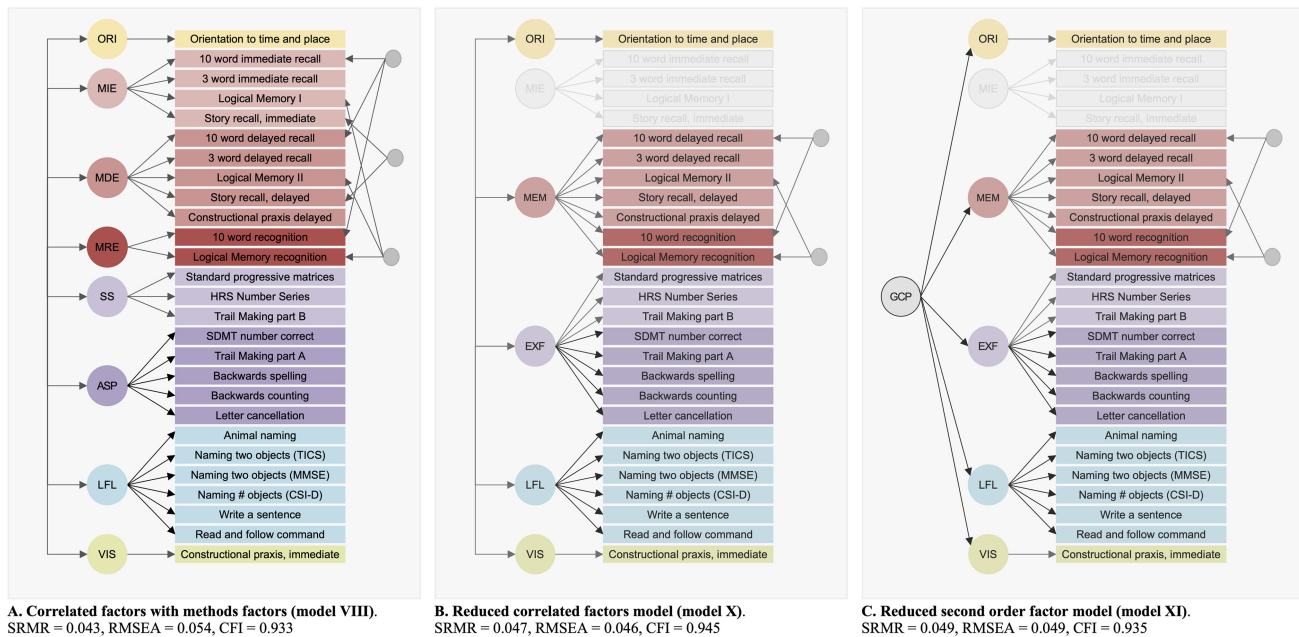


Figure 1. Summaries of model structure for models VIII, X, and XI. Figure 1 displays structure and model fit information for model VII (Panel A) displays a correlated factors model will all initially included indicators), model X (Panel B, correlated factors model removing immediate episodic memory and using a single factor for delayed and recognition memory and a single factor for set shifting and attention/speed), and Model XI, (Panel C) a second-order factor model based on Model X. Abbreviations: ORI, orientation; MEI, immediate episodic memory; MDE, delayed episodic memory; MRE, recognition memory; SS, set shifting; ASP, attention/speed; LFL, language and fluency; VIS, visuospatial; MEM, memory; EXF, executive functioning; GCP, general cognitive performance; SRMR, standardized root mean squared residual; RMSEA, root mean squared error of approximation; CFI, confirmatory fit index; TICS, Telephone Interview for Cognitive Status; MMSE, Mini-Mental State Examination; HRS, Health and Retirement Study; CSI-D, Community Screening Instrument for Dementia; SDMT, Symbol Digit Modalities Test.

We also wanted to evaluate the fit of a second-order factor model based on our final multidimensional model (Figure 1, panel C). We expect the second-order factor model will be useful for harmonization with other of the IPS, especially if partner studies do not include all of the items or subtests within a domain and/or if investigators wish to test the assumption of measurement invariance across IPS (Nichols et al., 2022). Results of this model fit are shown in Table 4 (model XI). This model estimates five fewer parameters, but the cost to model χ^2 is 259, suggesting a significant reduction in model fit. Nevertheless, the model returns similar estimates as for model X for SRMR (0.049), RMSEA (0.049), but (as with Model X) the CFI falls below typical thresholds for adequate fit (CFI = 0.935).

