



HCAP Network

Harmonized Cognitive Assessment Protocol

HCAP Protocol Guidance

Lindsay Kobayashi, University of Michigan
Sarah Kwiatek, University of Michigan
Miguel Arce Rentería, Columbia University
Emily Briceño, University of Michigan
Tsai-Chin Cho, University of Michigan
Alden Gross, Johns Hopkins University
Emma Nichols, University of Southern California
Lindsay Ryan, University of Michigan
David Weir, University of Michigan
Ken Langa, University of Michigan

Survey Research Center
Institute for Social Research
University of Michigan
Ann Arbor, Michigan

October 2026

Funding: The HCAP Network is supported by the National Institute on Aging through the Division of Social and Behavioral Research (NIA U24AG065182).

Suggested Citation: Kobayashi, L. C., Kwiatek, S.M., Arce Rentería, M., Briceño, E. M., Cho, T.C., Gross, A. L., Nichols, E., Ryan, L. R., Weir, D. R., & Langa, K. M. (2026). *HCAP Protocol Guidance*. University of Michigan.

Acknowledgements: We thank Amanda Sonnega and Eva Giatas for their support in preparing this document.

Table of Contents:	
Introduction	5
Cognitive Domains in the HCAP Battery	6
Overview of the HCAP	7
Principles of cross-cultural HCAP adaptation	10
Overview of HCAP Cognitive Tests	12
Memory	12
10-word recall (immediate, delayed, and recognition)	12
3-word recall (immediate and delayed)	14
Story recall (long story and short story; immediate, delayed, and recognition)	14
Constructional praxis (delayed)	15
Executive Function	16
Backward counting	17
Cancellation test (letters or symbols)	18
Serial Subtractions, Backward Spelling, Backward Day Naming	19
Symbols and Digits	20
Trail Making Test Parts A & B	20
Similarities & differences	21
Judgement and problem solving	21
Everyday calculations	22
Digit span (forward and backward)	22
Hand movement sequencing	22
Go/No-Go	22
Token test	23
HRS number series	23
Raven's standard progressive matrices	24
Spatial working memory (forward and backward)	24
Language	24
Brief naming items	25
Boston Naming Test	25
Language comprehension items	26
Following commands	26
Local market	26

Animal naming	26
Follow a multi-step command	27
Follow a 1-step command	27
Language repetition	28
Write a sentence	28
Phonemic fluency	29
Visuospatial	29
Constructional praxis (copying)	30
Interlocking Shapes	30
Clock drawing	31
Orientation	31
Place	31
Time	32
Name political leader	32
Missing data coding	34
Recommended further reading	35
Cohort profile papers	36
Citations	37

Introduction

The HCAP Toolbox provides information and guidance for study teams collecting primary data using the Harmonized Cognitive Assessment Protocol (HCAP). The HCAP Toolbox consists of this protocol guidance document and an accompanying item repository. The item repository summarizes the cognitive test items included in HCAP studies across the network that have completed data collection. Together, these materials are intended to support study team decisions regarding the selection, adaptation, and administration of tests by providing clear information and guidance for these processes based on the field experiences of existing HCAP studies. The long-term goal of the HCAP Toolbox, as a shared resource produced by the HCAP Network, is to provide efficient support for high-quality and harmonized HCAP data collection across diverse global settings. The HCAP Network is funded by the US National Institute on Aging at the US National Institutes of Health (U24 AG065182, PI: Langa, Kobayashi).

