Harmonisation of later-life cognitive function across national contexts: results from the Harmonized Cognitive Assessment Protocols

Alden L Gross,

Chihua Li,

Emily M Briceño,

Miguel Arce Rentería,

Richard N Jones,

Kenneth M Langa,

Jennifer J Manly,

Emma Nichols,

David Weir,

Rebeca Wong,

Lisa Berkman,

Jinkook Lee[*],

Lindsay C Kobayashi[*]

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (A L Gross PhD, E Nichols PhD); Center on Aging and Health, Johns Hopkins University, Baltimore, MD, USA (A L Gross, E Nichols); Center for Social Epidemiology and Population Health, Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA (C Li PhD, L C Kobayashi PhD); Survey Research Center, University of Michigan Institute for Social Research, Ann Arbor, MI, USA (C Li, Prof K M Langa MD, Prof D Weir

Correspondence to: Assoc Prof Alden L Gross, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA, agross14@jhu.edu.

See **Online** for appendix

PhD, L C Kobayashi); Department of Physical Medicine and Rehabilitation (E M Briceño PhD) and Department of Internal Medicine (Prof K M Langa), University of Michigan Medical School, Ann Arbor, MI, USA; Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University College of Physicians and Surgeons, New York, NY, USA (M Arce Rentería PhD, Prof J J Manly PhD); Department of Psychiatry and Human Behavior, Warren Alpert Medical School, Brown University, Providence, RI, USA (Prof R N Jones ScD); Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI, USA (Prof K M Langa); Center for Economic and Social Research (E Nichols, Prof J Lee PhD) and Department of Economics (Prof J Lee), University of Southern California, Los Angeles, CA, USA; School of Public and Population Health, and Sealy Center on Aging, University of Texas Medical Branch, Galveston, TX, USA (Prof R Wong PhD); Harvard Center for Population and Development Studies and Department of Social and Behavioral Sciences, Harvard T H Chan School of Public Health, Cambridge, MA, USA (Prof L Berkman PhD)

Summary

Background—The Harmonized Cognitive Assessment Protocol (HCAP) is an innovative instrument for cross-national comparisons of later-life cognitive function, yet its suitability across diverse populations is unknown. We aimed to harmonise general and domain-specific cognitive scores from HCAP studies across six countries, and evaluate reliability and criterion validity of the resulting harmonised scores.

Methods—We statistically harmonised general and domain-specific cognitive function scores across publicly available HCAP partner studies in China, England, India, Mexico, South Africa, and the USA conducted between October, 2015 and January, 2020. Participants missing all cognitive test items in a given HCAP were excluded. We used an item banking approach that leveraged common cognitive test items across studies and tests that were unique to studies. We generated harmonised factor scores to represent a person's relative functioning on the latent factors of general cognitive function, memory, executive function, orientation, and language using confirmatory factor analysis. We evaluated the marginal reliability, or precision, of the factor scores using test information plots. Criterion validity of factor scores was assessed by regressing the scores on age, gender, and educational attainment in a multivariable analysis adjusted for these characteristics.

Findings—We included 21 144 participants from the six HCAP studies of interest (11 480 women [54·3%] and 9664 [45·7%] men), with a median age of 69 years (IQR 64–76). Confirmatory factor analysis models of cognitive function in each country fit well: 31 (88·6%) of 35 models had adequate or good fit to the data (comparative fit index ≥ 0·90, root mean square error of approximation ≤ 0·08, and standardised root mean residual ≤ 0·08). Marginal reliability of the harmonised general cognitive function factor was high (>0·9) for 19 044 (90·1%) of 21 144 participant scores across the six countries. Marginal reliability of the harmonised factor was above 0·85 for 19 281 (91·2%) of 21 142 participant factor scores for memory, 7805 (41·0%) of 19 015 scores for executive function, 3446 (16·3%) of 21 103 scores for orientation, and 4329 (20·5%) of 21 113 scores for language. In each country, general cognitive function scores were lower with older age and higher with greater levels of educational attainment.

Interpretation—We statistically harmonised cognitive function measures across six large population-based studies of cognitive ageing. These harmonised cognitive function scores empirically reflect comparable domains of cognitive function among older adults across the six countries, have high reliability, and are useful for population-based research. This work provides a foundation for international networks of researchers to make improved inferences and direct comparisons of cross-national associations of risk factors for cognitive outcomes in pooled analyses.

Introduction

Dementia, which has cognitive decline as its hallmark symptom, poses major public health, clinical, and policy challenges. Although much research on risk and protective factors for dementia has been done in high-income countries, it is estimated that by 2050, three-quarters of the 152 million people with dementia will be living in low-income and middle-income countries.[1–3] Differences in the distribution of potential risk factors (and of cultural and demographic factors that affect the risk of dementia) across countries make cross-national research imperative.

To facilitate cross-national comparisons of later-life cognitive outcomes, assessment instruments need to provide valid measures of cognitive function across populations with diverse cultural, educational, social, economic, and political contexts. The Harmonized Cognitive Assessment Protocol (HCAP) has been developed and implemented in international partner studies of the US Health and Retirement Study (HRS).[4] To date, the HCAP network represents the largest concerted global effort to conduct harmonised, large-scale, population-representative studies of cognitive ageing and dementia. The HCAP aims to provide a detailed assessment of cognitive function of older adults that is flexible, yet comparable across populations in countries with diverse cultural, educational, social, economic, and political contexts. The HCAP network ultimately intends to provide comparable estimates of dementia and mild cognitive impairment prevalence across countries, and to use cross-national variation in key risk and protective factors to better understand determinants of cognitive ageing and dementia.[4]

Although the HCAP was designed collaboratively to ensure its comparability across countries, necessary adaptations were made to individual test items, test administration, and scoring procedures to accommodate different languages, cultures, and levels of literacy and numeracy in respondents.[5] The effects of these adaptations on the performance, reliability, and validity of the HCAP cognitive test items are only beginning to be understood,[5,6] and might limit cross-national comparisons. We aimed to statistically harmonise the HCAP instruments fielded in China, England, India, Mexico, South Africa, and the USA. To achieve this harmonisation, we assigned cognitive test items to domains, identified which test items were common and which were unique across countries, derived harmonised factor scores for general and specific cognitive domains, and estimated the reliability and validity of the harmonised factor scores.

Methods

Study design and participants

The HRS in the USA and its international partner studies are large, population-based studies of ageing. Between October, 2015 and January, 2020, six such studies administered HCAPs to participants: the HRS, the English Longitudinal Study on Ageing (ELSA), the Longitudinal Aging Study in India (LASI), the Mexican Health and Aging Study (MHAS), the China Health and Retirement Longitudinal Study (CHARLS), and Health and Aging in Africa: A Longitudinal Study in South Africa (HAALSI). Details on HCAP administration in each cohort, participant eligibility, dates of recruitment, and sample sizes are provided in the appendix (pp 14–15).[4,7–12] Hereafter, we refer to the HCAP studies from the parent cohort studies (HRS-HCAP, ELSA-HCAP, LASI-Diagnostic Assessment of Dementia [LASI-DAD], MHAS Cognitive Aging Ancillary Study [Mex-Cog], CHARLS-HCAP, and HAALSI-HCAP).