Factor score estimation

We obtained factor score estimates using the EAP method and Bayesian PVs (Asparouhov & Muthén, 2010). Estimates were derived from the single factor models for specific domains (corresponding to models IID, IIIA, V, Table 3), and from the second-order model for the general domain (model XI, Table 4). Results of these procedures are illustrated in Supplement Figure S2, which illustrates the density of EAP factor score estimate (black line) and the density of PVs (gray-filled density). What is interesting to note is where the EAP density is less than the PV density. These regions of the latent trait indicate areas where the standard error of measurement is high, or the reliability is low, for the estimated latent trait level.

As can be seen in Supplement Figure S2, the *memory*, *executive functioning*, and the *general cognitive performance* latent trait

estimates seem reasonably precise across the range of cognitive ability. The range of ability is based on the standard normal distribution. The values for language/fluency are less precise, particularly in the tails of the distribution. This is not surprising given this domain is comprised mostly of components with two or three response categories, except for *animal naming*. This specific domain was also the one with lowest reliability (Table 3).

Discussion

We applied factor analysis methods to a pre-selected set of neuropsychological performance indicators included in the HCAP battery of tests fielded by the HRS in 2016. Using an *a priori* conceptualization of the structure of the cognitive tests, we conducted a set of analyses guided by model fit and parameter estimates to derive a model of reasonably well-fitting domains of cognitive performance in a general and broadly inclusive sample of community dwelling older adults. Our final measurement model maps well on to well-characterized domains of *memory*, *executive functioning*, *language/fluency*, and *visuospatial* functioning.

Limitations of our approach and results are worth mentioning. Firstly, we commenced our analysis with a definite view as to the structure we were likely to observe among the tests included in the HCAP battery. While we allowed some aspects of model fit and factor correlation to drive final model structure, it is likely that if we had started with a different set of assumptions, made different decisions about which scores from tests to include, how to treat those variables analytically, or followed a more exploratory factor analysis method, we may have settled on a different factor structure in our final models. One of the strengths of basing this research

within the HRS is that the data are public access and as such other investigators can obtain the source variables and derive alternative solutions. Another limitation is that our analysis does not make use of complex sampling weights, as these were not defined for the HCAP sample at the time of analysis. It will be important to replicate these results when those sampling weights are available.

As a final limitation, readers are reminded that factor structures based on a general community sample, in which most participants are functioning within the range of cognitively normal, may not translate to clinical or impaired populations (c.f., Delis et al., 2003). It is likely that far more creative and informative use of the extensive HCAP battery could be used to identify persons with an existing or developing cognitive disorder.

In terms of utility for identifying persons with probable dementia within a DSM-5 neurocognitive disorder framework, our results hold promise. We have identified 5 domains of functioning (*orientation, memory, executive functioning, language/fluency, visuo-spatial*). While two of these domains are assessed with single items. The remaining 3 are assessed with multiple items and have high reliability on the basis of the omega coefficient, but the marginal reliability (as indicated by the standard error of measurement) around 2 standard deviations below the mean is 0.83 for *memory*, 0.92 for *executive*, and 0.56 for *language/fluency*. These results are encouraging for *executive* and *memory*, but suggest the measurement of the *language/fluency* domain is in need of further development.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S135561772300019X>

Acknowledgments. This study was not preregistered.

Funding statement. The HRS is funded by the National Institute on Aging (DW, U01-AG009740), with supplemental funding from the Social Security Administration. The HCAP is funded by the National Institute on Aging (KL, U01-AG058499).

Conflicts of interest. The authors report no conflicts of interest.

