USC Dornsife Center for Economic and Social Research

USC Schaeffer Leonard D. Schaeffer Center for Health Policy & Economics

# Considerations around the measurement of cognition in large-scale cross-national surveys: Lessons from the Health and Retirement International Network of Surveys (HRS INS) and the Harmonized Cognitive Assessment Protocol (HCAP)

*Emma Nichols & Jinkook Lee*

*Paper No: 2024-012*

## CESR-SCHAEFFER WORKING PAPER SERIES

**Considerations around the measurement of cognition in large-scale cross-national surveys: Lessons from the Health and Retirement International Network of Surveys (HRS INS) and the Harmonized Cognitive Assessment Protocol (HCAP)**

Emma Nichols[1,2] & Jinkook Lee[1,3]

1. Center for Economic and Social Research, University of Southern California, Los Angeles, CA
2. Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA
3. Department of Economics, University of Southern California, Los Angeles, CA

## Introduction

With dementia prevalence expected to triple by 2050 (1), continued resources are needed to enhance our understanding of factors related to disease risk, resilience, and disparities. Though progress has been made in developing drugs that have been successful in removing the proteins hypothesized to underlie Alzheimer's disease (2,3), large questions remain about how well these treatments work (i.e., are observed benefits clinically meaningful), and for whom they work, especially in the face of dangerous side-effects (4,5). While research in search of effective treatments continues, evidence on life-course modifiable risk factors also points to the potential for policy and lifestyle interventions that may be able to shift the needle at the population level (6). Accumulating evidence supports the role of a variety of lifecourse risk factors related to cardiometabolic disease (e.g., hypertension, cholesterol, diabetes), sensory impairment (e.g., vision and hearing loss), and social and physical environment (e.g., education, social isolation, air pollution) (7,8). However, much of the existing evidence supporting this literature comes from geographically restricted studies in high-income settings (9).

Though geographically restricted samples can provide important insights, generalizability can be a concern; in contrast, population-representative samples can provide evidence for the same population that would be impacted by large-scale policy changes or other evidence-based interventions. In the same way that results from one area of a country may not generalize to a national level, findings from one country are even less likely to generalize internationally due to differences in culture and context, including the range of exposures people experience across the lifecourse (10,11). Beyond the importance of conducting regionally relevant research, collecting data across contexts to facilitate cross-national comparisons can lead to novel insights through the comparison of settings with differences in exposure distributions or differences in effect modifiers. Furthermore, comparisons of associations across settings with different contexts and therefore different confounding structures can provide an important form of evidence triangulation and strengthen causal arguments (12).

The Health and Retirement (HRS) study was originally established in the early 1990s to study the economics of retirement and health among older adults in the United States (US). Over time, the HRS has expanded and grown to include over 37,000 individuals and is nationally representative of those over 50, becoming a key data resource for research on cognitive aging in the US (13). HRS has also become the model for a growing number of studies in the HRS International Network of

Studies (HRS INS) which seek to replicate the design, structure, and content included in the HRS in order to facilitate comparative research around the world (14–17). One of the key strengths of these studies is the breadth of information collected across a wide range of topics; however, the time devoted to any given topic is limited, therefore restricting the feasibility of administering more detailed scales and assessments with high measurement precision. Prior evaluations of the cognitive battery concluded that included tests were reasonable given constraints, but raised concerns about the reliability and validity of the battery (18). In 2001, the HRS study team began data collection for the Aging Demographics and Memory Study (ADAMS), a sub-study of the HRS designed to collect more detailed cognitive measures and conduct a full clinical evaluation for dementia. While successful in its primary goals, the expensive protocol included in ADAMS limited sample size and cohort diversity.

Building on lessons from the ADAMS sub-study, the Harmonized Assessment Protocol (HCAP) was the second sub-study designed to augment the cognitive measures included in the HRS (19). In addition to serving as a key data resource in HRS, one of the key goals from the inception of HCAP was to launch a network of partner sub-studies to enable international comparisons of cognitive aging across countries (19). The initial wave of data collection using the HCAP has now been completed in the US, England, Mexico, India, China, Chile, and Europe, with the second round of data collection already underway in a handful of countries (19–22).

Data collection efforts and analyses of cognitive data in the HRS INS and the HCAP studies have spurred substantial research and led to wide-ranging insights, not only on associations between risk factors and cognitive outcomes but on topics related to the measurement of cognition as well (23–25). In particular, the comparable measurement of cognition cross-nationally given differences in language, context, literacy, and education has been the topic of considerable attention and research (26–28). The goal of the present paper is to summarize key lessons learned in developing and administering cognitive assessments cross-nationally using both the brief HRS INS cognitive battery and the HCAP. Guidance and insights from past data collection efforts can be used to improve existing data collection efforts and inform the design of new studies.

**Overall key considerations**

Based on experiences and prior research within the HRS INS and HCAP networks, we highlight three overarching considerations that researchers should keep in mind to enhance the comparability and performance of cognitive batteries for use in cross-national research:

1. Consistency across studies and over time is key
2. Cognitive tests should be acceptable and feasible in both high-income and lower- and middle-income countries (LMIC)
3. Longer test batteries perform better

Though straightforward, these guiding principles should be considered when making key decisions on the items to include and the resources to devote to cognitive testing. Further nuance on each of these topics provides additional insights that may help guide choices.

