

1 Running title: HCAP Harmonization
2 Harmonization of Later-Life Cognitive Function Across National Contexts: Results from the
3 Harmonized Cognitive Assessment Protocols (HCAPs)

4
5 Alden L. Gross, PhD¹ Chihua Li, PhD^{2,3} Emily M. Briceno, PhD⁴ Miguel Arce Rentería, PhD⁵
6 Richard N. Jones, ScD⁶ Kenneth M. Langa, MD^{2,7,8} Jennifer J. Manly, PhD⁵ Emma L. Nichols,
7 PhD⁹ David Weir, PhD² Rebeca Wong, PhD¹⁰ Lisa Berkman, PhD¹¹ Jinkook Lee*, PhD⁹ Lindsay
8 C. Kobayashi*, PhD^{2,3}

9 *Co-senior authors

10

11 ¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore,
12 MD, USA; Center on Aging and Health, Johns Hopkins University, Baltimore, MD, USA

13 ²Center for Social Epidemiology and Population Health, Department of Epidemiology, University
14 of Michigan School of Public Health, Ann Arbor, MI, USA

15 ³Survey Research Center, University of Michigan Institute for Social Research, Ann Arbor, MI,
16 USA

17 ⁴Department of Physical Medicine & Rehabilitation, University of Michigan Medical School, Ann
18 Arbor, MI, USA

19 ⁵Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of
20 Neurology, Columbia University College of Physicians and Surgeons, New York City,
21 NY, USA

22 ⁶Department of Psychiatry and Human Behavior, Warren Alpert Medical School, Brown
23 University, Providence RI, USA

24 ⁷Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

25 ⁸Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI, USA

26 ⁹Center for Economic and Social Research and Department of Economics, University of
27 Southern California, Los Angeles, CA, USA

28 ¹⁰University of Texas Medical Branch, Sealy Center on Aging, Galveston TX, United States

29 ¹¹Harvard Center for Population and Development Studies and Department of Social and
30 Behavioral Sciences, TH Chan School of Public Health, Cambridge MA.

31

32 **Corresponding author:** Alden L. Gross, agross14@jhu.edu; Phone: 410-474-3386;
33 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 616 N. Wolfe
34 St., Baltimore, MD 21205, USA

35

36 Words: 268/300 (Abstract); 3495/3500 (text); 4 tables; 3 Figures; 3 Supplemental Tables; 1
37 Supplemental Figure

38

39

Abstract

40 **Background:** The Harmonized Cognitive Assessment Protocol (HCAP) is an innovative
41 instrument for cross-national comparisons of later-life cognitive function, yet its suitability across
42 diverse populations is unknown. We aimed to harmonize general and domain-specific cognitive
43 scores from HCAPs across six countries, and evaluate precision and criterion validity of the
44 resulting harmonized scores.

45 **Methods:** We statistically harmonized general and domain-specific cognitive function across the
46 six publicly available HCAP partner studies in the United States, England, India, Mexico, China,
47 and South Africa (N=21,141). We used an item banking approach that leveraged common
48 cognitive test items across studies and tests that were unique to studies, as identified by a
49 multidisciplinary expert panel. We generated harmonized factor scores for general and domain-
50 specific cognitive function using serially estimated graded-response item response theory (IRT)
51 models. We evaluated precision of the factor scores using test information plots and criterion
52 validity using age, gender, and educational attainment.

53 **Findings:** IRT models of cognitive function in each country fit well. We compared measurement
54 reliability of the harmonized general cognitive function factor across each cohort using test
55 information plots; marginal reliability was high ($r > 0.90$) for 93% of respondents across six
56 countries. In each country, general cognitive function scores were lower with older ages and
57 higher with greater levels of educational attainment.

58 **Interpretation:** We statistically harmonized cognitive function measures across six large,
59 population-based studies of cognitive aging in the US, England, India, Mexico, China, and South
60 Africa. Precision of the estimated scores was excellent. This work provides a foundation for
61 international networks of researchers to make stronger inferences and direct comparisons of
62 cross-national associations of risk factors for cognitive outcomes.

63 **Funding:** National Institute on Aging (R01 AG070953, R01 AG030153, R01 AG051125, U01
64 AG058499; U24 AG065182; R01AG051158)

65

66 Introduction

67 Alzheimer's disease and related dementias, for which cognitive decline is the hallmark
68 symptom, are a major global public health, clinical, and policy challenge. Although much
69 research on risk and protective factors for dementia has been conducted in high-income
70 countries it is anticipated that three-quarters of the 152 million persons with dementia will be
71 living in low- and middle-income countries in the coming decades.¹⁻³ Differences in the
72 distributions of potential risk factors and cultural and demographic factors that impact dementia
73 across countries makes cross-national research imperative.

74
75 To facilitate cross-national comparisons of later-life cognitive outcomes, measurement
76 instruments must validly measure cognitive function across populations with diverse cultural,
77 educational, social, economic, and political contexts. To that end, the Harmonized Cognitive
78 Assessment Protocol (HCAP) has been developed and implemented in International Partner
79 Studies (IPS) of the US Health and Retirement Study (HRS).⁴ The HCAP network represents
80 the largest concerted global effort to-date to conduct harmonized large-scale population-
81 representative studies of cognitive aging and dementia.

82
83 Although the HCAP was designed collaboratively to ensure its comparability across countries,
84 necessary adaptations were made to its individual test items, test administrations, and scoring
85 procedures to accommodate different languages, cultures, and levels of literacy and numeracy
86 of its respondents.¹⁵ The impacts of these adaptations on the performance, reliability, and
87 validity of the HCAP cognitive test items are only beginning to be understood^{15,20}, which may
88 limit cross-national utility of the HCAP battery. The goal of this study was to conduct statistical
89 harmonization of the HCAP instruments fielded in the United States, England, India, Mexico,
90 China, and South Africa. Statistical harmonization involved assigning cognitive test items to
91 domains, determining which test items were common and which were unique across countries,
92 deriving harmonized factor scores for general and specific cognitive domains, and estimating
93 the reliability and validity of the harmonized factor scores.

94

95 Methods

96 Participants

97 The Health and Retirement Study in the United States (HRS) and its International Partner
98 Studies (IPS) are large, population-based studies of aging. Between 2016 and 2019, six such
99 studies administered Harmonized Cognitive Assessment Protocols (HCAPs) to participants from

100 each core IPS. They included the HRS, the English Longitudinal Study on Ageing (ELSA), the
101 Longitudinal Aging Study in India (LASI), the Mexican Health and Aging Study (MHAS), the
102 China Health and Retirement Longitudinal Study (CHARLS), and Health and Aging in Africa: A
103 Longitudinal Study of an INDEPTH Community in South Africa (HAALSI). Details of HCAP
104 administration in each cohort, eligibility, timing, and sample sizes are summarized in
105 **Supplemental Table 1.**⁴⁻⁹ The HCAP aims to provide a detailed assessment of cognitive
106 function of older adults that is flexible, yet comparable across populations in countries with
107 diverse cultural, educational, social, economic, and political contexts. The HCAP network
108 ultimately intends to provide comparable estimates of dementia and mild cognitive impairment
109 prevalence across countries, and to exploit cross-national variation in key risk and protective
110 factors to better understand the determinants of later-life cognition, cognitive aging, and
111 dementia.⁶¹

112
113 The HCAPs in the US and Mexico randomly sampled participants from the core studies who did
114 not need a proxy interview in the previous core interview wave, and HRS-HCAP further included
115 a random sampling of N=219 participants interviewed by proxy in the 2016 HRS core wave.^{4,7}
116 To ensure adequate sample sizes of participants with dementia, HCAPs in England, India, and
117 South Africa recruited participants with low cognitive function.^{5,8,9} All parent studies were
118 nationally representative, with the exception of HAALSI in South Africa, which is a
119 representative sample from the Agincourt sub-district in northeastern South Africa.¹⁰ All
120 participants consented to research and IRBs at local institutions approved each IPS and its
121 respective HCAP.