The HCAP Toolbox is not meant to replace source material for cognitive tests. We do not recommend administering cognitive tests based solely on the information presented here. The source materials provide details that are needed for administration and scoring that are not included here (e.g., interviewer instructions, stimuli that accompany the tests). In addition, careful consideration of the study population's educational, linguistic, social, and cultural background is needed when deciding which cognitive tests to include, and whether and how they should be adapted, administered, and scored. The HCAP Toolbox is thus meant to provide a comprehensive summary of the cognitive tests used in HCAP studies around the world and how they have been adapted in their administration, scoring, and coding of responses. Topics covered in this protocol guidance include the following: 1) an overview of the HCAP, 2) key principles of cross-cultural adaptation of cognitive test items in the HCAP, 3) rationale for each of the five cognitive domains measured in the HCAP, 4) an overview of all items that have been included in ongoing HCAP studies that have publicly available data and protocols, with redactions for copyright purposes where appropriate, along with recommendations from the HCAP Network about adaptations and administration 5) recommendations for harmonized missing data coding, and 6) further recommended readings.

We envision this HCAP Toolbox as a living document that will be updated as we continue to learn from ongoing and newer HCAP studies as they are fielded, and from findings regarding reliability, validity, and statistical co-calibration of collected HCAP data. Information about HCAP studies will be added to the toolbox as their data and protocols become publicly available.

Non-cognitive test item content (e.g., general health, subjective cognitive functioning, depression screener) and informant interview measures are not addressed in these materials.

Cognitive Domains in the HCAP Battery

All HCAP studies include both direct cognitive tests and informant-based assessments of functional independence. This dual approach enables evaluation of cognitive performance alongside everyday functioning, particularly activities of daily living. HCAP studies can implement rules to determine when a respondent has completed enough of the direct cognitive testing to be included as a “partial” interview. For example, the HRS-HCAP includes all cases that at least complete the MMSE, the first test in their respondent protocol. Based on neuropsychological theory and prior psychometric work, the cognitive factor structure of the HCAP tests falls into five cognitive domains: memory, executive functioning/attention, language, visuospatial functioning, and orientation. These five cognitive domains are relevant to later-life cognitive health and are impacted by various neurodegenerative diseases such as Alzheimer’s disease and related dementias (ADRD). Additional information on the clinical relevance of each cognitive domain can be found below in subsequent sections. The inclusion of at least five cognitive domains in the HCAP facilitates the cross-national harmonization of cognitive factor scores and diagnostic approaches to MCI and dementia. Although prior HCAP studies have developed algorithmic approaches with these five domains (Manly et al., 2022, Mejia-Arango et al., 2026), other analyses have used only four domains (A. Gross et al., 2023). Not all cognitive domains have been able to be harmonized across available HCAP studies around the world (e.g., see Visuospatial, below). As such, investigators should balance the specific needs of their target population and those of harmonization with other HCAP around the world.

This five-domain framework is intended as an organizing structure rather than a definitive taxonomy of cognitive domains. It reflects a psychometrically valid approach to structuring HCAP content in a manner that supports cross-study and cross-national harmonization using existing instruments. Although these domains capture broad areas central to dementia classification and comparison across settings, they do not exhaust the range of cognitive abilities that may be relevant in all populations. For example, attention is conceptually a separate domain from executive functioning, yet measures capturing attention are largely grouped with executive functioning in the HCAP. More narrowly defined constructs, such as set-shifting or processing speed, are also often subsumed within executive functioning. Accordingly, the five-domain structure of the HCAP should not necessarily be interpreted as sufficient or required for comprehensive cognitive assessment in every context. Differential emphasis may be warranted depending on cultural norms, educational exposures, and expectations for everyday cognitive performance. The domain structure thus serves as a harmonization framework that supports the application of dementia classification algorithms, while allowing flexibility for context-specific adaptation.