References

- Alzheimer's Association. (2019). 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 15, 321–387. doi: [10.1016/j.jalz.2019.01.010](https://doi.org/10.1016/j.jalz.2019.01.010)
- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders, (DSM-5): American Psychiatric Publishers.
- Armitage, S. (1946). Analysis of certain psychological tests used for the evaluation of brain damage. *Psychological Monographs*, 60, 1–48. doi: [10.1037/h0093567](https://doi.org/10.1037/h0093567)
- Asparouhov, T., and Muthén, B. (2010). Plausible values for latent variables using Mplus. Retrieved from <http://www.statmodel.com/download/Plausible.pdf>
- Asparouhov, T., and Muthén, B. (2018, 2 May). SRMR in Mplus. Retrieved from <http://www.statmodel.com/download/SRMR2.pdf>
- Brandt, J., Spencer, M., and Folstein, M. (1988). The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 1, 111–117. <https://journals.lww.com/cogbehavneurol/toc/1988/00120>
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Publications.
- Cutler, S. J. (2015). Worries about getting Alzheimer's: Who's concerned? *American Journal of Alzheimer's Disease & Other Dementias*, 30, 591–598. doi: [10.1177/1533317514568889](https://doi.org/10.1177/1533317514568889)
- Delis, D. C., Jacobson, M. W., Bondi, M. W., Hamilton, J. M., and Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: lessons from memory assessment. *Journal of the International Neuropsychological Society*, 9, 936–946. doi: [10.10170S1355617703960139](https://doi.org/10.10170S1355617703960139)
- Fisher, G. G., McArdle, J. J., McCammon, R. J., Sonnega, A., and Weir, D. R. (2014). *New measures of fluid intelligence in the HRS: Quantitative reasoning, verbal reasoning, verbal fluency* (DR-027). Retrieved from Ann Arbor: <http://hrsonline.isr.umich.edu/sitedocs/userg/dr-027b.pdf>
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). 'Mini-mental state.' A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198. doi: [10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6)
- Hall, K., Gao, S., Emsley, C. L., Ogunniyi, A. O., Morgan, O., and Hendrie, H. C. (2000). Community screening interview for dementia (CSI 'D'); performance in five disparate study sites. *International Journal of Geriatric Psychiatry*, 15, 521–531. doi: [10.1002/1099-1166\(200006\)15:6<521::aid-gps182>3.0.co;2-f](https://doi.org/10.1002/1099-1166(200006)15:6<521::aid-gps182>3.0.co;2-f)
- Hall, K., Hendrie, H., Brittain, H., Norton, J., Rodgers, D., Prince, C., Pillaym, N., Blue, A. W., Kaufert, J. M., and Nath, A. (1993). The development of a dementia screening interview in two distinct languages. *International Journal of Methods in Psychiatric Research*, 3, 1–28. doi: [10.1002/1099-1166\(200006\)15:6%3C521::aid-gps182%3E3.0.co;2-f](https://doi.org/10.1002/1099-1166(200006)15:6%3C521::aid-gps182%3E3.0.co;2-f)
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
- Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., and Langa, K. M. (2013). Monetary Costs of Dementia in the United States. *The New England journal of medicine*, 368, 1326–1334. doi: [10.1056/NEJMsa1204629](https://doi.org/10.1056/NEJMsa1204629)
- Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological methods & research*, 44, 486–507. doi: [10.1177/0049124114543236](https://doi.org/10.1177/0049124114543236)
- Lachman, M. E., Agrigoroaei, S., Tun, P. A., and Weaver, S. L. (2013). Monitoring cognitive functioning: Psychometric properties of the Brief Test of Adult Cognition by Telephone. *Assessment*, 21, 404–417. doi: [10.1177/1073191113508807](https://doi.org/10.1177/1073191113508807)
- Langa, K., Plassman, B. L., Wallace, R. B., Herzog, A. R., Heeringa, S. G., Ofstedal, M. B., Burke, J. R., Fisher, G. G., Fultz, N. H., Hurd, M. D., Potter, G. G., Rodgers, W. L., Steffens, D. C., Weir, D. R., and Willis, R. J. (2005). The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology*, 25, 181–191. doi: [10.1159/000087448](https://doi.org/10.1159/000087448)
- Langa, K., Ryan, L., McCammon, R., Jones, R. N., Manly, J., Levine, D. A., Sonnega, A., Farron, M., and Weir, D. (2020). The Health and Retirement Study Harmonized Cognitive Assessment Protocol (HCAP) Project: Study Design and Methods. *Neuroepidemiology*, 54, 64–74. doi: [10.1159/000503004](https://doi.org/10.1159/000503004)
- Mejia-Arango, S., Nevarez, R., Michaels-Obregon, A., Trejo, B., Mendoza-Alvarado, L., Sosa-Ortiz, A. L., Martinez-Ruiz, A., and Wong, R. (2020). The Mexican Cognitive Aging Ancillary Study (Mex-Cog): Study Design and Methods. *Archives of Gerontology and Geriatrics*, 91, 104210. doi: [10.1016/j.archger.2020.104210](https://doi.org/10.1016/j.archger.2020.104210)
- Meng, Q., Wang, H., Strauss, J., Langa, K. M., Chen, X., Wang, M., Qu, Q., WeiKuang, W., and Zhang, N. (2019). Validation of neuropsychological tests for the China Health and Retirement Longitudinal study harmonized cognitive assessment protocol. *International Psychogeriatrics*, 31, 1709–1719. doi: [10.1017/S1041610219000693](https://doi.org/10.1017/S1041610219000693)
- Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., and Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39(9), 1159–1165. doi: [10.1212/WNL.39.9.1159](https://doi.org/10.1212/WNL.39.9.1159)
- Nichols, E. L., Cadar, D., Lee, J., Jones, R. N., and Gross, A. L. (2022). Linear linking for related traits (LLRT): A novel method for the harmonization of cognitive domains with no or few common items. *Methods*, 204, 179–188. doi: [10.1016/j.ymeth.2021.11.011](https://doi.org/10.1016/j.ymeth.2021.11.011)
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill College Division.
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: why and how to abandon Cronbach's alpha and the route towards more