122
123 Variables
124 Cognitive test battery. Details of the original battery of 17 cognitive tests in HCAP are available
125 in Langa et al.⁴ By design, each HCAP study administered as close to the same battery of tests
126 as was feasible. We granularized these batteries to 30-51 cognitive test indicators in each
127 HCAP, as shown in Figure 1 and Supplemental Table 2.

128
129 Each cognitive test item was assigned to a domain based on *a priori* theory, combined with
130 empirical analyses demonstrating which test items fit well into a domain using factor analysis
131 methods.¹¹⁻¹³ Assigning test items to domains is essential to statistical harmonization, also
132 referred to as co-calibration, as this process relies on the presence of equivalent or comparable
133 cognitive test items across one or more studies. If cognitive test items are presumed to be the

134 same across HCAP studies, but are in fact different (e.g., a different test; the same test with
135 different stimuli, administration, or scoring procedures), such methodological differences could
136 contribute to artifactual differences in the observed cognitive scores between studies. These
137 artifactual differences would imperil the quality of cross-national inferences drawn from the
138 derived summary cognitive scores.

139
140 To determine the comparability of cognitive test items across HCAPs, we convened an expert
141 panel of neuropsychologists, epidemiologists, persons with cultural/linguistic expertise, and
142 psychometricians with working knowledge of cross-cultural neuropsychology and administration
143 of the HCAPs to conduct pre-statistical harmonization of cognitive test items. This group used
144 available materials including codebooks, interviewer training manuals, and personal
145 communication with study investigators and coordinators to document differences in test item
146 content and administration across HCAPs and to determine which differences were substantial
147 and whether cultural or language demands differed for each test. Considerations made for each
148 cognitive test item have been described previously.^{14,15} Using the HRS HCAP as the reference,
149 two neuropsychologists rated items from all other HCAPs as a confident linking item that is very
150 likely to be comparable, a tentative linking item, or a non-linking item based on available
151 information.¹⁵

152
153 **Covariates**

154 Age, gender, and highest educational attainment were collected in core IPS interviews. We
155 scaled educational attainment in each country to the 2011 International Standard Classification
156 of Education.¹⁶

157
158 **Analysis plan**

159 **Descriptive analyses.** We described demographic characteristics and cognitive tests using
160 means with standard deviations and counts with percentages. We identified overlapping and
161 unique cognitive test items by HCAP.

162
163 **HCAP-specific factor analyses.** We estimated confirmatory factor analysis (CFA) models for
164 cognitive domains of general cognitive function, memory, executive function, orientation, and
165 language separately in each HCAP study, without regard to items in common across studies.
166 The goal of this series of psychometric models was to illustrate that similar organizations of
167 cognitive test items fit well across countries.^{12,17} We ascertained model fit using three standard

168 absolute fit statistics: the Root Mean Square Error of Approximation (RMSEA), Comparative Fit
169 Index (CFI), and Standardized Root Mean Residual (SRMR).¹⁸ When possible, we attempted to
170 improve model fit through the use of bifactor models to address additional correlations between
171 theoretically similar items (e.g., Trail Making Test, parts A and B, or immediate and delayed
172 recall).^{11,19} Using the combination of these three fit statistics, we characterized model fit as
173 perfect, good, adequate, or poor, using previously described rubric.¹²

174

175 **Statistical harmonization via item banking.** Following the estimation of CFAs within each
176 HCAP study, we statistically harmonized scores for each cognitive domain across countries
177 using an item banking approach.²⁰ A flowchart in **Supplemental Figure 1** illustrates this
178 approach. For each cognitive domain, we serially estimated CFAs in each study, sequentially
179 fixing model parameters for items determined to be comparable to those in previous studies to
180 their corresponding values from previous studies. The order of studies was HRS-HCAP, ELSA-
181 HCAP, LASI-DAD, Mex-Cog, CHARLS-HCAP, and HAALSI-HCAP. LASI-DAD was split into
182 literate (N=1777, 43%) and illiterate (N=2139, 57%) subgroups due to administration differences
183 in some tests. Because there is no natural scaling in latent variable space, the mean and
184 variance of the factor score (general cognitive function, memory, executive function, orientation,
185 or language) were set to 0 and 1, respectively, beginning with the HRS-HCAP as the reference.
186 The factor score is estimated based on all of the items in the CFA. The CFA models estimated
187 two relevant parameters for each cognitive test item: factor loadings and item thresholds (for
188 categorical items) or intercepts (for continuous items). Factor loadings characterize how strongly
189 correlated a given cognitive test item is with the other cognitive test items in the model. In
190 general, loadings between 0.3 and 0.9 indicate an item is meaningfully related to the other
191 items without overwhelming others in the model.^{11,19} Item thresholds characterize the location
192 along the factor at which the cognitive test item provides maximal information of underlying
193 cognitive function.

194

195 Loadings and thresholds/intercepts from the CFA models in HRS-HCAP were saved for use in
196 serially estimated CFAs in subsequent HCAPs (**Supplemental Figure 1**). After estimating a
197 CFA model for the HRS-HCAP study, we next estimated a CFA model in ELSA-HCAP, in which
198 item parameters for cognitive test items in common with the HRS-HCAP were constrained to
199 those observed in the HRS-HCAP, and the mean and variance of the underlying trait were freely
200 estimated. The same process was repeated for all other HCAPs. Parameters for cognitive test
201 items from a given HCAP study that were not yet in the item bank were freely estimated, then

202 saved in the item bank for use when the next HCAP study was added to the item bank. In a final
203 factor score-estimating CFA model for each cognitive domain, we estimated a CFA in the
204 pooled sample of all HCAP studies, in which all parameters were fixed to previous values. We
205 evaluated marginal reliability of the measurement models of each domain, calculated from the
206 standard error of the measurement, in each HCAP.

207

208 **Differential item functioning.** The validity of the cross-national harmonization of cognitive
209 function depends on the availability of common, equivalent cognitive test items across studies.
210 While our expert panel identified equivalent linking items, it is possible to miss test differences
211 that may have not been documented, that are due to unforeseen cultural differences, or for
212 which there was insufficient documentation available. Thus, we statistically tested for differential
213 item functioning (DIF) among candidate equivalent linking items between the HRS-HCAP and
214 each study, by cognitive domain. We used multiple indicator, multiple cause (MIMIC) models to
215 evaluate DIF by HCAP study membership.^{15,24} Briefly, we first tested DIF amongst cognitive test
216 items rated as confident linking items. Next, we tested for DIF among cognitive items rated as
217 tentative linking items, treating as anchors the confident items that showed no DIF in the prior
218 analysis. The magnitude of DIF attributable to a given cognitive test item is represented by an
219 odds ratio (OR) for an item on an indicator for study membership; we considered non-negligible
220 DIF as an OR outside the range of 0.66 to 1.5.²⁵ Large impact of DIF on participant's domain-
221 level scores, called salient DIF, was evaluated by taking the difference between DIF-adjusted
222 and non-DIF-adjusted scores, via enabling items that showed DIF to have different
223 measurement model parameters across studies, and counting how many participant scores
224 would differ by more than 0.3 SD units.²⁶

225

226 **Validation.** To evaluate construct validity, we evaluated the patterns of factor scores by age,
227 gender, and educational attainment by regressing general cognitive function on each of these
228 characteristics, adjusting for the other characteristics. We hypothesized better cognitive function
229 on average at younger ages and with more educational attainment.^{21,22} With respect to gender,
230 we hypothesized women are more disadvantaged in LMIC settings compared to men, given
231 known gender-based societal inequalities in these settings that apply to determinants of later-life
232 cognitive health such as educational opportunities.²³

233

234 Descriptive analyses were conducted using Stata (Version 17, Stata Corp, College Station, TX).
235 Factor analysis was conducted using Mplus.²⁷

236

237

Results

238 **Descriptive analyses.** There were N=21,141 participants across the six HCAP studies (**Table**
239 **1**).