The HCAP studies in England, Mexico, South Africa, and the USA randomly sampled participants from the core studies who did not have a proxy interview in the core study interview. The HRS-HCAP further included a random sample of 219 participants interviewed by proxy in the 2016 HRS core interview.[4] CHARLS-HCAP selected all participants aged 60 years or older from the CHARLS study. The LASI-DAD study employed a stratified, random sampling design, recruiting participants from the main LASI study. To ensure adequate sample sizes of participants with dementia, the HCAP studies in England, India, and South Africa oversampled participants with low cognitive function.[7,10,11] The minimum age for inclusion was 50 years in HAALSI-HCAP, 55 years in Mex-Cog, 60 years in LASI-DAD and CHARLS-HCAP, and 65 years in HRS-HCAP and ELSA-HCAP. All parent studies were nationally representative, with the exception of HAALSI, which is a sample from the Agincourt subdistrict in northeastern South Africa.[12] Written informed consent was obtained for participants in all studies and institutional review boards approved each international partner study and its respective HCAP study.

Variables

Details of the original battery of 17 cognitive tests in the HCAP have been described by Langa and colleagues.[4] By design, each HCAP study administered as close to the same battery of tests as feasible. We granularised these batteries to the cognitive test indicators that were administered in each HCAP study. Between 30 and 48 indicators were administered in each study, with 78 different indicators used in total (figure 1, appendix pp 16–18).

The HCAP battery was designed to assess key cognitive domains affected by cognitive ageing, including memory, executive function, orientation, and language.[4] Each cognitive test item was assigned to a domain on the basis of well accepted neuropsychological and psychological theories[13,14] and empirical analyses showing which test items fit well into a domain.[6,15–17] Assigning test items to domains is essential to statistical harmonisation (also referred to as co-calibration) and this process relies on the presence of equivalent or comparable cognitive test items across one or more studies. If cognitive test items are

presumed to be the same across HCAP studies but are in fact different (eg, a different test, or the same test with different stimuli, administration, or scoring procedures), the resulting methodological differences could contribute to artifactual differences in observed cognitive scores across studies, which would reduce the quality of cross-national inferences from derived summary cognitive scores.

To establish the comparability of cognitive test items across HCAPs, we convened a multidisciplinary panel of individuals (including authors ALG, MAR, EMB, and LCK) with cultural or linguistic expertise, including neuropsychologists, epidemiologists, and psychometricians with working knowledge of cross-cultural neuropsychology and administration of HCAPs to conduct prestatistical harmonisation of cognitive test items (appendix pp 2–7). Using the HRS-HCAP as the reference and based on available information, two neuropsychologists (EMB and MAR) rated each item from other HCAPs as a confident linking item (ie, no known issues violating item equivalence), a tentative linking item (ie, possible issues with item equivalence), or a non-linking item (ie, unique or novel items with known issues violating item equivalence; appendix pp 6–7).[5] All indicators (both linking and non-linking) were included in the final models to estimate general and domain-specific cognitive function.

Age, gender, and highest educational attainment covariates were collected in core international partner study interviews. Each study recorded gender as male or female; for the purposes of this study, we use the term gender because the variable as used here reflects participant identity, not necessarily biological sex. We scaled educational attainment in each country according to the 2011 International Standard Classification of Education (appendix p 19).[18] Data on race and ethnicity are not reported because ethnic minority groups in each country are categorised in country-specific ways, making cross-country comparisons unfeasible.

Statistical analysis

We described demographic characteristics and cognitive tests using means with standard deviations, medians with IQR, and counts with percentages. We identified overlapping and unique cognitive test items across the HCAP studies.

To illustrate empirically that similar organisations of cognitive test items fit well across countries before imposing assumptions about cross-national linking items between studies, we estimated confirmatory factor analysis models, consistent with graded-response and continuous-response item response theory models, for cognitive domains of general and domain-specific cognitive function separately for each HCAP study (appendix pp 8–10).[16,19,20] We assessed model fit for the confirmatory factor analysis models using the indices of root mean square error of approximation, comparative fit index, and standardised root mean residual (appendix p 8).

We statistically harmonised scores for each cognitive domain across the HCAP studies using an item banking procedure.[6] A flowchart of this approach is presented in the appendix (p 34). For each cognitive domain, we serially estimated confirmatory factor analyses in each study, sequentially fixing model parameters for linking items to their corresponding

values from other studies. The confirmatory factor analysis models estimated two relevant parameters for each cognitive test item: factor loadings, and item thresholds (for categorical items) or intercepts (for continuous items). The interpretation of these parameters is described in the appendix (p 8).

To begin the item banking procedure, model parameters from a confirmatory factor analysis model in HRS-HCAP for a given cognitive domain were saved for use in confirmatory factor analyses in subsequent HCAPs (appendix p 34). For example, after estimating a confirmatory factor analysis for the HRS-HCAP study, we estimated a confirmatory factor analysis in ELSA-HCAP for the given domain, in which item parameters for cognitive test items in common with the HRS-HCAP were fixed to those observed in the HRS-HCAP, and the mean and variance of the underlying trait were freely estimated. The process was repeated for each HCAP study. The order of studies was HRS-HCAP, ELSA-HCAP, LASI-DAD, Mex-Cog, CHARLS-HCAP, and HAALSI-HCAP (appendix pp 11–13). LASI-DAD was divided into literate (n=1777 [43·4%]) and illiterate (n=2319 [56·6%]) subgroups due to administration differences in some tests (eg, some tests originally intended to be delivered in writing were instead given as verbal instruction or cues for individuals who were illiterate; appendix p 5). During the item banking procedure, parameters for cognitive test items from a given HCAP study that were not yet in the item bank were freely estimated, then saved in the item bank for use when a subsequent study with the same item was included. For the purposes of estimating cognitive factor scores, as the quantitative summaries of an observation's relative location on a latent trait (ie, cognition), in a final confirmatory factor analysis model for each cognitive domain, we estimated a confirmatory factor analysis in the pooled sample of all HCAP studies, in which all parameters were fixed to previous values stored in the item bank. Because no natural scaling in latent variable space exists, the mean and variance of latent traits (ie, general cognitive function, memory, executive function, orientation, and language) were set to 0 and 1, respectively, beginning with HRS-HCAP as the reference. In addition to factor scores, using test information plots we evaluated marginal reliability (also known as internal consistency reliability) of the factor scores from confirmatory factor analysis measurement models for each cognitive domain in each HCAP study, with marginal reliability expressed as a value between 0 and 1 and calculated from one minus the squared standard error of the measurement of each observation.

The validity of cross-national harmonisation of cognitive function depends on the availability of common, equivalent cognitive test items across studies. Although our expert panel identified equivalent linking items, test differences that might not have been documented, that were due to unforeseen cultural differences, or for which insufficient documentation was available might have been missed. Thus, we statistically tested for differential item functioning (DIF) among linking items presumed to be equivalent between the HRS-HCAP and each study, by cognitive domain. We used multiple indicator, multiple cause (MIMIC) models to evaluate DIF for linking items in each HCAP study. We evaluated DIF in LASI-DAD using the subsample of literate respondents due to the availability of more linking items than in the subsample of illiterate respondents. In addition to underlying cognitive function for a given domain, MIMIC models were adjusted for age and gender.[5,21] In the MIMIC models, we used item ratings (ie, confident or tentative linking) made during the course of prestatistical harmonisation to avoid subjectivity in choice of anchor items.