Consistency across studies and over time

Assessment consistency over time allows for longitudinal modeling and the estimation of cognitive decline, strengthening the evidence on associations between risk factors and cognitive outcomes. Consistency over time can also help researchers account for learning or practice effects, which are often strongest between the first and second assessments, waning over time (29). Although some researchers have previously suggested that small tweaks to tests may help alleviate practice effects, this can complicate efforts to model practice effects and account for them statistically (30). Ultimately, consistent estimation of cognition over time using the same cognitive tests is the most straightforward way to facilitate the accurate modeling of longitudinal trajectories, accounting for nuisance factors such as practice effects.

Consistency across studies is key in facilitating comparisons across studies as well. The application of modern psychometric methods rooted in item response theory (IRT) for the harmonization of cognitive functioning across studies has been increasingly used across study consortia (31,32) and relies on the existence of common or "linking" items across studies. Adequate and statistically appropriate linkage requires not only that items are identical, but that administration differences are minimized as well. For example, though all studies had administered 10-word recall when conducting statistical harmonization of the studies with the HCAP battery, investigators chose to separate 10-word recall into two items: one version in HRS-HCAP, ELSA-HCAP, and CHARLS-HCAP which gave both visual and auditory stimuli, and another version in LASI-DAD, Mex-Cog, and, Chile-Cog which gave only auditory stimuli (31). Beyond differences in the different tests used, these administration differences, though small, can pose substantial barriers to the successful harmonization of cognition across settings.

Acceptability and feasibility across contexts

In line with the goal of increased consistency across countries, it is optimal to select cognitive items that can be validly administered across all contexts of interest. Cognitive tests with substantial literacy or numeracy requirements may work well in high-income contexts, but pose challenges in LMICs (28,33). Items on orientation to place or time are less relevant in contexts where individuals have lower awareness of Western calendars or geographic boundaries beyond their village (28,34). A number of studies in LMICs using the HCAP battery have also made adjustments to accommodate populations with low literacy and numeracy, using for example symbol cancellation instead of letter cancellation (35). Should these tests function adequately in high-income contexts, adoption of the versions used in LMICs would serve to further strengthen comparability and linkage across countries.

Beyond objective cognitive testing, even in informant reports of cognition, care needs to be taken to ensure the items are culturally relevant and perform similarly across contexts (36). Prior work in the LASI-DAD study has shown that some items of the Informant Questionnaire for Cognitive Decline in the Elderly had limited cultural relevance, leading to high endorsement of the "doesn't do" missing data category for various items (37). Imputation of missing data in these items help increase comparability, but existing evidence has also suggested that the performance of the scale may be preserved even after dropping items with lower cultural relevance across settings (37).

Length of the test battery

The Spearman-Brown Prophecy Formula and classical test theory describe the magnitude of improvements to the reliability of latent trait estimates with the addition of new items to an existing scale (38,39). Given scale reliability influences the standard error of measurement for estimates of cognitive functioning, increasing test length and therefore reliability is a key mechanism by which it is possible to reduce measurement error, increase the ratio of signal to noise, and increase power in subsequent analyses based on estimates of cognitive functioning. Though the length of assessment overall is less of a concern in analyses using the HCAP battery, given the approximately 1-hour cognitive assessment, this is still a concern for the estimation of specific cognitive domains such as visuospatial functioning, for which the HCAP battery only included a couple of tests, limiting the precision of estimates (40).

Furthermore, longer batteries may allow for higher precision across the full range of cognitive functioning, as it is easier to include tests with varying levels of difficulty. This can be important if capturing both dementia as well as more mild forms of impairment are key study goals. For studies focused on either cross-sectional differences between groups or the estimation  of longitudinal trajectories, it is important to have good measurement properties and high precision across the range of cognitive functioning to ensure that findings are due to underlying abilities rather than psychometric attributes of the measures of interest (41).

**Evidence and specific lessons for the design of brief survey instruments for measuring cognition in cross-national studies**

Existing studies within the HRS INS have included measures of orientation, memory, language/fluency, executive functioning, and visuospatial functioning (Table 1). Only orientation to time (orientation) and immediate and delayed word recall (memory) have been consistently administered across all studies. Of language tests, object naming (10/14 studies) and animal naming (8/14 studies) were most consistently administered. However, object naming assessments can differ in the specific objects included, impacting comparability. Serial 7s (12/14 studies) is the most common item on executive functioning, whereas other cognitive test items on executive functioning have been included more sporadically. Visuospatial functioning assessments were included in only a handful of studies in the HRS INS.

Although the historical precedent of what has been previously administered in the HRS INS should inform future decisions in attempts to maximize alignment over time and geographically, other factors should be considered in tandem. One challenge with many binary items in the existing batteries is low variability (e.g., because the items are easy, very few people answer incorrectly) (28). This is a particular challenge with assessments of orientation and object naming, which are commonly administered across the HRS INS. These assessments were initially developed for use in clinical settings, or as cognitive screening instruments for dementia (42,43). Although such assessments are helpful in identifying individuals with very severe cognitive impairment or dementia, they are not typically helpful for understanding more mild impairment or cognitive decline. In population-based settings and studies, where relatively few people have severe impairments, these items will only be helpful in quantifying cognition for a select number of people with extreme impairment. While the identification of those with dementia or severe cognitive impairment may be of particular interest, warranting the focus on capturing low cognitive performance, extremely low performance on continuous assessments of cognition (e.g., word recall or animal naming) also points towards severe cognitive impairment or dementia (44). Similar, continuous assessments of cognition can provide information across a range of cognitive functioning and are helpful not only for identifying dementia but also for the measurement of subtle decline or mild deficits.