240

241 **Figure 1** (and **Supplemental Table 2**) displays the cognitive test items, stratified by cognitive
242 domain to which tests were assigned. Of the 78 cognitive test items administered, 12 were
243 judged by experts to be comparably administered in every HCAP. Overall, 15 distinct test items
244 were assigned to the orientation domain, 14 distinct test items were assigned to the memory
245 domain, 26 distinct cognitive test items were assigned to the executive functioning domain, and
246 23 distinct cognitive test items were assigned to the language domain. For a given test item in
247 each column, the presence of factor loadings from the item banking approach reflect decisions
248 about the comparability of items made during prestatistical harmonization. For example, for
249 orientation, we determined that asking for one's municipality in HAALSI-HCAP was comparable
250 to asking for one's district in LASI-DAD. Notably, our prestatistical team decided *a priori* that the
251 CERAD word recall test was administered differently in the HRS-HCAP and ELSA-HCAP as
252 compared to LASI-DAD, Mex-Cog, and HAALSI-HCAP because participants in the former two
253 countries were presented with the words both verbally and visually, but in the latter three
254 countries, participants were presented with the words only verbally. Moreover, while all studies
255 presented the words verbally, there was variation in the order in which words were presented
256 (i.e., alternating per trial vs fixed).

257

258 **HCAP-specific factor analyses.** **Table 2** displays model fit statistics for measurement models
259 of each of the five cognitive domains, by each of the seven study groups (six HCAP studies with
260 LASI-DAD stratified by literacy). Of these 35 measurement models, 31% (11 models) were of
261 perfect or good fit, 60% (21 models) were of adequate fit, and the remaining 9% (3 models)
262 were of poor fit. Two of the three poorly fitting models were in the general cognitive function
263 domain. Ultimately, we proceeded with these factor structures because most model fits were
264 good or adequate.²⁸

265

266 **Statistical harmonization via item banking.** The factor scores for general cognitive function,
267 memory, language, and executive function were approximately normally distributed in each
268 study (**Figure 2**). In contrast, the orientation factor showed a strong ceiling effect in each study
269 (**Figure 2**). These ceiling effects are explained by low reliabilities (internal consistency based on

270 the standard error of the measurement model) of the factor scores for the orientation domain.
271 The orientation factor provided low precision above scores of 0 as more than half of participants
272 in HRS-HCAP and a plurality of those in other studies were at the ceiling because they
273 answered all orientation items correctly (**Figure 3**). In comparison, across a broad range of
274 values, the reliabilities of the general cognitive function and memory factors are uniformly high
275 (above $r=0.9$) for each HCAP between scores of -4 and 2, which encompasses over 90% of
276 respondent scores for those domains. The language factor exhibited higher reliability at lower
277 levels of language ability compared to higher levels, reflecting that almost all the language
278 items, with the exception of animal fluency, tended to be easier questions about naming. For
279 executive function, reliability was high for all studies except CHARLS-HCAP, which had just two
280 test items measuring executive function.

281
282 **Differential item functioning.** Evidence of DIF by study was present across cognitive test
283 items (**Supplemental Table 3**). Of 78 unique cognitive test items, 23 showed DIF between
284 HRS-HCAP and another study. Of these, 12 assessed language. With respect to the impact of
285 DIF, 16.3% (N=290) of LASI-DAD orientation scores, 51.9% (N=326) of HAALSI-HCAP
286 orientation scores, and 68.4% (N=6,668) of CHARLS-HCAP language factor score estimates
287 demonstrated salient DIF (i.e., estimates differed by 0.3 units or more before vs. after
288 accounting for DIF). Subsequent analyses, removing each item as a linking item one at a time,
289 revealed that orientation to year was entirely responsible for all salient or impactful DIF in
290 orientation for both LASI-DAD and HAALSI-HCAP (performance on this item was much lower in
291 these studies, controlling for underlying orientation ability). Most of the salient DIF in CHARLS
292 HCAP's language domain could be attributed to differences in two items: naming a described
293 cactus and reading and following a command. After removing these items as linking items
294 between CHARLS HCAP and other studies (**Table 3**), 16.8% of participant scores were affected
295 by adjustment for DIF in the language domain for CHARLS-HCAP. Otherwise, less than 6% of
296 scores for any domain in any study was considerably impacted by DIF adjustment. **Figure 1**
297 reflects decisions from DIF analyses to relax assumptions of item equivalence for orientation to
298 year between LASI-DAD and HAALSI-HCAP with other studies, and for naming a described
299 cactus and reading and following a command between CHARLS-HCAP and other studies.

300
301 **Validation.** Patterns of cognitive function aligned with hypothesized expectations: general
302 cognitive function scores were, on average, lower at older ages and higher with greater
303 education (**Table 4**). Women had higher average general cognitive function scores than men in

304 the HRS-HCAP and ELSA-HCAP, but lower average scores than men in LASI-DAD, Mex-Cog,
305 CHARLS-HCAP, and HAALSI-HCAP.

306

307 Discussion

308 We investigated the performance of common and unique cognitive test items administered to
309 21,141 older adults across six large harmonized studies of aging in the US, England, India,
310 Mexico, China, and South Africa. We demonstrated these cognitive test items empirically reflect
311 comparable domains of cognitive function among older adults living across these countries, they
312 are reliable and valid measures of cognitive function, and useful for population-based research.
313 Most importantly, we overcame differences in test administration due to language, literacy, and
314 numeracy to statistically harmonize general and domain-specific cognitive function across these
315 countries.

316

317 Over the past decade, a growing number of cross-national studies have examined risk factors of
318 cognitive function decline and dementia, mostly using data from the HRS and its IPS.^{29–33} Risk
319 factors examined in these studies have included socioeconomic characteristics, health
320 behaviors, physical and mental health conditions, and telomere length.^{21,34–55} However, none of
321 them conducted in-depth statistical harmonization of cognitive test items.

322

323 High-quality, harmonized scores for general and domain-specific cognitive function are crucial
324 tools to promote valid cross-national comparisons of predictors and outcomes of cognitive aging
325 in a rapidly aging world.¹⁷ A recent Lancet Commission report identified 12 risk factors that had
326 strong evidence of a causal risk for dementia: low education, hearing impairment, traumatic
327 brain injury, hypertension, diabetes, excessive alcohol use, obesity, smoking, depression, social
328 isolation, physical inactivity, and air pollution.⁵⁶ The harmonized cognitive function scores
329 generated here can be used in pooled analyses to evaluate whether these risk factors have
330 similar effects on cognitive function across global settings. Such knowledge could facilitate
331 identification of contextual risk-modifying factors that could be intervened upon to reduce the
332 risk of dementia in certain populations.⁵⁷ Further, common cognitive phenotypes could be used
333 to improve the quality of population attributable risk estimates. Finally, common cognitive
334 phenotypes could be used in dementia algorithms that are applied cross-nationally to generate
335 prevalence and incidence estimates that are truly comparable across national settings.⁵⁸

336

337 Prestatistical harmonization accompanied by statistical testing for differential item functioning
338 (DIF) were two essential steps to the harmonization goal of this study. DIF can be introduced by
339 methodological differences in test administration or scoring across studies, in addition to
340 population-level differences that may alter responses to equivalent test items (e.g., differences
341 in literacy, numeracy, and language). We evaluated the comparability of cognitive test items
342 using a multidisciplinary team, which was a crucial component of this harmonization work.
343 However, an expert's ability to identify measurement differences in cognitive test items across
344 languages and cultures in items depends on the quality of available study documentation and
345 level of expertise regarding the population under study. We are confident in our pre-statistical
346 harmonization given the available documentation and our team's level of expertise, but
347 adequate documentation is crucial. Statistical DIF testing identified only four of 20 domain-by-
348 country categories in which the DIF made a difference for more than 10% of the sample,
349 however in these cases the DIF proved critical to the estimation of scores.