We first tested DIF using MIMIC models among cognitive test items rated as confident linking items, and then tested for DIF among cognitive items rated as tentative linking items, treating as anchors in MIMIC models the confident items that showed no DIF in the previous analysis. The magnitude of DIF attributable to a given cognitive test item was represented by an odds ratio (OR) for the regression of an item on a variable for study membership; we considered DIF to be of a large magnitude (non-negligible) when the OR was outside the range of 0·66–1·50.[22] A large effect of DIF on a participant's domain-level scores, or salient DIF, was evaluated by taking the difference between DIF-adjusted and non-DIF-adjusted scores. DIF-adjusted scores allowed items shown to have non-negligible DIF to have different measurement model parameters across studies. We counted how many participant scores in each HCAP study differed by more than 0·3 SD units.[5,23] Both confident and tentative linking items that empirically showed no DIF and scenarios in which the detected DIF was not salient were considered as linking items in the final statistical harmonisation for general and domain-specific factors. As a general rule to reduce the influence of DIF, we aimed to have 10% or less observations with salient DIF.

To assess validity, we evaluated patterns of factor scores by age, gender, and educational attainment by regressing general and domain-specific cognitive function factor scores on each of these characteristics, adjusting for the other characteristics in a multivariable regression. We hypothesised that cognitive function would generally be better at younger ages and at higher levels of educational attainment.[24,25] With respect to gender, we hypothesised that women would have poorer cognitive function in low-income and middle-income settings compared with men, given established gender-based societal inequalities in these settings that apply to determinants of later-life cognitive health such as educational, occupational, and other life opportunities.[26,27]

Before modelling, we recoded responses of "don't know" as incorrect.[19] Other item-level missingness in cognitive test items (ie, missingness due to refusals and other reasons) did not preclude inclusion into this study because confirmatory factor analyses were estimated using the full information maximum likelihood estimator, which assumes data are missing conditional on other variables in the model (eg, cognitive tests). Participants missing all cognitive test items in an HCAP were excluded from this study.

Descriptive analyses were conducted with Stata (version 17). Factor analyses were conducted with Mplus (version 8.7).[28] In analyses related to criterion validation, we report statistical significance at an α level of 0·05 but no interpretations in this manuscript rely solely on a threshold for p values.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

We included 21 144 participants from the six HCAP studies (table 1). This sample comprised 11 480 women (54·3%) and 9664 (45·7%) men, and median age was 69 years (IQR 64–76; range 50–108).

Cognitive test items were organised by the cognitive domain to which tests were assigned (figure 1, appendix pp 16–18). Of 78 items, 15 (19·2%) were assigned to the orientation domain, 14 (17·9%) to the memory domain, 26 (33·3%) to executive function domain, and 23 (29·5%) to the language domain (figure 1). Most factor loadings were between 0·3–0·9, suggesting adequate intercorrelations among items, given that this range generally indicates meaningful relation to other items.[15,29] Item thresholds, which describe the location of each item relative to others, ranged from –4·5 to 4·5 SDs along the latent traits, and are provided in the appendix (pp 20–22). For a given test item (figure 1, appendix pp 16–18), the presence of equal factor loadings across rows reflects decisions regarding whether items were linking items or non-linking items, made during prestatistical harmonisation (with equal factor loadings indicating linking items). Of the 78 cognitive test items administered, 13 (16·7%) were comparably administered in every HCAP study, comprising three items for orientation, two for memory, and eight for language (including items for either the literate or illiterate subgroup, or both subgroups, of LASI-DAD). No items in the executive function domain were comparably administered across all the HCAP studies.

Confirmatory factor analysis models of cognitive function in each country fit well: 31 (88·6%) of 35 models had adequate or good fit to the data (comparative fit index 0·90, root mean square error of approximation 0·08, and standardised root mean residual 0·08; appendix pp 9–10). After statistical harmonisation via item banking, the factor scores for general cognitive function, memory, language, and executive function were normally distributed in each HCAP study (figure 2). By contrast, the orientation factor showed a strong ceiling effect in each study (figure 2). These ceiling effects were explained by low marginal reliabilities (based on the standard error of the measurement model) of factor scores for orientation. The orientation factor provided low reliability above scores of 0, as 2156 (64·6%) of 3340 participants in HRS-HCAP and many participants in the other studies (4984 [28·1%] of 17 763) had maximum scores because they answered all orientation items correctly (figure 3). By comparison, reliabilities of the general cognitive function and memory factors were uniformly high for each HCAP study. For general cognitive function, marginal reliability was greater than 0·9 for 19 044 (90·1%) of 21 144 participant scores. Marginal reliability of the harmonised factor was above 0·85 for 19 281 (91·2%) of 21 142 participant factors scores for memory, 7805 (41·0%) of 19 015 scores for executive function, 3446 (16·3%) of 21 103 scores for orientation, and 4329 (20·5%) of 21 113 factor scores for language. The language factor showed greater reliability at lower levels of language ability than at higher levels, reflecting that most language items, with the exception of animal fluency, tended to be easier. For executive function, reliability was high for all studies except CHARLS-HCAP, which had just two test items measuring executive function.

Regarding DIF testing, 21 (26·9%) of the 78 cognitive test items showed a large magnitude of DIF (corresponding OR outside the range 0·66–1·50) between HRS-HCAP and another

study (appendix pp 23–31). Of these 21, ten (47·6%) assessed language. With respect to the effect of this observed DIF on participant factor scores, we identified most of the salient DIF (ie, score estimates differing by 0·3 units or more before *vs* after accounting for DIF) in three domain-by-country cases. Specifically, 290 (16·3%) of 1777 LASI-DAD orientation factor scores among literate participants, 326 (51·7%) of 631 HAALSI-HCAP orientation factor scores, and 6668 (68·4%) of 9755 CHARLS-HCAP language factor scores showed salient DIF. Subsequent analyses, removing each item as a linking item successively, showed that orientation to year was entirely responsible for detected salient DIF in the orientation domain for both LASI-DAD and HAALSI-HCAP (controlling for underlying orientation ability, performance on this item was lower in these studies). Most of the salient DIF in the CHARLS-HCAP language factor scores could be attributed to differences in two items after controlling for underlying language ability: naming a described cactus (cacti are unfamiliar in China) and following a read command (due to illiteracy). After removing these items as linking items between CHARLS-HCAP and other studies and retaining them as non-linking items, 1637 (16·8%) of 9755 participant scores were affected by adjustment for DIF in the language domain for CHARLS-HCAP (table 2). Thus, although the CHARLS-HCAP survey was conducted in Mandarin, DIF in the language cognitive domain compared with HRS-HCAP was affected more by familiarity with a cactus and by literacy than by language differences between Mandarin and English. Between 0% and 5·1% of scores for any domain in any other study had salient DIF (table 2). The heatmap of overlapping items (figure 1) reflects decisions from DIF analyses to relax assumptions of item equivalence for the three aforementioned items in LASI-DAD, HAALSI-HCAP, and CHARLS-HCAP.

Patterns of cognitive function aligned with hypothesised expectations: general cognitive function mean scores were generally lower at older ages and higher with greater educational attainment (table 3). Women had higher mean general cognitive function scores than men in the HRS-HCAP, but lower mean scores than men in LASIDAD, Mex-Cog, CHARLS-HCAP, and HAALSI-HCAP, consistent with hypothesised between-country differences in gender inequalities. Findings with respect to age and education were similar for specific domains (appendix pp 32–33).