Continuous measures of cognition with the ability to discriminate across a wide range of cognitive functioning can provide critical information for cross-country comparisons when acceptable across settings and administered comparatively across countries. The utility of such measures is highlighted by the extent to which the existing literature on cross-national comparisons of cognitive functioning in the HRS INS relies on the word recall and animal naming tasks (45–49). As a continuous measure of cognition that is informative across the range of cognitive functioning, and which has been administered comparably across countries, word recall in particular provides an important link between studies and facilitates comparisons. However, it is important to note limitations; word recall is a memory measure, and does not capture other aspects of general cognitive functioning, nor should the measure be interpreted as general cognitive functioning (50). The addition of other continuous measures administered consistently across countries in the HRS INS core batteries may bolster the ability to make more generalized conclusions.

Although adding or modifying existing test batteries in the HRS INS will have important and meaningful effects on future cross-sectional research or prospective longitudinal research, benefits on longitudinal research including information from prior waves may be limited. Though methodological approaches based in IRT methods could theoretically allow for the estimation of cognitive functioning on a constant metric when cognitive tests change over time (51), the performance of such methods requires the reliability of the measure to be unchanged by addition or subtract of cognitive test items. When these approaches were testing in the English Longitudinal Study on Aging to incorporate additional cognitive tests adding in Wave 7, changes to the reliability led to spurious longitudinal trends (50), suggesting that larger batteries with greater precision across the full range of cognitive functioning are needed to appropriately apply these methods to assessments of cognitive functioning in the HRS INS.

**Evidence and specific lessons for the design of longer survey instruments for measuring cognition in cross-national studies**

The approximately 1-hour HCAP battery contains a larger set of consistent items across studies and greater coverage of the orientation, memory, executive functioning, and language domains (Table 2). The language domain had the largest number of items administered across all or almost all studies, although many items in this domain and in the orientation domain have issues with low variability (i.e., most respondents get these items correct). The memory domain had the largest number of high-quality items administered consistently across all or almost all studies, including word recall (3 items), the two story recall tests (5 items) (Mex-Cog did not include story recognition), and delayed constructional praxis (28,31). However, CHARLS-HCAP did not include either story recall test. Executive functioning has fewer items in common across HCAPs, though all studies have symbol or letter cancellation. It is important to note, however, that letter cancellation (administered mostly in high-income contexts) and symbol cancellation (administered mostly in low-income contexts) cannot be considered comparable despite their similarities (31). Although the HCAP battery is substantially longer than what is included in the HRS INS, assessments of visuospatial functioning in HCAP are brief. This highlights the importance of comprehensive content coverage across domains of interest, as it is difficult to estimate latent constructs with fewer than three indicators (52). When estimating visuospatial functioning using confirmatory factor analysis, the small number of items available in HRS-HCAP posed challenges, and investigators decided to drop the copying polygons item, leaving visuospatial functioning represented by constructional praxis alone (40).

Although the a priori efforts and emphasis on the development of a harmonized network of studies has paid off by leading to greater harmonization of cognitive items administered in studies with the HCAP battery compared to the HRS INS, adjustments may still help in enhancing comparability across studies. Many of the largest changes to the HCAP battery were made in adjusting the battery of tests for use in LMICs, and in populations with low levels of literacy, numeracy, and exposure to test-taking. In addition to changing letter cancellation to symbol cancellation, important adjustments have been introduced in the trail-making test (using colors and shapes, rather than numbers and letters) (53), and naming assessments (adjusting the objects used) (35), among others. A number of new executive functioning tests that do not rely on numeracy, such as the token test and the Go-No-Go test have also been introduced to bolster the assessment of executive functioning in LMICs (35). These types of adjustments are necessary to ensure adequate performance of the batteries across settings but do affect comparability. To the extent possible, the adoption of some of these assessments in high-income countries as well would improve the ability to statistically harmonize estimates of cognitive functioning between countries.

One important addition to the longer HCAP battery is the inclusion of informant assessments of cognitive and physical functioning by a close friend or family member. Informant reports can be particularly helpful in assessing the physical consequences of cognition or the impacts of cognition on everyday functioning, as those with cognitive impairment are unlikely to be able to report their own cognitive failures (54). Supporting the notion that informant reports contain important and meaningful information, the use of screening procedures using both the Mini-Mental State Examination (MMSE) and the Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE) has been shown to outperform the use of either scale in isolation (55,56). Further, in studies that have developed algorithmic approaches to classify dementia using the HCAP battery, data from informant reports plays an important role in applying rules from the Diagnostic and Statistical Manual of Mental Disorders to existing survey data (57,58). Although measurement error in informant reports is likely, and evidence exists that characteristics of the informant, unrelated to cognitive functioning, may influence reporting (36), the sources of bias in informant reports are likely different than those observed in objective cognitive testing, which may help in triangulating evidence to better understand and characterize cognitive functioning.