350
351 Strengths of this study include nationally representative sampling (in the case of HAALSI,
352 regionally representative sampling) and comprehensive cognitive phenotyping with a common
353 protocol. All data are publicly available (see **Supplemental Table 1**). Our harmonization
354 approach based on item banking is readily scalable: as data from more HCAPs are released or
355 become available, and as longitudinal data from existing HCAPs become available, they can be
356 readily added to our item bank to be harmonized alongside the data shown here. Alongside
357 these strengths, there are notable limitations. The quality of the linking between studies is best
358 when there are more cognitive test items with richer distributions. This poses challenges when
359 domains largely include relatively easy dichotomous items (e.g. language and orientation). A
360 further limitation is that while we identified DIF, it was outside the scope of this study to
361 characterize possible reasons for DIF across HCAP batteries in each item. This is a worthwhile
362 aim for future research, especially as test batteries are adapted to additional countries and
363 contexts.

364
365 In conclusion, the HCAP suite of cognitive test items administered in the US, England, India,
366 Mexico, China, and South Africa reflects a common structure of general- and domain-specific
367 cognitive function across these diverse countries. Despite common protocols, there were
368 necessary item adaptations to account for language, literacy, numeracy, and cultural differences
369 across participating countries. Statistical harmonization involving an item-banking approach with
370 identification of common and unique items allowed for the construction of reliable and valid

371 factor scores that account for these differences. Future cross-national comparisons of risk
372 factors for cognitive aging outcomes, estimates of dementia prevalence and incidence, and
373 estimates of population attributable fractions of risk factors should consider using harmonized
374 factor scores to improve the quality of their analyses.

375

376

377 **References**

- 378 1. Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, et al. Estimation of the
379 global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the
380 Global Burden of Disease Study 2019. *Lancet Public Health* 2022;**7**(2).
- 381 2. Patterson C. World Alzheimer Report 2018 - The state of the art of dementia research: New
382 frontiers. *Alzheimer's Disease International (ADI): London, UK* 2018;
- 383 3. Prince M, Wimo A, Guerchet M, Gemma-Claire A, Wu YT, Prina M. World Alzheimer Report 2015:
384 The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends.
385 *Alzheimer's Disease International* 2015;
- 386 4. Langa KM, Ryan LH, McCammon RJ, Jones RN, Manly JJ, Levine DA, et al. The Health and
387 Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and
388 Methods. *Neuroepidemiology* 2020;**54**(1).
- 389 5. Bassil DT, Farrell MT, Wagner RG, Brickman AM, Glymour MM, Langa KM, et al. Cohort Profile
390 Update: Cognition and dementia in the Health and Aging in Africa Longitudinal Study of an
391 INDEPTH community in South Africa (HAALSI dementia). *Int J Epidemiol* 2022;**51**.
- 392 6. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: The China health and retirement
393 longitudinal study (CHARLS). *Int J Epidemiol* 2014;**43**(1).
- 394 7. Mejia-Arango S, Nevarez R, Michaels-Obregon A, Trejo-Valdivia B, Mendoza-Alvarado LR, Sosa-
395 Ortiz AL, et al. The Mexican Cognitive Aging Ancillary Study (Mex-Cog): Study Design and
396 Methods. *Arch Gerontol Geriatr* 2020;**91**.
- 397 8. Lee J, Khobragade PY, Banerjee J, Chien S, Angrisani M, Perianayagam A, et al. Design and
398 Methodology of the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia (LASI-
399 DAD). *J Am Geriatr Soc* 2020;**68**(S3).
- 400 9. Cadar D, Abell J, Matthews FE, Brayne C, David Batty G, Llewellyn DJ, et al. Cohort Profile Update:
401 The Harmonised Cognitive Assessment Protocol Sub-study of the English Longitudinal Study of
402 Ageing (ELSA-HCAP). *Int J Epidemiol* 2021;**50**.
- 403 10. Xavier Gómez-Olivé F, Montana L, Wagner RG, Kabudula CW, Rohr JK, Kahn K, et al. Cohort
404 profile: Health and ageing in Africa: A longitudinal study of an indepth community in South Africa
405 (HAALSI). *Int J Epidemiol* 2018;**47**.
- 406 11. Mukherjee S, Choi SE, Lee ML, Scollard P, Trittschuh EH, Mez J, et al. Cognitive domain
407 harmonization and cocalibration in studies of older adults. *Neuropsychology* 2022;
- 408 12. Gross AL, Khobragade PY, Meijer E, Saxton JA. Measurement and Structure of Cognition in the
409 Longitudinal Aging Study in India—Diagnostic Assessment of Dementia. *J Am Geriatr Soc*
410 2020;**68**(S3).
- 411 13. Arce Rentería, M., Manly, J., Vonk, J., Mejia Arango, S., Michaels Obregon, A., Samper-Ternent,
412 R., ... Tosto, G. (2022). Midlife Vascular Factors and Prevalence of Mild Cognitive Impairment in

- 413 Late-Life in Mexico. *Journal of the International Neuropsychological Society*, 28(4), 351-361.
414 doi:10.1017/S1355617721000539.
- 415 14. Briceño EM, Gross AL, Giordani BJ, Manly JJ, Gottesman RF, Elkind MSV, et al. Pre-Statistical
416 Considerations for Harmonization of Cognitive Instruments: Harmonization of ARIC, CARDIA, CHS,
417 FHS, MESA, and NOMAS. *Journal of Alzheimer's Disease* 2021;**83**(4).
- 418 15. Briceño EM, Arce Rentería M, Gross AL, Jones RN, Gonzalez C, Wong R, et al. A cultural
419 neuropsychological approach to harmonization of cognitive data across culturally and
420 linguistically diverse older adult populations. *Neuropsychology* 2022;
- 421 16. UNESCO Institute for Statistics. International standard classification of education : ISCED 2011.
422 *UNESCO Institute for Statistics* 2012;(84)
- 423 17. Kobayashi LC, Gross AL, Gibbons LE, Tommet D, Sanders RE, Choi SE, et al. You Say Tomato, I Say
424 Radish: Can Brief Cognitive Assessments in the U.S. Health Retirement Study Be Harmonized With
425 Its International Partner Studies? *J Gerontol B Psychol Sci Soc Sci* 2021;**76**(9).
- 426 18. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional
427 criteria versus new alternatives. *Structural Equation Modeling* 1999;**6**(1).
- 428 19. Gibbons RD, Bock RD, Hedeker D, Weiss DJ, Segawa E, Bhaumik DK, et al. Full-information item
429 bifactor analysis of graded response data. *Appl Psychol Meas* 2007;**31**(1).
- 430 20. Vonk JM, Gross AL, Zammit AR, Bertola L, Avila JF, Jutten RJ, et al. Cross-national harmonization
431 of cognitive measures across HRS HCAP (USA) and LASI-DAD (India). *PLoS One* 2022;**17**(2).
- 432 21. Clouston SAP, Smith DM, Mukherjee S, Zhang Y, Hou W, Link BG, et al. Education and Cognitive
433 Decline: An Integrative Analysis of Global Longitudinal Studies of Cognitive Aging. *The Journals of*
434 *Gerontology: Series B* 2020;**75**(7):e151–60.
- 435 22. Lenehan ME, Summers MJ, Saunders NL, Summers JJ, Vickers JC. Relationship between education
436 and age-related cognitive decline: a review of recent research. *Psychogeriatrics* 2015;**15**(2):154–
437 62.
- 438 23. Li R, Singh M. Sex differences in cognitive impairment and Alzheimer's disease. *Frontiers in*
439 *Neuroendocrinology* 2014;**35**.
- 440 24. Jones RN. Identification of measurement differences between English and Spanish language
441 versions of the Mini-Mental State Examination: Detecting differential item functioning using
442 MIMIC modeling. *Med Care* 2006;**44**(11 SUPPL. 3).
- 443 25. Zwick R. A REVIEW OF ETS DIFFERENTIAL ITEM FUNCTIONING ASSESSMENT PROCEDURES:
444 FLAGGING RULES, MINIMUM SAMPLE SIZE REQUIREMENTS, AND CRITERION REFINEMENT. *ETS*
445 *Research Report Series* 2012;(1).
- 446 26. Crane PK, Gibbons LE, Narasimhalu K, Lai JS, Cella D. Rapid detection of differential item
447 functioning in assessments of health-related quality of life: The Functional Assessment of Cancer
448 Therapy. *Quality of Life Research* 2007;**16**(1).