Overview of the HCAP

Key characteristics of HCAP studies are summarized in Table 1. There is a need to administer cognitive tests in a way that supports both cross-national comparability and local validity. The HCAP is not a rigid checklist of identical cognitive tests that are administered the same way around the world, but instead a flexible measurement architecture built around five core cognitive domains with a commitment to cross-national harmonization. As articulated by Gross and colleagues (Gross et al., 2026), the cognitive components of an HCAP battery can be organized around the five cognitive domains of orientation, memory, executive functioning, language, and visuospatial functioning. We recommend that each domain include at least three to four test items with strong psychometric properties, such as polytomous or continuous items with precision of measurement along the full distribution of ability, that can serve as linking items across studies. A linking item is a cognitive test item that is thought to perform with similar difficulty and discrimination across HCAP study populations. A good example of a linking item is animal naming, which is administered similarly without adaptations (aside from linguistic translation) and believed to perform similarly across HCAP studies. The presence of linking items across HCAP studies enables what we refer to as “statistical co-calibration” or “statistical harmonization”, using modern psychometric methods to place items on a common measurement scale, preserving cross-national comparability while allowing measurement precision across the full range of cognitive ability. Regarding nomenclature, we refer to “cognitive tests” and “cognitive test items” separately because some cognitive tests can yield more than one item (e.g., the Trail Making Test consists of Part A and Part B, which are treated as separate items during statistical co-calibration).

Table 1. Considerations and recommended characteristics of an HCAP study

Consideration	Characteristics
What are the goals of an HCAP study?	To generate comparable, population-based estimates of dementia prevalence and incidence, and to reliably measure continuously distributed cognitive function and decline, all while accounting for contextual factors such as education, language, literacy, and socioeconomic background.
How standardized should an HCAP study be?	Include a core set of three to four continuously distributed test items per domain that can ideally serve as linking items for cross-national harmonization. Supplement the HCAP battery with additional tests that reflect the cultural, linguistic, or research needs of their populations.

What tests should constitute a core or fixed set for an HCAP battery?	Assess five core cognitive domains: memory, executive function, language, visuospatial ability, and orientation. Specific test recommendations are made in the text. Cognitive test items should have good psychometric properties (e.g., reliability and precision across the range of the distribution of ability), and validity for measuring cognition in a local context. Prioritize either continuous items or ordinal items with many categories. Linking items: include 3-4 per domain to facilitate cross-national harmonization. Include an informant (family or friend) interview for all participants. This is particularly important for participants too impaired for direct cognitive testing.
How long should an HCAP battery be?	Testing should take on average at least 60 minutes, if not longer.
What setting is needed for a study to be considered an HCAP study?	Sample participants in such a way as to facilitate the ability to draw inferences to some defined, desired target population. Information on risk and protective factors, demographics, and other contextual factors (literacy, education, etc.), either from the HRS family parent study or from a shortened interview administered alongside HCAP.
Consistent administration, scoring, and missing data coding of HCAP tests	Key differences between existing HCAP studies -- in test administration, scoring of responses, and coding of missing data -- have been identified, which threaten cross-national harmonization efforts. Care should be taken to coordinate these issues centrally via the HCAP network.

Below, we list “core” cognitive tests for each domain that we recommend including within an HCAP instrument. While we recognize that not all of these items will be appropriate for inclusion in every HCAP study, they all have strong psychometric properties, are often continuously distributed, and have demonstrated feasibility of administration across diverse study contexts. They are thus more likely to meet criteria as high-quality linking items across HCAP studies and, when administered with appropriate adaptations, should provide locally valid measures of cognition, while enhancing harmonization across the network:

- **Memory:** multi-trial word list learning with immediate and delayed recall (e.g., CERAD word list), two story recall tasks (one shorter and one longer) to assess immediate and delayed recall of contextual information, visuoconstructional recall.
- **Executive functioning:** Trail Making Test (with letters/numbers or circles/squares), the Go/No-Go task, digit span, letter or symbol cancellation, reasoning measures (e.g., Raven's Progressive Matrices), and judgment/problem-solving items.

- **Language:** animal naming as a measure of semantic fluency, supplemented by confrontation naming, comprehension (e.g., Token Test or multi-step commands), and functional language items.
- **Visuospatial functioning:** constructional praxis or interlocking pentagons, with literacy adaptations as needed, such as matchstick copying.
- **Orientation:** orientation items to time and place are included to capture severe cognitive impairment, although they have limited discrimination at levels of functioning seen in community-dwelling samples (i.e., ceiling effects).

Across domains, we recommend giving priority to items that discriminate well across a broad range of ability levels, thereby strengthening cross-study linking.