**Lessons to inform the administration of cognitive testing in cross-national studies**

Beyond the choice of which cognitive tests to administer, there are a multitude of additional details and processes related to the administration and adaptation of cognitive tests that can be incredibly impactful in ensuring the validity and comparability of cognitive testing. The process of cultural and linguistic adaptation in particular should be an ongoing consideration throughout the design and implementation of cognitive testing. In any new setting, pilot testing to ensure cultural relevance is necessary, and participants piloting cognitive tests should ideally mirror the population of interest, with diversity across factors such as education, cognitive ability, language, and geography, all of which may impact testing. This type of diversity is necessary to flag issues related to specific sub-populations or issues in the comparisons of different groups. Studies also need to stay nimble and amenable to change in the case of emerging issues. Due to concerns about the lack of familiarity with cacti in the Indian context, LASI-DAD investigators chose to change this item to naming a coconut in Wave 1 (35). However, inspection of Wave 1 data and conversations with field investigators highlighted important state-level differences – the proportion who answered correct was considerably higher in coastal states in southern India than in mountainous landlocked states in the north (59). Based on this data and feedback, investigators chose to change this item to naming a tree in Wave 2 (59). Comprehensive data monitoring, both as data is collected during waves, and after data collection is complete is important to ensure the validity of cognitive testing, to allow studies to remain responsive to problems that emerge, and to improve the quality of data.

Beyond the cognitive scores directly, missing data on cognitive testing also should be considered. Nuances underlying missing data codes can impact the interpretation and ultimately the scoring of data on cognition. In other types of survey data, all missing data are treated the same, but a missing data code of "don't know" on cognitive testing is typically recoded as 0, as not knowing the answer to a cognitive test implies that they respondent could not answer. In contrast, "refusal" or other missing codes indicate more general forms of missingness and are often imputed (60,61). Given the importance of this distinction, it is imperative that studies allow for these different forms of missingness and train interviewers in their appropriate use. Even in informant questionnaires, the inclusion or exclusion of missing data option such as the "doesn't do" option on the IQCODE questionnaire, has been shown to have detectable impacts on comparisons of mean IQCODE scores (36).

Supplemental questions or assessments to better understand characteristics that may impact the testing environment or informant reports can also provide key insights and enable the application of post-hoc corrections or imputations to improve data quality. Sensory impairment, including vision and hearing loss, as well as physical impairments, including motor issues or tremor, can impact cognitive testing and effect missingness or scores. Both assessments of interviewer observations and objective testing to measure sensory and physical impairments can be helpful in determining the impacts of potential deficits on cognitive testing. As prior research has shown that the characteristics of informants can impact informant reports independent of the respondent's cognitive functioning (36,62–65), it is also important to ensure that this data is captured and to adjust for this potential source of bias in subsequent analyses.

On a larger scale, the mode of administration of cognitive testing in surveys also warrants consideration. Although historically most cognitive testing is done in face-to-face interviews and the HCAP battery is designed to be and is implemented as a face-to-face assessment, the core

HRS INS sometimes utilize different modes of data collection to reduce costs and increase retention. Though HRS has successfully implemented data collection via telephone and web-based surveys to achieve these goals (13), mode shifts can have large impacts on cognitive testing (66,67). Different modes allow for the use of different strategies, including looking up the answers to questions, or writing down word lists, which may make cognitive tests easier. Although face-to-face interviews are preferred for cognitive testing, when utilizing different modes of data collection, the use of calibration samples where both versions of the assessment are administered or the use of randomization to assign interview mode are critical tools to allow the estimation of mode effects and the correction of this source of bias.

**Conclusion**

Despite challenges in the collection of cognitive data across countries in the HRS INS and studies using the HCAP battery, these surveys are an important and valuable data resource for researchers focused on cognitive aging. The implementation of harmonized data collection efforts in these studies has yielded many important lessons in understanding the best practices and key considerations for the measurement of cognition, with focus on harmonization and cross-national comparisons. Choices in the design and implementation of cognitive testing will ultimately require weighing a variety of factors, including consistency across studies and time, feasibility and acceptability, length of the battery, and resource limits. Though there is no single correct way to approach the goal of balancing these factors, key considerations and lessons learned from the implementation of cognitive testing in the HRS INS and in studies using the HCAP to date can provide lessons on the implications of different choices, and emphasize the importance of attention to all components of data collection efforts, including pilot testing and data monitoring, missing data, and collection of information on characteristics that might influence cognitive testing. These considerations and recommendations, based on experiences administering the cognitive batteries in HRS INS and HCAP, should be used to guide choices in upcoming data collection efforts, with future lessons learned from these upcoming data collection efforts continuously considered in updating and refining protocols for cross-national assessments of cognition in future HRS INS waves, studies using the HCAP, and other new data collection efforts.

**References**


1. Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, 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. The Lancet Public Health. 2022 Feb 1;7(2):e105–25.

2. Sims JR, Zimmer JA, Evans CD, Lu M, Ardayfio P, Sparks J, et al. Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. JAMA. 2023 Aug 8;330(6):512–27.

3. van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in Early Alzheimer's Disease. N Engl J Med. 2023 Jan 5;388(1):9–21.

4. Walsh S, Merrick R, Milne R, Nurock S, Richard E, Brayne C. Considering challenges for the new Alzheimer's drugs: Clinical, population, and health system perspectives. Alzheimer's & Dementia [Internet]. [cited 2024 Aug 29];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.14108

5. Digma LA, Winer JR, Greicius MD. Substantial Doubt Remains about the Efficacy of Anti-Amyloid Antibodies. Journal of Alzheimer's Disease. 2024 Jan 1;97(2):567–72.