- 449 27. Muthen LK, Muthen BO. Mplus user's guide: 1998-2012. Los Angeles, CA: *Muthen & Muthen*
450 2012;8
- 451 28. Langa KM, Llewellyn DJ, Lang IA, Weir DR, Wallace RB, Kabeto MU, et al. Cognitive health among
452 older adults in the United States and in England. *BMC Geriatr* 2009;**9**(1):23.
- 453 29. Skirbekk V, Loichinger E, Weber D. Variation in cognitive functioning as a refined approach to
454 comparing aging across countries. *Proc Natl Acad Sci U S A* 2012;**109**(3): 770-4.
- 455 30. Savva GM, Maty SC, Setti A, Feeney J. Cognitive and physical health of the older populations of
456 England, the United States, and Ireland: international comparability of the Irish Longitudinal
457 Study on Ageing. *J Am Geriatr Soc* 2013; **61 Suppl 2**: S291-8.
- 458 31. Hong I, Reistetter TA, Díaz-Venegas C, Michaels-Obregon A, Wong R. Cross-national health
459 comparisons using the Rasch model: findings from the 2012 US Health and Retirement Study and
460 the 2012 Mexican Health and Aging Study. *Quality of Life Research* 2018;**27**(9): 2431-41.
- 461 32. de La Fuente J, Caballero FF, Verdes E, Rodríguez-Artalejo F, Cabello M, de La Torre-Luque A, et
462 al. Are younger cohorts in the USA and England ageing better? *Int J Epidemiol* 2019;**48** (6):1906-
463 13.
- 464 33. Lyu J, Lee C, Dugan E. Risk factors related to cognitive functioning: A cross-national comparison of
465 U.S. and korean older adults. *Int J Aging Hum Dev* 2014;**79**(1):81-101.
- 466 34. Weir D, Lay M, Langa K. Economic development and gender inequality in cognition: A comparison
467 of China and India, and of SAGE and the HRS sister studies. *J Econ Ageing* 2014;**4**: 114-25.
- 468 35. Angrisani M, Lee J, Meijer E. The gender gap in education and late-life cognition: Evidence from
469 multiple countries and birth cohorts. *J Econ Ageing* 2020;**16**.
- 470 36. Wu YT, Daskalopoulou C, Muniz Terrera G, Sanchez Niubo A, Rodríguez-Artalejo F, Ayuso-Mateos
471 JL, et al. Education and wealth inequalities in healthy ageing in eight harmonised cohorts in the
472 ATHLOS consortium: a population-based study. *Lancet Public Health*. 2020;**5**(7): e386-e94.
- 473 37. Faul JD, Ware EB, Kabeto MU, Fisher J, Langa KM. The Effect of Childhood Socioeconomic
474 Position and Social Mobility on Cognitive Function and Change among Older Adults: A
475 Comparison between the United States and England. *Journals of Gerontology - Series B*
476 *Psychological Sciences and Social Sciences* 2021;**76**(Suppl 1): S51-s63.
- 477 38. Stefler D, Prina M, Wu YT, Sánchez-Niubò A, Lu W, Haro JM, et al. Socioeconomic inequalities in
478 physical and cognitive functioning: Cross-sectional evidence from 37 cohorts across 28 countries
479 in the ATHLOS project. *J Epidemiol Community Health* (1978) 2021;**75**(10): 980-6.
- 480 39. Yu X, Langa KM, Cho TC, Kobayashi LC. Association of Perceived Job Insecurity With Subsequent
481 Memory Function and Decline Among Adults 55 Years or Older in England and the US, 2006 to
482 2016. *JAMA Netw Open* 2022;**5**(4): e227060.
- 483 40. Ma Y, Liang L, Zheng F, Shi L, Zhong B, Xie W. Association Between Sleep Duration and Cognitive
484 Decline. *JAMA Netw Open* 2020;**3**(9): e2013573.

- 485 41. Yoneda T, Lewis NA, Knight JE, Rush J, Vendittelli R, Kleineidam L, et al. The importance of
486 engaging in physical activity in older adulthood for transitions between cognitive status
487 categories and death: A coordinated analysis of 14 longitudinal studies. *J Gerontol A Biol Sci Med*
488 *Sci.* 2021;**76**(9): 1661-7.
- 489 42. Kelly A, Calamia M, Koval A, Terrera GM, Piccinin AM, Clouston S, et al. Independent and
490 interactive impacts of hypertension and diabetes mellitus on verbal memory: A coordinated
491 analysis of longitudinal data from England, Sweden, and the United States. *Psychol Aging*
492 2016;**31**(3): 262-73.
- 493 43. Maharani A, Dawes P, Nazroo J, Tampubolon G, Pendleton N. Visual and hearing impairments are
494 associated with cognitive decline in older people. *Age Ageing* 2018;**47**(4): 575-81.
- 495 44. Duggan EC, Piccinin AM, Clouston S, Koval A v., Robitaille A, Zammit AR, et al. A Multi-study
496 Coordinated Meta-analysis of Pulmonary Function and Cognition in Aging. *J Gerontol A Biol Sci*
497 *Med Sci.* 2019;**74**(11): 1793-804.
- 498 45. Yu Z bin, Zhu Y, Li D, Wu MY, Tang ML, Wang JB, et al. Association between visit-to-visit variability
499 of HbA1c and cognitive decline: a pooled analysis of two prospective population-based cohorts.
500 *Diabetologia* 2020;**63**(1): 85-94.
- 501 46. Jindra C, Li C, Tsang RSM, Bauermeister S, Gallacher J. Depression and memory function –
502 evidence from cross-lagged panel models with unit fixed effects in ELSA and HRS. *Psychol Med*
503 2022;**52**(8):1428–36.
- 504 47. Sutin AR, Luchetti M, Stephan Y, Terracciano A. Purpose in Life and Motoric Cognitive Risk
505 Syndrome: Replicable Evidence from Two National Samples. *J Am Geriatr Soc* 2021;**69**(2): 381-8.
- 506 48. Ma Y, Hua R, Yang Z, Zhong B, Yan L, Xie W. Different hypertension thresholds and cognitive
507 decline: a pooled analysis of three ageing cohorts. *BMC Med* 2021;**19**(1): 287.
- 508 49. Zammit AR, Piccinin AM, Duggan EC, Koval A, Clouston S, Robitaille A, et al. A Coordinated Multi-
509 study Analysis of the Longitudinal Association between Handgrip Strength and Cognitive Function
510 in Older Adults. *J Gerontol A Biol Sci Med Sci* 2021;**76**(2): 229-41.
- 511 50. Li C, Zhu Y, Ma Y, Hua R, Zhong B, Xie W. Association of Blood Pressure With Cognitive Decline,
512 Dementia, and Mortality. *J Am Coll Cardiol* 2022;**79**(14):1321–35.
- 513 51. Sutin AR, Luchetti M, Stephan Y, Strickhouser JE, Terracciano A. The association between
514 purpose/meaning in life and verbal fluency and episodic memory: A meta-analysis of >140,000
515 participants from up to 32 countries. *Int Psychogeriatr* 2022; **34**(3): 263-73.
- 516 52. Zhu Y, Li C, Xie W, Zhong B, Wu Y, Blumenthal JA. Trajectories of depressive symptoms and
517 subsequent cognitive decline in older adults: A pooled analysis of two longitudinal cohorts. *Age*
518 *Ageing* 2022;**51**(1).
- 519 53. Zhu Y, Li C, Wu T, Wang Y, Hua R, Ma Y, et al. Associations of cumulative depressive symptoms
520 with subsequent cognitive decline and adverse health events: Two prospective cohort studies. *J*
521 *Affect Disord* 2023;**320**:91–7.