Discussion

We investigated the performance of harmonised cognitive factors based on common and unique cognitive test items administered to 21 144 older adults across six harmonised studies of ageing conducted in China, England, India, Mexico, South Africa, and the USA. We used psychometric methods to ensure these factors reflected comparable domains of cognitive function and were reliable and valid measures. Differences in test administration due to language, literacy, and numeracy were addressed to statistically harmonise general and domain-specific cognitive function factors across the six cohorts. As such, these factors should be useful for population-based cross-national research.

High-quality, harmonised scores for general and domain-specific cognitive function are crucial tools to promote valid cross-national comparisons of age-related cognitive changes in a rapidly ageing population worldwide.[19] The present study describes the derivation and validation of cognitive scores to facilitate pooled cross-country investigations. One of our

aims was to make the use of these factor scores accessible to individuals without specific expertise in psychometrics. In the past decade, an increasing number of cross-national studies have examined risk factors for cognitive function decline and dementia, many using data from the HRS and its international partner studies.[26,31–35] A recent *Lancet* Commission identified 12 major risk factors for dementia: low education, hearing impairment, traumatic brain injury, hypertension, diabetes, excessive alcohol use, obesity, smoking, depression, social isolation, physical inactivity, and air pollution.[36] The harmonised cognitive function scores generated in this study can be used in pooled analyses to evaluate whether these risk factors have similar effects on cognitive function across global settings. Such knowledge could facilitate identification of contextual risk-modifying factors that could be intervened upon to reduce the risk of dementia in specific populations.[37] Furthermore, common cognitive phenotypes could be used to improve the quality of population-attributable risk estimates and to generate prevalence and incidence estimates in dementia algorithms that are truly comparable across national settings. Importantly, theoretical and methodological considerations must be made regarding how future cross-national research will be conducted to leverage these cognitive factors. One methodological issue is the use of sampling weights—most HCAP surveys provide complex sampling weights to facilitate nationally representative inferences. We did not incorporate sampling weights in the present study because our aim was to develop cognitive scores for people in the sample and not the target population of interest; however, future research that uses these scores in select samples or populations might consider the use of sampling weights to make population-representative inferences.

Prestatistical harmonisation and statistical testing for DIF were two essential steps for the harmonisation process in this study. DIF can be introduced by methodological differences in test administration or scoring across studies, in addition to population-level differences that might alter responses to equivalent test items (eg, differences in literacy, numeracy, or language). We evaluated the comparability of cognitive test items using a multidisciplinary team, which was a crucial component of this harmonisation work. However, the ability to identify measurement differences in cognitive test items across languages and cultures depends on the quality of available study documentation, adequate documentation, and the level of expertise regarding the population under study. Unknown sources of differential performance on items across populations and subpopulations might exist and we might have been unable to identify these through DIF testing. Regardless, the principles underlying our statistical harmonisation procedure would apply even if new sources of DIF are identified within studies. Statistical DIF testing identified only three of 20 domain-by-country categories in which DIF influenced scores in more than 10% of a study sample; for these cases, our DIF adjustments enable the estimation of comparable scores in the presence of the observed differences. The relatively low comparability observed in the orientation and language domains might be attributable to insufficient scrutiny of the cultural relevance of seemingly simple questions, such as orientation to year, which we consider in the appendix (p 3).

Strengths of this study include nationally (or, in the case of HAALSI, regionally) representative sampling and comprehensive cognitive phenotyping with a common protocol. All data are publicly available. Our harmonisation approach based on item banking is readily

scalable: as data from more HCAPs are released or become available, and as longitudinal data from existing HCAPs become available, they can be readily added to our item bank to be harmonised alongside the data therein. However, this study also has important limitations. The quality of the linking between studies is best with a high number of cognitive test items with rich distributions. This need poses challenges when domains largely include relatively easy dichotomous items that have ceiling effects (eg, language and orientation). A further limitation is that, although we identified DIF, it was outside the scope of this study to characterise possible explanations for DIF across HCAP batteries for each test item. Studying the causes of DIF is a worthwhile aim for future research, especially as test batteries are adapted for other countries and contexts. We used MIMIC models for the detection of DIF, and the accuracy of DIF varies considerably across different methods, particularly with respect to the identification of DIF magnitude for particular items.[38] However, the presence of salient DIF in a battery, regardless of item attribution, is usually evident irrespective of approach. In addition, the identification of equivalent items across studies was done on the basis of expert reviews of available documentation, coupled with statistical DIF testing; the expert review step was ultimately a subjective process. A final limitation is that the minimum ages in the included studies ranged from 50 years to 65 years. Given that age is a major determinant of cognitive performance, analyses that were intended to identify cross-study DIF might have detected differences attributable to age instead of country. Although DIF was detected for many items, salient DIF in at least 10% of participants in a sample was only noted in three domain-by-country categories; thus, age is not likely to have been a primary driver of DIF in this study.

In conclusion, the HCAP suite of cognitive test items administered in China, England, India, Mexico, South Africa, and the USA reflects a common structure of general and domain-specific cognitive function across these diverse countries. Despite common protocols, necessary item adaptations were used to account for language, literacy, numeracy, and cultural differences across participating countries. Statistical harmonisation involving an item banking approach with identification of common and unique items allowed for the construction of reliable and valid factor scores that account for these differences. Future cross-national comparisons of risk factors for cognitive ageing outcomes, estimates of dementia prevalence and incidence, and estimates of population-attributable fractions of risk factors should consider using harmonised factor scores to improve the quality of their analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data sharing

Data and codebooks for cognitive tests from each Harmonized Cognitive Assessment Protocol (HCAP) are publicly available on study-specific websites. The HCAP data are available as follows: for the US Health and Retirement Study, at https://hrsdata.isr.umich.edu/dataproducts/; for the English Longitudinal Study on Ageing, at https://www.elsa-project.ac.uk/hcap; for the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia, at https://g2aging.org/?section=lasi-dad-downloads; for the Mexican Cognitive Aging Ancillary Study, at https://www.mhasweb.org/DataProducts/AncillaryStudies.aspx; for the China Health and Retirement Longitudinal Study, at https://charls.charlsdata.com/pages/data/111/zh-cn.html ; and for the Health and Aging in Africa: a Longitudinal Study in South Africa, at https://haalsi.org/data. Data and documentation for cognitive composite variables described in this study will be shared as user-contributed datasets on these websites. Data are shared via these websites after signing of a data access agreement.

References

1. Nichols E, Steinmetz JD, Vollset SE, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022; 7: e105–25. [PubMed: 34998485]

2. Alzheimer's Disease International, Patterson C. World Alzheimer Report 2018. The state of the art of dementia research: new frontiers. 2018. https://www.alzint.org/resource/world-alzheimer-report-2018/ (accessed Sept 7, 2023).

3. Alzheimer's Disease International, Prince M, Wimo A, et al. World Alzheimer Report 2015. The global impact of dementia: an analysis of prevalence, incidence, cost and trends. August, 2015. https://www.alzint.org/resource/world-alzheimer-report-2015/ (accessed Sept 7, 2023).