6. Walsh S, Govia I, Peters R, Richard E, Stephan BCM, Wilson NA, et al. What would a population-level approach to dementia risk reduction look like, and how would it work? Alzheimer's & Dementia. 2023 Jul;19(7):3203–9.

7. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. The Lancet [Internet]. 2024 Jul 31 [cited 2024 Aug 1];0(0). Available from: https://www-thelancet-com.libproxy2.usc.edu/journals/lancet/article/PIIS0140-6736(24)01296-0/fulltext

8. Whalley LJ, Dick FD, McNeill G. A life-course approach to the aetiology of late-onset dementias. The Lancet Neurology. 2006 Jan 1;5(1):87–96.

9. Kalaria R, Maestre G, Mahinrad S, Acosta DM, Akinyemi RO, Alladi S, et al. The 2022 symposium on dementia and brain aging in low- and middle-income countries: Highlights on research, diagnosis, care, and impact. Alzheimer's & Dementia. 2024;20(6):4290–314.

10. Lee J, Phillips D, Wilkens J, Gateway to Global Aging Data Team. Gateway to Global Aging Data: Resources for Cross-National Comparisons of Family, Social Environment, and Healthy Aging. J Gerontol B Psychol Sci Soc Sci. 2021 Jun 8;76(Suppl 1):S5–16.

11. Ailshire J, Carr D. Cross-National Comparisons of Social and Economic Contexts of Aging. J Gerontol B Psychol Sci Soc Sci. 2021 Jun 8;76(Suppl 1):S1–4.

12. Matthay EC, Hagan E, Gottlieb LM, Tan ML, Vlahov D, Adler NE, et al. Alternative causal inference methods in population health research: Evaluating tradeoffs and triangulating evidence. SSM - Population Health. 2020 Apr 1;10:100526.

13. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). International Journal of Epidemiology. 2014 Apr 1;43(2):576–85.

14. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of ageing. Int J Epidemiol. 2013 Dec;42(6):1640–8.

15. Wong R, Michaels-Obregon A, Palloni A. Cohort Profile: The Mexican Health and Aging Study (MHAS). Int J Epidemiol. 2017 Apr 1;46(2):e2.

16. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol. 2014 Feb;43(1):61–8.

17. Sonnega A, Weir DR. The Health and Retirement Study: A Public Data Resource for Research on Aging. Open Health Data. 2014 Oct 16;2:e7–e7.

18. Lachman M, Spiro A. Critique of cognitive measures in the health retirement study (HRS) and the asset and health dynamics among the oldest old (AHEAD) study. HRS Data Monitoring Board Report [Internet]. 2002 [cited 2024 Aug 29]; Available from: https://www.researchgate.net/profile/Avron-Spiro/publication/242096659_Critique_of_Cognitive_Measures_in_the_Health_Retirement_Study_HRS_and_the_Asset_and_Health_Dynamics_among_the_Oldest_Old_AHEAD_Study/links/0c96052869f0bd6f6a000000/Critique-of-Cognitive-Measures-in-the-Health-Retirement-Study-HRS-and-the-Asset-and-Health-Dynamics-among-the-Oldest-Old-AHEAD-Study.pdf

19. Langa KM, Ryan LH, McCammon RJ, Jones RN, Manly JJ, Levine DA, et al. The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. Neuroepidemiology. 2020;54(1):64–74.

20. Lee J, Banerjee J, Khobragade PY, Angrisani M, Dey AB. LASI-DAD study: a protocol for a prospective cohort study of late-life cognition and dementia in India. BMJ Open. 2019 Jul 31;9(7):e030300.

21. Mejia-Arango S, Nevarez R, Michaels-Obregon A, Trejo-Valdivia B, Mendoza-Alvarado LR, Sosa-Ortiz AL, et al. The Mexican Cognitive Aging Ancillary Study (Mex-Cog): Study Design and Methods. Archives of Gerontology and Geriatrics. 2020 Nov 1;91:104210.

22. Cadar D, Abell J, Matthews FE, Brayne C, Batty GD, Llewellyn DJ, 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 Jul 9;50(3):725–726i.

23. Blankson AN, McArdle JJ. A Brief Report on the Factor Structure of the Cognitive Measures in the HRS/AHEAD Studies. Journal of Aging Research. 2014;2014(1):798514.

24. Weir D, Faul J, Langa K. Proxy interviews and bias in the distribution of cognitive abilities due to non-response in longitudinal studies: a comparison of HRS and ELSA. Longit Life Course Stud. 2011 May;2(2):170–84.

25. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of Cognition Using Surveys and Neuropsychological Assessment: The Health and Retirement Study and the Aging,

Demographics, and Memory Study. J Gerontol B Psychol Sci Soc Sci. 2011 Jul;66B(Suppl 1):i162–71.

26. Chan KS, Gross AL, Pezzin LE, Brandt J, Kasper JD. Harmonizing Measures of Cognitive Performance Across International Surveys of Aging Using Item Response Theory. J Aging Health. 2015 Dec 1;27(8):1392–414.

27. Kobayashi LC, Gross AL, Gibbons LE, Tommet D, Sanders RE, Choi SE, et al. You Say Tomato, I Say Radish: Can Brief Cognitive Assessments in the U.S. Health Retirement Study Be Harmonized With Its International Partner Studies? The Journals of Gerontology: Series B. 2021 Nov 1;76(9):1767–76.