- 522 54. Zhan Y, Clements MS, Roberts RO, Vassilaki M, Druliner BR, Boardman LA, et al. Association of
523 telomere length with general cognitive trajectories: a meta-analysis of four prospective cohort
524 studies. *Neurobiol Aging* 2018;**69**: 111-6.
- 525 55. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention,
526 intervention, and care: 2020 report of the Lancet Commission. *The Lancet* 2020;**396**.
- 527 56. Rose G. Sick individuals and sick populations. *Int J Epidemiol.* 2001;**30**(3).
- 528 57. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: A
529 systematic review and metaanalysis. *Alzheimer's and Dementia* 2013;**9**.
- 530 58. Gross AL, Crane PK, Gibbons LE, Manly JJ, Romero H, Thomas M, et al. Effects of education and
531 race on cognitive decline: An integrative study of generalizability versus study-specific results.
532 *Psychol Aging* 2015;**30**(4).
- 533 59. Taasobshirazi G, Wang S. THE PERFORMANCE OF THE SRMR, RMSEA, CFI, AND TLI: AN
534 EXAMINATION OF SAMPLE SIZE, PATH SIZE, AND DEGREES OF FREEDOM. *Journal of Applied*
535 *Quantitative Method* 2016;**11**(3).
- 536 60. Brown T.A., Confirmatory factor analysis for applied research. *Guilford publications* 2015;**5**.
- 537 61. Langa, K. M., Ryan, L. H., McCammon, R. J., Jones, R. N., Manly, J. J., Levine, D. A., Sonnega, A.,
538 Farron, M., & Weir, D. R. (2020). The Health and Retirement Study Harmonized Cognitive
539 Assessment Protocol Project: Study Design and Methods. *Neuroepidemiology*, 54(1), 64-74.
540 <https://doi.org/10.1159/000503004>

Table 1. Sample characteristics of included HCAP studies (N=21,141)

Characteristics	Overall sample	HRS-HCAP	ELSA-HCAP	LASI-DAD	Mex-Cog	CHARLS - HCAP	HAALSI-HCAP
Sample size, n	21141	3347	1273	4096	2042	9755	628
Age, years, mean (SD)	72.7 (8.9)	76.6 (7.5)	75.8 (7.0)	69.7 (7.6)	68.1 (9.0)	68.5 (6.5)	69.3 (11.5)
Female gender, n (%)	16404 (55.7)	2020 (60.4)	700 (55.0)	2207 (53.9)	1203 (58.9)	4960 (50.8)	387 (61.6)
Education, n (%)							
No or Early Childhood Education	8862 (42.0)	22 (0.7)	3 (0.2)	2558 (62.5)	1023 (50.5)	4909 (50.3)	347 (55.3)
Primary education (US grades 1-6)	3674 (17.4)	131 (3.9)	0 (0.0)	527 (12.9)	452 (22.3)	2355 (24.1)	209 (33.3)
Lower secondary (US grades 7-9)	3160 (15.0)	454 (13.6)	486 (39.3)	314 (7.7)	317 (15.7)	1562 (16.0)	27 (4.3)
Upper secondary (US grades 10-12)	3465 (16.4)	1773 (53.0)	303 (24.5)	505 (12.3)	60 (3.0)	792 (8.1)	32 (5.1)
Any college	1924 (9.1)	965 (28.8)	446 (36.0)	192 (4.7)	172 (8.5)	137 (1.4)	12 (1.9)

Table 2. Model fit statistics of CFAs for each cognitive domain in each study: Results from HCAP studies (N=21,141)

Cognitive domain	Study	Number of items	RMSEA	CFI	SRMR	Bifactor structure	Summary of fit
General cognition	HRS HCAP	45	0.035	0.925	0.078	Yes	Adequate
General cognition	ELSA HCAP	41	0.027	0.968	0.089	Yes	Poor
General cognition	LASI-DAD - literate	48	0.033	0.904	0.066	Yes	Adequate
General cognition	LASI-DAD - illiterate	48	0.036	0.904	0.063	Yes	Adequate
General cognition	Mex-Cog	40	0.040	0.932	0.072	Yes	Adequate
General cognition	CHARLS HCAP	31	0.032	0.949	0.051	No	Adequate
General cognition	HAALSI HCAP	51	0.043	0.913	0.122	Yes	Poor
Memory	HRS HCAP	11	0.045	0.980	0.023	Yes	Good
Memory	ELSA HCAP	11	0.060	0.971	0.038	Yes	Adequate
Memory	LASI-DAD - literate	11	0.046	0.978	0.027	Yes	Good
Memory	LASI-DAD - illiterate	11	0.049	0.965	0.031	Yes	Good
Memory	Mex-Cog	10	0.048	0.985	0.033	Yes	Good
Memory	CHARLS HCAP	5	0.047	0.984	0.020	No	Good
Memory	HAALSI HCAP	9	0.026	0.995	0.018	Yes	Good
Orientation	HRS HCAP	10	0.028	0.971	0.064	No	Adequate
Orientation	ELSA HCAP	9	0.010	0.999	0.052	No	Adequate
Orientation	LASI-DAD - literate	10	0.053	0.924	0.077	Yes	Adequate
Orientation	LASI-DAD - illiterate	10	0.049	0.945	0.064	Yes	Adequate
Orientation	Mex-Cog	8	0.062	0.924	0.066	No	Adequate
Orientation	CHARLS HCAP	10	0.043	0.968	0.051	No	Adequate
Orientation	HAALSI HCAP	10	0.032	0.989	0.069	No	Adequate
Language	HRS HCAP	14	0.020	0.971	0.071	No	Adequate
Language	ELSA HCAP	12	0.007	0.997	0.070	Yes	Adequate
Language	LASI-DAD - literate	14	0.034	0.897	0.070	Yes	Adequate
Language	LASI-DAD - illiterate	14	0.032	0.949	0.050	Yes	Adequate
Language	Mex-Cog	13	0.016	0.986	0.073	Yes	Adequate
Language	CHARLS HCAP	13	0.029	0.960	0.046	No	Good
Language	HAALSI HCAP	16	0.039	0.971	0.127	Yes	Poor
Executive functioning	HRS HCAP	8	0.076	0.973	0.020	Yes	Adequate

Executive functioning	ELSA HCAP	8	0.080	0.966	0.020	Yes	Adequate
Executive functioning	LASI-DAD - literate	10	0.029	0.989	0.024	No	Good
Executive functioning	LASI-DAD - illiterate	10	0.038	0.975	0.034	No	Good
Executive functioning	Mex-Cog	7	0.043	0.995	0.018	Yes	Good
Executive functioning	CHARLS HCAP	2	0.000	1.000	0.000	No	Perfect
Executive functioning	HAALSI HCAP	12	0.076	0.922	0.059	Yes	Adequate

Legend. CFI: confirmatory fit index; RMSEA: root mean square error of approximation; SRMR: standardized root mean squared residual.