4. Langa KM, Ryan LH, McCammon RJ, et al. The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: study design and methods. Neuroepidemiology 2020; 54: 64–74. [PubMed: 31563909]

5. Briceño EM, Arce Rentería M, Gross AL, et al. A cultural neuropsychological approach to harmonization of cognitive data across culturally and linguistically diverse older adult populations. Neuropsychology 2023; 37: 247–57. [PubMed: 35482625]

6. Vonk JMJ, Gross AL, Zammit AR, et al. Cross-national harmonization of cognitive measures across HRS HCAP (USA) and LASI-DAD (India). PLoS One 2022; 17: e0264166. [PubMed: 35213581]

7. Bassil DT, Farrell MT, Wagner RG, et al. Cohort profile update: cognition and dementia in the Health and Aging in Africa Longitudinal Study of an INDEPTH community in South Africa (HAALSI dementia). Int J Epidemiol 2022; 51: e217–26. [PubMed: 34871405]

8. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014; 43: 61–68. [PubMed: 23243115]

9. Mejia-Arango S, Nevarez R, Michaels-Obregon A, et al. The Mexican Cognitive Aging Ancillary Study (Mex-Cog): study design and methods. Arch Gerontol Geriatr 2020; 91: 104210. [PubMed: 32781379]

10. Lee J, Khobragade PY, Banerjee J, et al. Design and methodology of the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia (LASI-DAD). J Am Geriatr Soc 2020; 68 (suppl 3): S5–10. [PubMed: 32815602]

11. Cadar D, Abell J, Matthews FE, et al. Cohort profile update: the Harmonised Cognitive Assessment Protocol sub-study of the English Longitudinal Study of Ageing (ELSA-HCAP). Int J Epidemiol 2021; 50: 725–26i. [PubMed: 33370436]

12. Gómez-Olivé FX, Montana L, Wagner RG, et al. Cohort profile: Health and Ageing in Africa: A Longitudinal Study of an INDEPTH community in South Africa (HAALSI). Int J Epidemiol 2018; 47: 689–90j. [PubMed: 29325152]

13. Keith TZ, Reynolds MR. Cattell–Horn–Carroll abilities and cognitive tests: what we've learned from 20 years of research. Psychol Sch 2010; 47: 635–50.

14. Lezak MD. Neuropsychological assessment. New York, NY: Oxford University Press, 2004.

15. Mukherjee S, Choi S-E, Lee ML, et al. Cognitive domain harmonization and cocalibration in studies of older adults. Neuropsychology 2023; 37: 409–23. [PubMed: 35925737]

16. Gross AL, Khobragade PY, Meijer E, Saxton JA. Measurement and structure of cognition in the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia. J Am Geriatr Soc 2020; 68 (suppl 3): S11–19. [PubMed: 32815599]

17. Arce Rentería M, Manly JJ, Vonk JMJ, et al. Midlife vascular factors and prevalence of mild cognitive impairment in late-life in Mexico. J Int Neuropsychol Soc 2022; 28: 351–61. [PubMed: 34376262]

18. UNESCO Institute for Statistics. International standard classification of education: ISCED 2011. 2012. https://uis.unesco.org/sites/default/files/documents/international-standardclassification-of-education-isced-2011-en.pdf (accessed Sept 7, 2023).

19. Kobayashi LC, Gross AL, Gibbons LE, et al. You say tomato, I say radish: can brief cognitive assessments in the US Health Retirement Study be harmonized with its international partner studies? J Gerontol B Psychol Sci Soc Sci 2021; 76: 1767–76. [PubMed: 33249448]

20. Jones RN, Manly JJ, Langa KM, et al. Factor structure of the Harmonized Cognitive Assessment Protocol neuropsychological battery in the Health and Retirement Study. J Int Neuropsychol Soc 2023; published online July 14. 10.1017/S135561772300019X

21. Jones RN. Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination. Detecting differential item functioning using MIMIC modeling. Med Care 2006; 44 (suppl 3): S124–33. [PubMed: 17060819]

22. Zwick R A review of ETS differential item functioning assessment procedures: flagging rules, minimum sample size requirements, and criterion refinement. ETS Res Rep Ser 2012; 2012: i–30.

23. Crane PK, Gibbons LE, Narasimhalu K, Lai J-S, Cella D. Rapid detection of differential item functioning in assessments of health-related quality of life: the Functional Assessment of Cancer Therapy. Qual Life Res 2007; 16: 101–14. [PubMed: 17111233]

24. Clouston SAP, Smith DM, Mukherjee S, et al. Education and cognitive decline: an integrative analysis of global longitudinal studies of cognitive aging. J Gerontol B Psychol Sci Soc Sci 2020; 75: e151–60. [PubMed: 31059564]

25. Lenehan ME, Summers MJ, Saunders NL, Summers JJ, Vickers JC. Relationship between education and age-related cognitive decline: a review of recent research. Psychogeriatrics 2015; 15: 154–62. [PubMed: 25516261]

26. Westrick AC, Avila-Rieger J, Gross AL, et al. Does education moderate gender disparities in later-life memory function? A cross-national comparison of harmonized cognitive assessment protocols in the United States and India. Alzheimers Dement 2023; published online July 25. 10.1002/alz.13404.

27. Li R, Singh M. Sex differences in cognitive impairment and Alzheimer's disease. Front Neuroendocrinol 2014; 35: 385–403. [PubMed: 24434111]

28. Muthén LK, Muthén BO. Mplus user's guide, 8th edn. Los Angeles, CA: Muthén & Muthén, 2017.

29. McDonald RP. Test theory: a unified treatment. Mahwah, NJ: Erlbaum, 1999.

30. Nunnally JC. Psychometric theory, 2nd edn. New York, NY: McGraw Hill, 1978.

31. Skirbekk V, Loichinger E, Weber D. Variation in cognitive functioning as a refined approach to comparing aging across countries. Proc Natl Acad Sci USA 2012; 109: 770–74. [PubMed: 22184241]

32. Savva GM, Maty SC, Setti A, Feeney J. Cognitive and physical health of the older populations of England, the United States, and Ireland: international comparability of the Irish Longitudinal Study on Ageing. J Am Geriatr Soc 2013; 61 (suppl 2): S291–98. [PubMed: 23662721]

33. Hong I, Reistetter TA, Díaz-Venegas C, Michaels-Obregon A, Wong R. Cross-national health comparisons using the Rasch model: findings from the 2012 US Health and Retirement Study and the 2012 Mexican Health and Aging Study. Qual Life Res 2018; 27: 2431–41. [PubMed: 29748824]

34. de la Fuente J, Caballero FF, Verdes E, et al. Are younger cohorts in the USA and England ageing better? Int J Epidemiol 2019; 48: 1906–13. [PubMed: 31873752]

35. Lyu J, Lee CM, Dugan E. Risk factors related to cognitive functioning: a cross-national comparison of US and Korean older adults. Int J Aging Hum Dev 2014; 79: 81–101.

36. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020; 396: 413–46. [PubMed: 32738937]

37. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013; 9: 63–75.e2. [PubMed: 23305823]

38. Teresi JA, Ramirez M, Lai J-S, Silver S. Occurrences and sources of differential item functioning (DIF) in patient-reported outcome measures: description of DIF methods, and review of measures of depression, quality of life and general health. Psychol Sci Q 2008; 50: 538. [PubMed: 20165561]

Research in context

Evidence before this study

A recent *Lancet* Commission reported 12 potentially modifiable risk factors for dementia. Most research on risk and protective factors for dementia has been done in high-income countries; however, an estimated three-quarters of the 152 million people with dementia by 2050 will be living in low-income and middle-income countries. The Harmonized Cognitive Assessment Protocol (HCAP) network is the largest concerted global effort to date to conduct harmonised large-scale population-representative studies of cognitive ageing and dementia. We searched PubMed from database inception up to Aug 3, 2023, for articles addressing cross-national harmonisation of cognitive data in older adults in large population cohorts. The search terms used were ("cross-national" OR "international") AND ("aging" OR "older adults") AND ("cognition" OR "cognitive" OR "cognitive ability" OR "cognitive function") AND ("statistical harmonization" OR "harmonization"). We found no previous studies that have applied rigorous psychometric methods to statistically co-calibrate cognition across more than two countries.

Added value of this study

We statistically harmonised the HCAP instruments fielded in China, England, India, Mexico, South Africa, and the USA, by means of an investigation of the performance of 78 cognitive test items administered to 21 144 older adults (aged 50–108 years) across six large studies of ageing. Harmonisation involved assigning cognitive test items to domains, establishing which test items were common and which were unique across countries, deriving harmonised factor scores for general and specific cognitive domains, and evaluating the reliability and validity of the harmonised factor scores.

Implications of all the available evidence

Despite differences in cognitive test administration due to language, literacy, and numeracy, we statistically harmonised general and domain-specific cognitive function across six countries. These harmonised cognitive function scores empirically reflect comparable domains of cognitive function among older adults living across these countries, are reliable and valid measures of cognitive function, and are useful for population-based research. They can be used in future pooled analyses to evaluate whether risk factors have similar effects on cognitive function across global settings. Such knowledge could facilitate the identification of contextual risk-modifying factors that could be intervened upon to reduce the risk of dementia in specific populations.
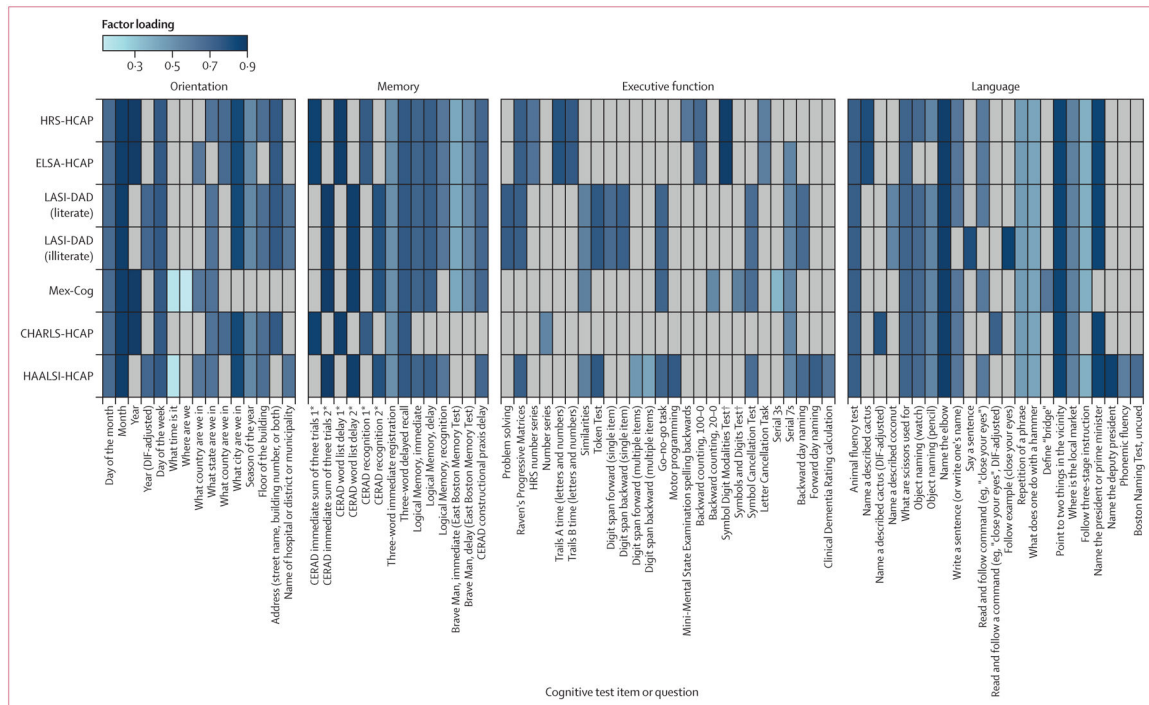
Figure 1: Heatmap of cognitive test item indicators and their overlap across each HCAP study
The heatmap shows factor loadings from domain-specific factor analyses. Loadings are standardised to have a range from –1 to 1, and can thus be interpreted as correlations between items and the underlying factor. All loadings presented are positive, indicating that items are positively correlated. The presence of factor loadings for a given test item in each column reflects decisions about the comparability of items made during prestatistical harmonisation and after testing for DIF. Grey cells indicate cognitive test items not administered in the HCAP battery for a given country. CERAD=Consortium to Establish a Registry for Alzheimer's Disease. CHARLS=China Health and Retirement Longitudinal Study. DIF=differential item functioning. ELSA=English Longitudinal Study on Ageing. HAALSI=Health and Aging in Africa: A Longitudinal Study in South Africa. HCAP=Harmonized Cognitive Assessment Protocol. HRS=US Health and Retirement Study. LASI-DAD=Longitudinal Aging Study in India-Diagnostic Assessment of Dementia. Mex-Cog=Mexican Health and Aging Study Cognitive Aging Ancillary Study. *1 and 2 notation reflects judgements made during prestatistical harmonisation on the similarities of the tests. †110 items, in which participants assign numbers to symbols on the basis of a key. ‡56 items, in which participants assign symbols provided on the basis of a key.
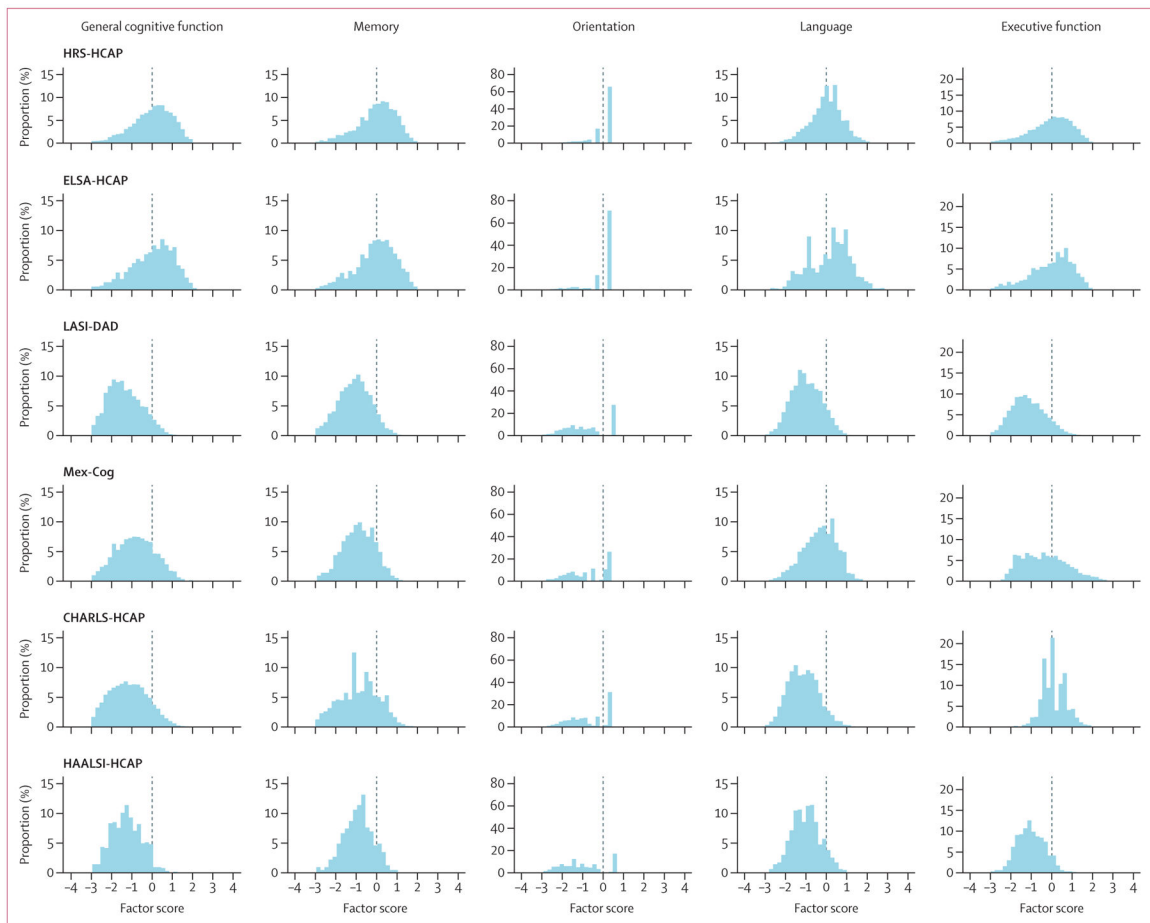
Figure 2: Distribution of harmonised general and domain-specific cognitive factor scores in the HCAP studies