28. Nichols E, Ng DK, Hayat S, Langa KM, Lee J, Steptoe A, et al. Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research. Alzheimer's & Dementia. 2023 Mar;19(3):1009–19.

29. Vivot A, Power MC, Glymour MM, Mayeda ER, Benitez A, Spiro A, et al. Jump, Hop, or Skip: Modeling Practice Effects in Studies of Determinants of Cognitive Change in Older Adults. Am J Epidemiol. 2016 Feb 15;183(4):302–14.

30. Gross AL, Inouye SK, Rebok GW, Brandt J, Crane PK, Parisi JM, et al. Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time. J Clin Exp Neuropsychol. 2012;34(7):758–72.

31. Gross AL, Li C, Briceño EM, Rentería MA, Jones RN, Langa KM, et al. Harmonisation of later-life cognitive function across national contexts: results from the Harmonized Cognitive Assessment Protocols. The Lancet Healthy Longevity. 2023 Oct 1;4(10):e573–83.

32. Mukherjee S, Link to external site this link will open in a new window, Choi SE, Lee ML, Scollard P, Trittschuh EH, et al. Cognitive domain harmonization and cocalibration in studies of older adults. Neuropsychology. 2023 May;37(4):409–23.

33. Hall KS, Hendrie HC, Brittain HM, Norton JA, et al. The development of a dementia screening interview in two distinct languages. International Journal of Methods in Psychiatric Research. 1993;3(1):1–28.

34. Ganguli M, Ratcliff G, Chandra V, Sharma S, Gilby J, Pandav R, et al. A hindi version of the MMSE: The development of a cognitive screening instrument for a largely illiterate rural elderly population in india. International Journal of Geriatric Psychiatry. 1995;10(5):367–77.

35. Lee J, Khobragade PY, Banerjee J, Chien S, Angrisani M, Perianayagam A, et al. Design and Methodology of the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia (LASI-DAD). J Am Geriatr Soc. 2020 Aug;68 Suppl 3(Suppl 3):S5–10.

36. Nichols E, Gross AL, Zhang YS, Meijer E, Hayat S, Steptoe A, et al. Considerations for the use of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) in cross-country comparisons of cognitive aging and dementia. Alzheimer's & Dementia [Internet]. [cited 2024 Jun 18];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.13895

37. Khobragade P, Nichols E, Meijer E, Varghese M, Banerjee J, Dey AB, et al. Performance of the Informant Questionnaire on Cognitive Decline for the Elderly (IQCODE) in a nationally representative study in India: the LASI-DAD study. International Psychogeriatrics. 2022 Jul 18;1–11.

38. Livingston SA, Carlson J, Bridgeman B. Test reliability-basic concepts (Research Memorandum No. RM-18-01) [Internet]. Princeton, New Jersey: Educational Testing Service; 2018 [cited 2023 Dec 19]. Available from: https://www.ets.org/Media/Research/pdf/RM-18-01.pdf

39. Angoff WH, Education AC on. Educational Measurement. American Council on Education; 1971. 796 p.

40. Jones RN, Manly JJ, Langa KM, Ryan LH, Levine DA, McCammon R, et al. Factor structure of the Harmonized Cognitive Assessment Protocol neuropsychological battery in the Health and Retirement Study. J Int Neuropsychol Soc. 2024 Jan;30(1):47–55.

41. Mungas D, Reed BR, Kramer JH. Psychometrically matched measures of global cognition, memory, and executive function for assesment of cognitive decline in older persons. Neuropsychology. 2003 Jul;17(3):380–92.

42. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975 Nov;12(3):189–98.

43. Brandt J, Spencer M, Folstein M. The Telephone Interview for Cognitive Status. Neuropsychiatry, Neuropsychology, & Behavioral Neurology. 1988;1(2):111–7.

44. Nichols E, Abd-Allah F, Abdoli A, Abualhasan A, Abu-Gharbieh E, Afshin A, et al. Use of multidimensional item response theory methods for dementia prevalence prediction: an example using the Health and Retirement Survey and the Aging, Demographics, and Memory Study. BMC Med Inform Decis Mak. 2021 Aug 11;21(1):241.

45. Mäcken J, Riley AR, Glymour MM. Cross-national Differences in the Association Between Retirement and Memory Decline. The Journals of Gerontology: Series B. 2021 Mar 1;76(3):620–31.

46. Skirbekk V, Loichinger E, Weber D. Variation in cognitive functioning as a refined approach to comparing aging across countries. Proceedings of the National Academy of Sciences. 2012 Jan 17;109(3):770–4.

47. Yu X, Langa KM, Cho TC, Kobayashi LC. Association of Perceived Job Insecurity With Subsequent Memory Function and Decline Among Adults 55 Years or Older in England and the US, 2006 to 2016. JAMA Network Open. 2022 Apr 13;5(4):e227060.

48. Sutin AR, Luchetti M, Stephan Y, Strickhouser JE, Terracciano A. The association between purpose/meaning in life and verbal fluency and episodic memory: a meta-analysis of >140,000 participants from up to 32 countries. International Psychogeriatrics. 2022 Mar;34(3):263–73.

49. Bonsang E, Skirbekk V, Staudinger UM. As You Sow, So Shall You Reap: Gender-Role Attitudes and Late-Life Cognition. Psychol Sci. 2017 Sep 1;28(9):1201–13.