Model fit was considered perfect if $CFI = 1$ and $RMSEA = 0$ and $SRMR = 0$, good if $CFI \geq 0.95$ and $RMSEA \leq 0.05$ and $SRMR \leq 0.05$, adequate if $CFI \geq 0.90$ and $RMSEA \leq 0.08$ and $SRMR \leq 0.08$, and poor if either $CFI < 0.9$ or $RMSEA > 0.08$ or $SRMR > 0.08$.

We chose this combination because each fit statistic has advantages and disadvantages. Together, these three statistics considered in conjunction minimize the risk of choosing a bad model. Although low SRMR implies low model residuals, it does not incorporate model complexity and may be partial to overly complex models or models with larger sample sizes. The RMSEA provides an index of model discrepancy per degree of freedom (which accounts for model complexity), but tends to improve with larger sample size. The CFI compares an estimated model with a hypothetical null baseline model.

Table 3. Number of participants in each study and for each domain whose scores show salient DIF: Results from HCAP studies (N=21,141)

Domain	HRS HCAP	ELSA HCAP	LASI-DAD	Mex-Cog	CHARLS HCAP	HAALSI
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Memory	Reference	2 (0.2%)	1 (0.1%)	105 (5.1%)	216 (2.2%)	0 (0%)
Orientation	Reference	22 (1.7%)	0 (0%)	0 (0%)	23 (0.3%)	0 (0%)
Language/fluency	Reference	57 (4.5%)	23 (1.3%)	50 (2.5%)	1637 (16.8%)	6 (1.0%)
Executive function	Reference	0 (0%)	0 (0%)	No overlap	No overlap	0 (0%)

Legend. Impact of differential item functioning (DIF) was calculated as the difference between DIF-adjusted and non-DIF-adjusted factor scores. This table shows the number of participants in each study and for each domain whose DIF-adjusted scores differed by more than 0.3 standard deviations from non-DIF-adjusted scores.

Table 4. Validation of the general cognitive function factor: Results from HCAP studies (N=21,141)

Covariate	Overall sample	HRS-HCAP	ELSA-HCAP	LASI-DAD	Mex-Cog	CHARLS-HCAP	HAALSI-HCAP
	Beta coefficient (SE)						
Female gender	0.017 (0.011)	0.19 (0.03)	0.06 (0.05)	-0.06 (0.03)	-0.10 (0.03)	-0.07 (0.02)	-0.22 (0.04)
Age group							
50-59 years	.40 (.043)	N/A	N/A	N/A	0.28 (0.05)	N/A	0.46 (0.07)
60-69 years	REF						
70-79 years	.14 (.014)	-0.30 (0.04)	-0.57 (0.07)	-0.31 (0.03)	-0.54 (0.05)	-0.25 (0.02)	-0.38 (0.07)
80-89 years	-.057 (.017)	-0.89 (0.04)	-1.23 (0.07)	-0.66 (0.04)	-1.17 (0.06)	-0.74 (0.04)	-0.72 (0.07)
90+ years	-.34 (.032)	-1.45 (0.07)	-1.80 (0.14)	-1.18 (0.09)	-1.61 (0.14)	-0.99 (0.17)	-0.95 (0.14)
Education							
No or Early Childhood education	-1.16 (0.017)	-0.55 (0.20)	-1.01 (0.58)	-0.96 (0.04)	-1.26 (0.05)	-1.13 (0.02)	-0.96 (0.12)
Primary education (US grades 1-6)	-0.36 (0.019)	-0.50 (0.09)	N/A	-0.26 (0.05)	-0.42 (0.05)	-0.35 (0.03)	-0.32 (0.12)
Lower secondary (US grades 7-9)	REF						
Upper secondary (US grades 10-12)	0.40 (0.020)	0.72 (0.05)	0.52 (0.07)	0.34 (0.05)	0.17 (0.10)	0.26 (0.03)	0.47 (0.16)
Any college	0.84 (0.023)	1.18 (0.05)	0.76 (0.07)	0.55 (0.06)	0.57 (0.07)	0.50 (0.07)	0.78 (0.21)

Legend. Beta coefficients represent overall and study-specific differences in general cognitive functioning between a given exposure grouping and the reference category. For age, persons aged 60-69 comprised the reference group. For education, persons with a lower secondary education comprised the reference group.

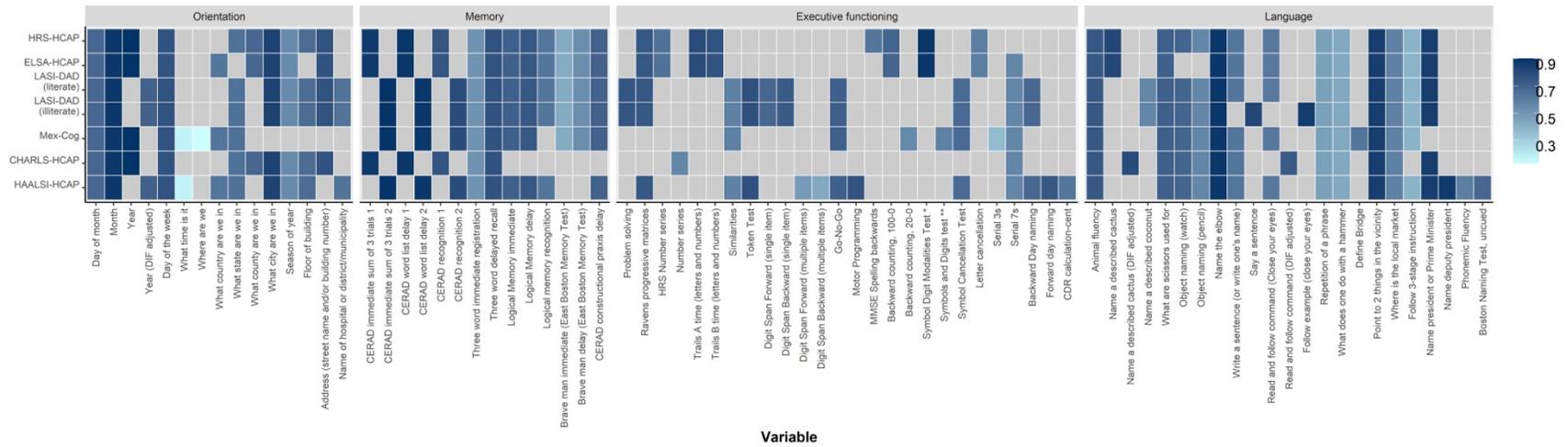


Figure 1. Heatmap of cognitive test items and their overlap across each study: Results from HCAP studies (N=21,141)