- comprehensive assessment of scale quality. *European Health Psychologist*, 16, 56–69. doi: [10.31234/osf.io/h47fv](https://doi.org/10.31234/osf.io/h47fv)
- Plassman, B. L., Langa, K. M., Fisher, G. G., Heeringa, S. G., Weir, D. R., Ofstedal, M. B., Burke, J. R., Hurd, M. D., Potter, G. G., Rodgers, W. L., Steffens, D. C., McArdle, J. J., Willis, R. J., and Wallace, R. B. (2008). Prevalence of cognitive impairment without dementia in the United States. *Annals of Internal Medicine*, 148, 427–434. doi: [10.7326/0003-4819-148-6-200803180-00005](https://doi.org/10.7326/0003-4819-148-6-200803180-00005)
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute of Educational Research.
- Raven, J. (1981). *Manual for Raven's progressive matrices and vocabulary scales. Research Supplement no. 1: The 1979 British standardization of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, together with comparative data from earlier studies in the UK, US, Canada, Germany, and Ireland*. Oxford, England: Oxford Psychologists Press.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696. doi: [10.1080/00273171.2012.715555](https://doi.org/10.1080/00273171.2012.715555)
- Richards, M., Kuh, D., Hardy, R., and Wadsworth, M. (1999). Lifetime cognitive function and timing of the natural menopause. *Neurology*, 53, 308–308. doi: [10.1212/WNL.53.2.308](https://doi.org/10.1212/WNL.53.2.308)
- Scherr, P. A., Albert, M. S., Funkenstein, H. H., Cook, N. R., Hennekens, C. H., Branch, L. G., White, L. R., Taylor, J. O., and Evans, D. A. (1988). Correlates of cognitive function in an elderly community population. *American Journal of Epidemiology*, 128, 1084–1101. doi: [10.1093/oxfordjournals.aje.a115051](https://doi.org/10.1093/oxfordjournals.aje.a115051)
- Schrank, F. A., and Flanagan, D. P. (2003). *WJ III clinical use and interpretation: Scientist-practitioner perspectives*. Academic Press.
- Smith, A. (1982). *Symbol digit modalities test*. Los Angeles: Western Psychological Services.
- Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W., and Weir, D. R. (2014). Cohort profile: the health and retirement study (HRS). *International journal of epidemiology*, 43, 576–585. doi: [10.1093/ije/dyu067](https://doi.org/10.1093/ije/dyu067)
- Steptoe, A., Breeze, E., Banks, J., and Nazroo, J. (2013). Cohort profile: the English longitudinal study of ageing. *International journal of epidemiology*, 42, 1640–1648. doi: [10.1093/ije/dys168](https://doi.org/10.1093/ije/dys168)
- Wechsler, D. (1987). *Wechsler Memory Scale-Revised manual*. San Antonio: Psychological Corporation.
- Weir, D., McCammon, R., Ryan, L., and Langa, K. (2014). *Cognitive test selection for the harmonized cognitive assessment protocol*. Retrieved from Ann Arbor: http://hrsonline.isr.umich.edu/sitedocs/userg/HCAP_testselection.pdf
- Wilson, R. S., Mendes De Leon, C. F., Barnes, L. L., Schneider, J. A., Bienias, J. L., Evans, D. A., and Bennett, D. A. (2002). Participation in cognitively stimulating activities and risk of incident Alzheimer disease. *JAMA*, 287, 742–748. doi: [10.1001/jama.287.6.742](https://doi.org/10.1001/jama.287.6.742)
- Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.