50. Nichols E, Jones RN, Gross AL, Hayat S, Zaninotto P, Lee J. Development and assessment of analytic methods to improve the measurement of cognition in longitudinal studies of aging through the use of substudies with comprehensive neuropsychological testing. Alzheimer's & Dementia [Internet]. [cited 2024 Aug 30];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.14175

51. Gross AL, Power MC, Albert MS, Deal JA, Gottesman RF, Griswold M, et al. Application of latent variable methods to the study of cognitive decline when tests change over time. Epidemiology. 2015 Nov;26(6):878–87.

52. Harrington D. Confirmatory Factor Analysis. Oxford University Press, USA; 2009. 133 p.

53. Humphreys GW, Duta MD, Montana L, Demeyere N, McCrory C, Rohr J, et al. Cognitive Function in Low-Income and Low-Literacy Settings: Validation of the Tablet-Based Oxford Cognitive Screen in the Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI). The Journals of Gerontology: Series B. 2017 Jan 1;72(1):38–50.

54. Jorm AF. Assessment of cognitive impairment and dementia using informant reports. Clinical Psychology Review. 1996 Jan 1;16(1):51–73.

55. Mackinnon A, Mulligan R. Combining Cognitive Testing and Informant Report to Increase Accuracy in Screening for Dementia. AJP. 1998 Nov;155(11):1529–35.

56. Knafelc R, Lo Giudice D, Harrigan S, Cook R, Flicker L, Mackinnon A, et al. The combination of cognitive testing and an informant questionnaire in screening for dementia. Age and Ageing. 2003 Sep 1;32(5):541–7.

57. Manly JJ, Jones RN, Langa KM, Ryan LH, Levine DA, McCammon R, et al. Estimating the Prevalence of Dementia and Mild Cognitive Impairment in the US: The 2016 Health and Retirement Study Harmonized Cognitive Assessment Protocol Project. JAMA Neurology. 2022 Dec 1;79(12):1242–9.

58. Gross AL, Nichols E, Angrisani M, Ganguli M, Jin H, Khobragade P, et al. Prevalence of DSM-5 mild and major neurocognitive disorder in India: Results from the LASI-DAD [Internet]. medRxiv; 2023 [cited 2024 Feb 8]. p. 2023.09.17.23295674. Available from: https://www.medrxiv.org/content/10.1101/2023.09.17.23295674v1

59. Khobragade PY, Petrosyan S, Weerman A, Hu P, Gross AL, Joyita B, et al. Design and Methodology of the Harmonized Diagnostic Assessment of Dementia for the Longitudinal Aging Study in India: Wave 2. Journal of the American Geriatrics Society. (Under Review.).

60. Chien S, Beaumaster S, Shao K, Meijer E, Angrisani M, Gross A, et al. Harmonized LASI-DAD Documentation [Internet]. Los Angeles: Center for Economic and Social Research; 2022 [cited 2023 Oct 20]. Available from: https://lasi-dad.org/public/documentation/codebooks/Harmonized%20LASI-DAD%20A.3.pdf

61. Chien S, King S, Young C, Liu Y, Meijer E, Gross A, et al. Harmonized ELSA-HCAP Documentation. Gateway to Global Aging Data; 2021.

62. Lin A, Brook J, Grill JD, Teng E. Participant-informant relationships affect quality of life ratings in incipient and clinical Alzheimer's disease. Am J Geriatr Psychiatry. 2017 Mar;25(3):297–307.

63. Loewenstein DA, Argüelles S, Bravo M, Freeman RQ, Argüelles T, Acevedo A, et al. Caregivers' judgments of the functional abilities of the Alzheimer's disease patient: a comparison of proxy reports and objective measures. J Gerontol B Psychol Sci Soc Sci. 2001 Mar;56(2):P78-84.

64. Hackett K, Mis R, Drabick DAG, Giovannetti T. Informant Reporting in Mild Cognitive Impairment: Sources of Discrepancy on the Functional Activities Questionnaire. J Int Neuropsychol Soc. 2020 May;26(5):503–14.

65. Conde-Sala JL, Reñé-Ramírez R, Turró-Garriga O, Gascón-Bayarri J, Juncadella-Puig M, Moreno-Cordón L, et al. Factors associated with the variability in caregiver assessments of the capacities of patients with Alzheimer disease. J Geriatr Psychiatry Neurol. 2013 Jun;26(2):86–94.

66. Smith JR, Gibbons LE, Crane PK, Mungas DM, Glymour MM, Manly JJ, et al. Shifting of Cognitive Assessments Between Face-to-Face and Telephone Administration: Measurement Considerations. The Journals of Gerontology: Series B. 2023 Feb 1;78(2):191–200.

67. Domingue BW, McCammon RJ, West BT, Langa KM, Weir DR, Faul J. The Mode Effect of Web-Based Surveying on the 2018 U.S. Health and Retirement Study Measure of Cognitive Functioning. The Journals of Gerontology: Series B. 2023 Sep 1;78(9):1466–73.