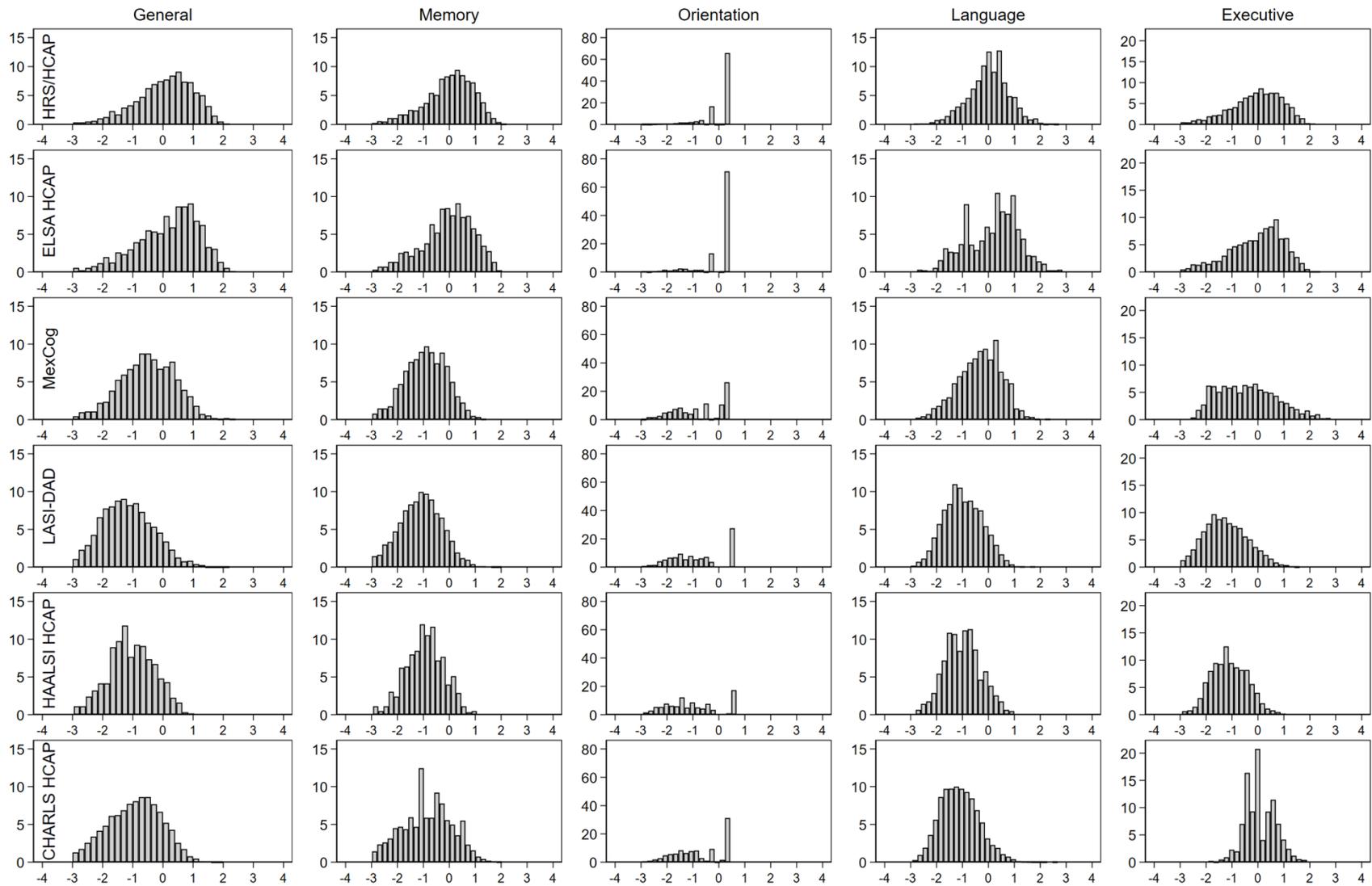


Figure 2. Distributions of harmonized general and domain-specific cognitive factor scores: Results from HCAP studies (N=21,141)

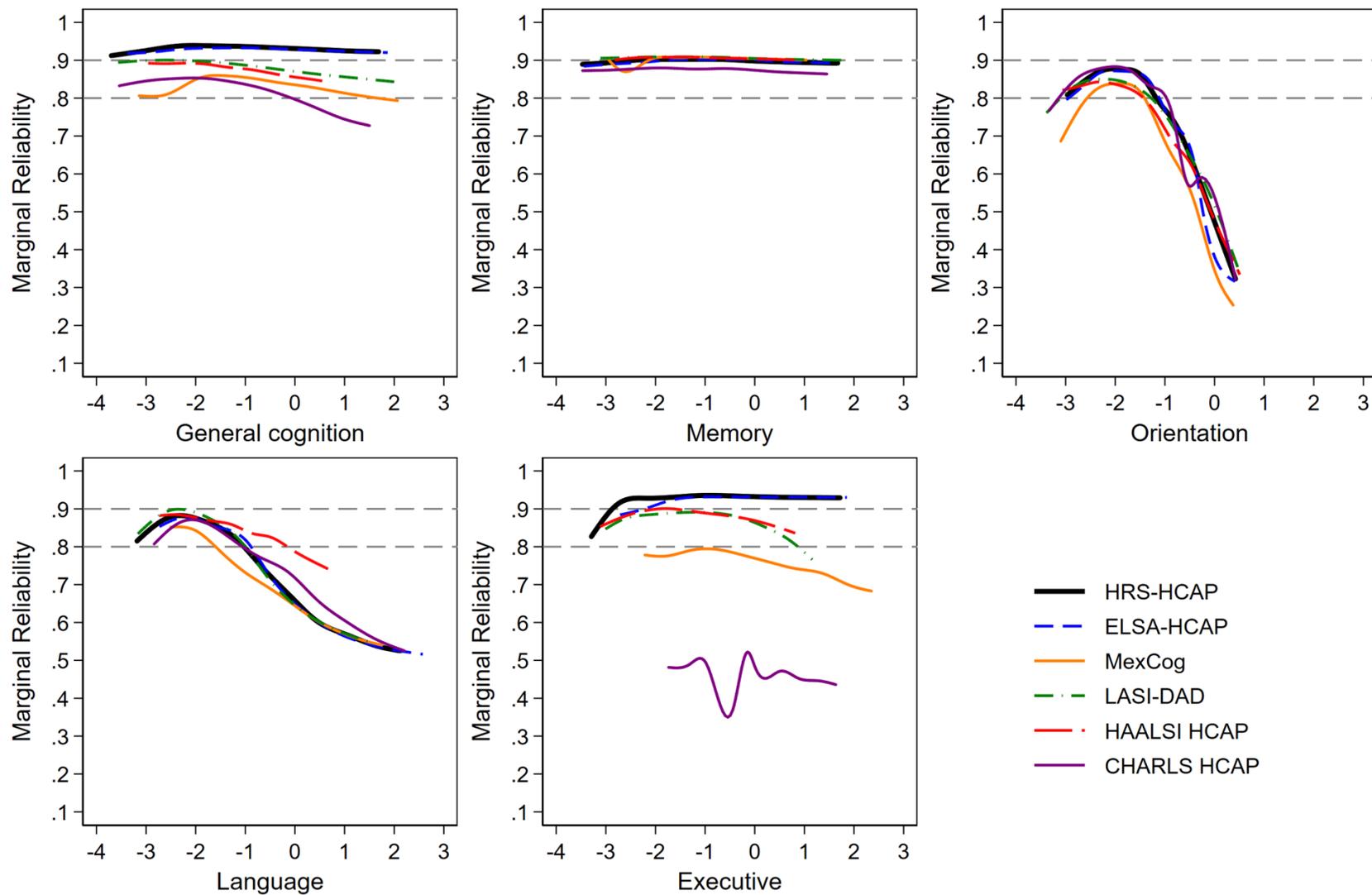


Figure 3. Plots of marginal reliability by study for overall and domain-specific cognitive performance: Results from HCAP studies (N=21,141)