Crucially, the HCAP is a fixed-flexible, adaptive architecture rather than a fixed battery. After including three to four linking items per cognitive domain, we encourage HCAP studies to add supplementary cognitive tests tailored to cultural, linguistic, educational, or research priorities (e.g., whether to accommodate low literacy, sensory limitations, local semantic categories, or specific research hypotheses). Decisions about additional items should consider validity in the local study context, feasibility of administration, administration time (the full battery should average at least ~60 minutes), and improving the overlap in content with other HCAP studies. All HCAP studies should also include an informant interview to capture functional impairment and support algorithmic dementia classification, particularly for participants unable to complete direct cognitive testing. Thus, well-designed HCAP studies should integrate (1) cognitive tests covering the five core domains; (2) multiple linking items per cognitive domain to facilitate cross-national harmonization; (3) careful documentation of administration, scoring, coding of tests, and interviewer training protocols; and (4) culturally responsive supplementation of additional cognitive tests within a harmonizable framework.

Principles of cross-cultural HCAP adaptation

The HCAP endeavors to optimize cross-national comparability in the assessment of cognitive function across older adult populations around the world. However, the implementation of an identical cognitive assessment protocol cross-nationally is not feasible or appropriate. The assessment of cognition requires that cognitive test items are accessible and appropriate for local settings. Linguistic, cultural, geographical, social, and educational exposures can influence the accessibility or appropriateness of a cognitive test item within a given setting. Adaptations may include substitution of culturally unfamiliar stimuli, modifications of administration instructions, and adjustments for literacy or numeracy differences. As such, careful adaptation of HCAP should be undertaken in each study to create a cognitive assessment protocol that is both locally valid and optimizes comparability with other studies in the HCAP network.

Balancing the needs of local validity and cross-national comparability often requires negotiation. The validity of a cognitive test item in the local setting should remain the highest priority. Test items that are not locally appropriate are not useful for the measurement of cognition, either within a specific study or for cross-national harmonization. As such, we strongly emphasize the critical importance of prioritizing construct validity in each setting by ensuring the adapted items tap the same cognitive domain at a comparable task difficulty level. In this document, we include a description of the cognitive construct that is intended to be measured by each cognitive test item in the HCAP.

We offer comprehensive recommendations for HCAP adaptation methodology elsewhere (Briceño et al., 2026). This work should be driven by a multidisciplinary team with expertise in the local language(s), in addition to the social, cultural, and geographical contexts represented in a study. The HCAP Toolbox includes details about various adaptations of items contributed by HCAP study teams, which may be useful starting points for some HCAP study teams. Here, we highlight several key components of the adaptation process:

- Prior to pre-testing, all necessary permissions and licenses for tests should be procured.
- All HCAP items, including items adapted by other teams, should undergo careful and thorough pre-testing before implementation in the field. When possible, we strongly recommend including a neuropsychologist with familiarity with the country context in which the HCAP will be administered. This expertise will help ensure that the underlying neuropsychological construct intended to be measured by a test item is retained in the course of adaptation and translation.
- The creation of clear and detailed administration and scoring guidance for all HCAP items, especially newly adapted HCAP items, is important. Recommended scoring procedures for some HCAP items are included in this document. For HCAP items that

are licensed/copyrighted, study teams are encouraged to obtain the publisher's standard scoring criteria after securing appropriate permissions.

- For each HCAP study, detailed administration and scoring guidance for each HCAP item should be integrated into an interviewer manual. This interviewer manual can then serve to facilitate consistent interviewer training, quality assurance, and subsequent cross-national harmonization.
- Rigorous interviewer training and certification are also critical for appropriate collection of HCAP data. The training should be led by an individual with expertise in cognitive test administration. Detailed recommendations for interviewer training are available elsewhere (Briceño et al., 2026).
- Throughout the data collection period, interview quality assurance procedures are strongly recommended. If feasible, we strongly recommend audio recording the HCAP interviews, with photographs of interview responses provided visually (e.g., respondent drawings). Periodic observation of interviews is also useful when feasible.