**Table 1.** Concordance table showing the cognitive tests included in at least one wave across the HRS INS studies.

| | HRS | MHAS | ELSA | SHARE | CRELES | KLoSA | JSTAR | TILDA | CHARLS | NICOLA | ELSI | LASI | SPS | MARS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Orientation*** | | | | | | | | | | | | | | |
| Orientation to Time | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Orientation to Space | | | | | X | X | X | X | X | X | | X | X | X |
| ***Memory*** | | | | | | | | | | | | | | |
| Immediate and delayed word recall | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Recall planned action | | | X | | | | | X | | | X | | | |
| Figure recall | | X | | | | | | X | | | | | | |
| ***Language/Fluency*** | | | | | | | | | | | | | | |
| Object naming | X | | X | X | | X | | X | X | X | X | X | | X |
| Animal naming | X | X | X | X | | | | X | X | | X | X | | |
| Writing or reading a sentence | | X | | | | X | | X | X | X | | X | | |
| Repeat a phrase | | | | | | X | | X | X | X | | | | |
| Close your eyes | | | | | | X | | X | X | X | | X | | |
| Country ruler | X | | X | | | | | X | X | | X | | | X |
| 3-stage task | | | | | X | X | | X | X | X | | X | X | |
| Medicine label comprehension | | | X | | | | | X | | | | | | |
| ***Executive functioning*** | | | | | | | | | | | | | | |
| Clock drawing | | | | X | | | | | | | | X | | |
| Serial 7's | X | X | X | X | X | X | X | X | X | X | | X | | X |
| Symbol cancellation | | X | X | | | | | | | | | | | |
| Backwards counting | X | X | X | X | | | | X | | | | X | | X |
| Backwards digit span | | | | | X | | | | | | | | X | |
| Number series | X | | X | | | | | | X | | | X | | |
| Numbers in everyday life | X | | X | X | | | X | | | | | X | | |
| ***Visuospatial*** | | | | | | | | | | | | | | |
| Overlapping pentagons | | | | | | X | | X | X | X | | X | | |
| Overlapping circles | | | | | X | | | | | | | | X | |

*SPS refers to the SPS-ENCAVIDAM sub-survey but is shortened to "SPS" for brevity

**Table 2.** Concordance table showing the cognitive tests included in at least one wave across HCAP studies.

| | HRS-HCAP | ELSA-HCAP | LASI-DAD | Mex-Cog | CHARLS-HCAP | Chile-Cog |
|---|---|---|---|---|---|---|
| ***Orientation*** | | | | | | |
| Day of the week | X | X | X | X | X | X |
| Day of month | X | X | X | X | X | X |
| Month | X | X | X | X | X | X |
| Season | X | X | X | | X | X |
| Year | X | X | X | X | X | X |
| What time is it | | | | X | | |
| What city are we in | X | X | X | | X | |
| What county are we in | X | X | | | X | |
| What province are we in | X | | X | X | X | X |
| What country are we in | | X | | X | | X |
| What floor are we on | X | | X | | X | X |
| What street are we on | | X | | | | X |
| What building are we in | | X | X | X | | |
| What address are we at | X | | X | | X | X |
| ***Memory*** | | | | | | |
| Word list immediate 3 trials | X | X | X | X | X | X |
| Word list delay | X | X | X | X | X | X |
| Word list recognition | X | X | X | X | X | X |
| Logical memory immediate | X | X | X | X | | X |
| Logical memory delay | X | X | X | X | | X |
| Logical memory recognition | X | X | X | | | |
| East Boston memory immediate | X | X | X | X | | X |
| East Boston memory delayed | X | X | X | X | | X |
| Constructional praxis delay | X | X | X | X | | X |
| 3-Word recall immediate | X | X | X | X | X | X |
| 3-Word recall delay | X | X | X | X | X | X |
| ***Executive Functioning*** | | | | | | |
| Symbol/letter cancellation | X | X | X | X | | X |
| Symbol digit modalities test | X | X | | | | |
| Symbols and digits test | | | | X | | X |
| Serial threes | | | | X | | |
| Serial sevens | | X | X | X | X | X |
| Spelling backwards | X | X | | | | |
| Backward day naming | | | X | | | |
| Backward counting | X | X | | | | X |
| Backwards numbers | | | | | | X |
| Backward counting from 20 | | | | X | | |
| Number series | X | X | | | X | |
| Digit span forward | | | X | | | |
| Digit span backward | | | X | | | |
| Trail-making test part A | X | X | X | | | |
| Trail-making test part B | X | X | X | | | |
| Raven's progressive matrices | X | X | X | | | |
| Go-No-Go | | | X | X | | X |
| Token test | | | X | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Similarities | | | X | X | | X |
| Problem solving | | | X | | | |
| ***Language*** | | | | | | |
| Animal fluency | X | X | X | X | X | X |
| Name cactus | X | X | | | X | |
| Name coconut | | | X | | | |
| Name tree | | | X | | | |
| Name scissors | X | X | X | X | X | X |
| Name prime minister | X | X | X | | | |
| Name deputy president | | | | | | |
| Name elbow | X | X | X | X | | X |
| Do with a hammer | X | X | X | X | X | X |
| Following instructions | X | X | X | X | | |
| Where is the local market? | X | X | X | X | X | X |
| Define bridge | | | | X | | X |
| Name watch | X | X | X | | X | X |
| Name pencil | X | X | X | X | X | X |
| Name shoe | | | | X | | |
| Write/say a sentence | X | X | X | X | X | X |
| Point to things | X | X | X | X | X | X |
| Read and follow command | X | X | X | X | X | X |
| Repetition of phrase | X | X | X | X | X | X |
| Following instructions 3-step | X | X | X | X | | X |
| ***Visuospatial Functioning*** | | | | | | |
| Constructional praxis | X | X | X | X | | X |
| Copy polygons | X | X | X | X | | X |
| Clock drawing | | | X | | | |