The histograms show general cognitive function (first column) and domain-specific cognitive factors (subsequent columns) by study. Bar heights in each histogram indicate relative frequency of observations with a value along the x-axis. CHARLS=China Health and Retirement Longitudinal Study. ELSA=English Longitudinal Study on Ageing. HAALSI=Health and Aging in Africa: a Longitudinal Study in South Africa. HCAP=Harmonized Cognitive Assessment Protocol. HRS=US Health and Retirement Study. LASI-DAD=Longitudinal Aging Study in India-Diagnostic Assessment of Dementia. Mex-Cog=Mexican Health and Aging Study Cognitive Aging Ancillary Study.
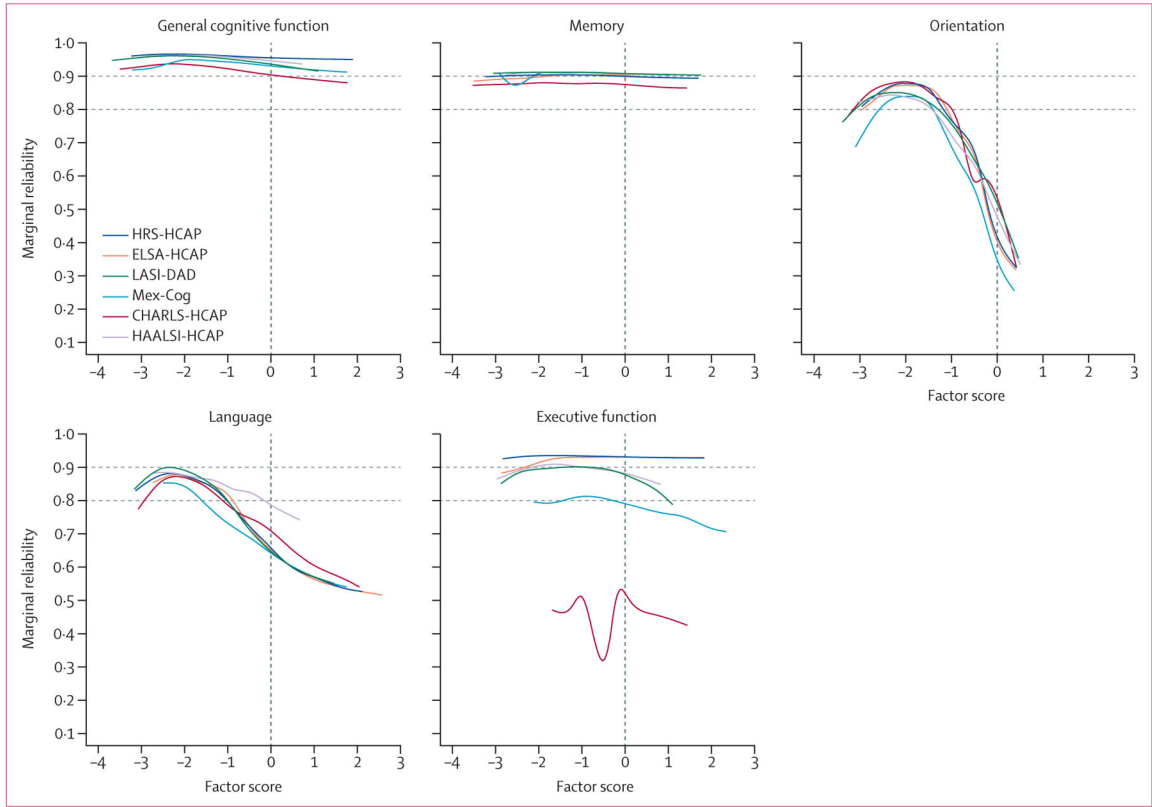
Figure 3: Marginal reliability for overall and domain-specific cognitive performance in each HCAP study

Plots show the marginal reliability of general and domain-specific factor scores for each study. This figure shows differences in the reliability of estimated factor scores as a function of corresponding abilities on the latent trait. Horizontal dashed lines at reliabilities of 0·8 and 0·9 indicate acceptable thresholds of reliability for basic research and high-stakes testing, respectively.[30] CHARLS=China Health and Retirement Longitudinal Study. ELSA=English Longitudinal Study on Ageing. HAALSI=Health and Aging in Africa: a Longitudinal Study in South Africa. HCAP=Harmonized Cognitive Assessment Protocol. HRS=US Health and Retirement Study. LASI-DAD=Longitudinal Aging Study in India-Diagnostic Assessment of Dementia. Mex-Cog=Mexican Health and Aging Study Cognitive Aging Ancillary Study.

Table 1:

Characteristics of included HCAP cohorts

	Total sample	HRS-HCAP	ELSA-HCAP	LASI-DAD	Mex-Cog	CHARLS-HCAP	HAALSI-HCAP
Sample size	21 144	3347	1273	4096	2042	9755	631
Age, years	69 (64–76)	76 (70–82)	75 (70–81)	70 (64–74)	68 (61–74)	67 (63–73)	69 (59–78)
Gender							
Female	11480 (54·3%)	2020 (60·4%)	700 (54·9%)	2207 (53·9%)	1203 (58·9%)	4960 (50·8%)	390 (61·8%)
Male	9664 (45·7%)	1327 (39·6%)	573 (45·0%)	1889 (46·1%)	839 (41·1%)	4795 (49·2%)	241 (38·2%)
Education * Highest educational attainment							
No education or early childhood education	8862 (41·9%)	22 (0·7%)	3 (0·2%)	2558 (62·5%)	1023 (50·1%)	4909 (50·3%)	347 (55·0%)
Primary	3677 (17·4%)	131 (3·9%)	0	527 (12·9%)	452 (22·1%)	2355 (24·1%)	212 (33·6%)
Lower secondary	3160 (14·9%)	454 (13·6%)	486 (38·2%)	314 (7·7%)	317 (15·5%)	1562 (16·0%)	27 (4·3%)
Upper secondary	3465 (16·4%)	1773 (52·9%)	303 (23·8%)	505 (12·3%)	60 (2·9%)	792 (8·1%)	32 (5·1%)
Post-secondary	1924 (9·1%)	965 (28·8%)	446 (35·0%)	192 (4·7%)	172 (8·4%)	137 (1·4%)	12 (1·9%)
Missing	56 (0·3%)	2 (0–1%)	35 (2·8%)	0	18 (0·9%)	0	1 (0·2%)

Data are n, n (%), or median (IQR). CHARLS=China Health and Retirement Longitudinal Study. ELSA=English Longitudinal Study on Ageing. HAALSI=Health and Aging in Africa: A Longitudinal Study in South Africa. HCAP=Harmonized Cognitive Assessment Protocol. HRS=US Health and Retirement Study. LASI-DAD=Longitudinal Aging Study in India-Diagnostic Assessment of Dementia. Mex-Cog=Mexican Health and Aging Study Cognitive Aging Ancillary Study.

*
Scaled to the 2011 International Standard Classification of Education (appendix p 19).[18]

Table 2:

Number of participants with scores with salient DIF by HCAP study and cognitive domain

	HRS-HCAP (n=3347)[*]	ELSA-HCAP (n=1273)	LASI-DAD (n=4096)	Mex-Cog (n=2042)	CHARLS-HCAP (n=9755)	HAALSI-HCAP (n=631)
Memory	Ref	2 (0·2%)	1 (<0·1%)	105 (5·1%)	216 (2·2%)	0
Orientation	Ref	22 (1·7%)	0	0	23 (0·2%)	0
Language	Ref	57 (4·5%)	23 (0·6%)	50 (2·4%)	1637 (16·8%)	6 (0.9%)
Executive function	Ref	0	0	NA	NA	0

Salience of DIF was calculated as the difference between DIF-adjusted and non-DIF-adjusted factor scores. The number of participants whose DIF-adjusted scores differed by more than 0·3 SDs from non-DIF-adjusted scores are shown here. Numbers in this table are corrected for the identified salient DIF that was detected in >10% of a study sample, with such items removed as linking items and retained as non-linking items. CHARLS=China Health and Retirement Longitudinal Study. DIF=differential item functioning. ELSA=English Longitudinal Study on Ageing. HAALSI=Health and Aging in Africa: a Longitudinal Study in South Africa. HCAP=Harmonized Cognitive Assessment Protocol. HRS=US Health and Retirement Study. LASI-DAD=Longitudinal Aging Study in India-Diagnostic Assessment of Dementia. Mex-Cog=Mexican Health and Aging Study Cognitive Aging Ancillary Study. NA=not applicable (no overlap).

[*] HRS-HCAP was the reference study against which DIF was evaluated for items in other studies.

Table 3:

Validation of the general cognitive function factor in HCAP studies

	Overall sample (n=21 141)	HRS-HCAP (n=3347)	ELSA-HCAP (n=1273)	LASI-DAD (n=4096)	Mex-Cog (n=2042)	CHARLS-HCAP (n=9755)	HAALSI-HCAP (n=631)
Sex							
Female	−0·00 (−0·03 to 0·02)	0·17 (0·12 to 0·23) *	0·05 (−0·05 to 0·15)	−0·19 (−0·23 to −0·15) *	−0·10 (−0·16 to −0·04) *	−0·08 (−0·12 to −0·05) *	−0·24 (−0·32 to −0·15) *
Male	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)
Age group, years							
50–59	0·34 (0·28 to 0·41) *	NA	NA	NA	0·14 (0·06 to 0·22) *	NA	0·21 (0·08 to 0·33) *
60–69	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)
70–79	−0·17 (−0·19 to −0·15) *	−0·27 (−0·34 to −0·20) *	−0·53 (−0·66 to −0·40) *	−0·23 (−0·28 to −0·19) *	−0·38 (−0·45 to −0·31) *	−0·22 (−0·26 to −0·19) *	−0·33 (−0·44 to −0·21) *
80–89	−0·56 (−0·59 to −0·52) *	−0·83 (−0·91 to −0·75) *	−1·16 (−1·30 to −1·02) *	−0·52 (−0·59 to −0·45) *	−0·89 (−0·99 to −0·79) *	−0·55 (−0·62 to −0·49) *	−0·59 (−0·72 to −0·45) *
90	−1·05 (−1·12 to −0·97) *	−1·45 (−1·58 to −1·32) *	−1·83 (−2·10 to −1·55) *	−0·89 (−1·02 to −0·75) *	−1·36 (−1·58 to −1·14) *	−0·74 (−1·03 to −0·44) *	−0·76 (−1·00 to −0·52) *
Highest educational attainment							
No education or early childhood education	−1·17 (−1·20 to −1·13) *	−0·57 (−0·92 to −0·23) *	−0·66 (−1·66 to 0·35)	−0·89 (−0·97 to −0·82) *	−1·12 (−1·20 to −1·03) *	−1·07 (−1·12 to −1·03) *	−0·67 (−0·89 to −0·45) *
Primary	−0·40 (−0·44 to −0·37) *	−0·59 (−0·75 to −0·44) *	NA	−0·22 (−0·31 to −0·13) *	−0·41 (−0·51 to −0·32) *	−0·31 (−0·36 to −0·26) *	−0·22 (−0·44 to −0·00) *
Lower secondary	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)
Upper secondary	0·50 (0·46 to 0·54) *	0·68 (0·60 to 0·77) *	0·39 (0·26 to 0·51) *	0·36 (0·27 to 0·45) *	0·10 (−0·09 to 0·28)	0·27 (0·20 to 0·33) *	0·33 (0·05 to 0·61) *
Post-secondary	0·98 (0·94 to 1·03) *	1·12 (1·03 to 1·21) *	0·61 (0·50 to 0·73) *	0·60 (0·49 to 0·71) *	0·56 (0·44 to 0·68) *	0·64 (0·50 to 0·78) *	0·64 (0·27 to 1·00) *

Values are β coefficients (95% CIs). β coefficients represent overall and study-specific mean differences in general cognitive function factor scores between a given exposure grouping and the reference category. A β coefficient below 0 represents worse cognitive function and a value above 0 represents higher cognitive function *vs* the reference group. Models for each exposure were mutually adjusted for other exposures in the table. CHARLS=China Health and Retirement Longitudinal Study. ELSA=English Longitudinal Study on Ageing. HAALSI=Health and Aging in Africa: a Longitudinal Study in South Africa. HCAP=Harmonized Cognitive Assessment Protocol. HRS=US Health and Retirement Study. LASI-DAD=Longitudinal Aging Study in India-Diagnostic Assessment of Dementia. Mex-Cog=Mexican Health and Aging Study Cognitive Aging Ancillary Study. NA=not applicable (ie, no observations in the group).

*
p<0·05.