Overview of HCAP Cognitive Tests

This section provides narrative insight that is not captured in the repository spreadsheet. Within each subsection for each cognitive domain, we define the cognitive domain and then list cognitive tests within that domain. For each test, we provide details on the history and purpose of adaptations, field experiences with the test, and lessons about test performance. We provide notes about administration and scoring that are relevant for improving harmonization across HCAP studies, especially when there are differing precedents from different studies in the network.

Where relevant, copyrighted test licensing information used by the HRS-HCAP is provided in this document, with additional details in the repository spreadsheet. As this document is organized by cognitive domains and individual test items, copyrighted assessments, like the MMSE, will be referenced in multiple places although only one license is needed to administer the overall MMSE. Due to copyright restrictions, some details about test item protocols are censored from this document and within the accompanying item repository spreadsheet.

Memory

Memory abilities are impacted across most neurodegenerative diseases and neurological conditions. Broadly speaking, most tests of memory assess abilities such as encoding, retrieval, and long-term storage. For instance, Alzheimer's disease affects encoding and storage aspects of memory abilities, while other diseases may target other aspects. The following tests measure both verbal and visual aspects of memory functioning. Verbal tests can also be considered as measuring learning of non-contextualized information (i.e., list learning) and contextualized information (i.e., story recall). Measurement of these diverse aspects of memory functioning leads to a robust, content-valid summary measure of this key cognitive domain.

10-word recall (immediate, delayed, and recognition)

This test is intended to measure the ability to learn and recall new information. A list of 10 familiar, high-imagery nouns is presented visually and verbally at the rate of 1 word every 2 seconds. There are three learning trials, in which the respondent is presented with the 10-word list and then given up to 90 seconds to recall as many words as possible. On each of the 2 subsequent trials, the 10 words are presented in a new random order. Following each trial, the subject tries to recall as many words as they can, in whatever order they can. The maximum number of correct responses is 30 across the 3 trials.

During the standard CERAD administration, the subject reads each word aloud as it is presented, and is then tested on immediate recall after each trial and ~5 minutes later on a delayed recall procedure followed immediately by a recognition task. The recognition task consists of presenting the participant with a list of 20 words, which includes all the words in the

first trials plus new words not included in the list. The respondent is asked whether each of the 20 words was on the original list.

The largest differences in administration across studies stem from the format of the word list presentation. In high-income countries (HICs), which tend to have high literacy rates, such as in the HRS, ELSA, and SHARE HCAPs, the words were presented visually, and respondents were asked to read each word aloud as it was shown to them. However, in many low- and middle-income countries (LMICs), such as in Mex-Cog, HAALSI-HCAP, and LASI-DAD, the words were presented in auditory format only. In HAALSI-HCAP, the word list was pre-recorded, and the interviewer played the audio recording to the respondent. Studies differ with regard to whether the respondent is asked to repeat each word as it is presented to them, or whether they listen passively as the list is read to them. Studies also differ with regard to whether the word list is presented in a random or consistent order across immediate recall trials. These differences lead to this item not being considered a linking item across these two types of administrations in statistical co-calibration efforts.

Studies also vary in the cognitive tests used to space out the immediate, delayed, and recognition trials. For example, in the SHARE-HCAP, animal naming, symbol cancellation, backward counting, and 10/66 naming items (hammer, elbow, shop, point) are between immediate and delayed recall, and immediate story recall (short and long story) is between delayed recall and recognition. In LASI-DAD, by contrast, digit span forward and backward as well as symbol cancellation are between immediate and delayed recall, whereas there were no tests between delayed word recall and recognition. The standard administration of the CERAD does not include any tests between delayed recall and recognition trials, but this may be necessary given the delay requirements of other tests included in the battery.

Other areas of difference between studies include the specific language translations of the words used in the word lists. We stress that word lists should NOT necessarily be directly translated, but rather when translating the words to new languages, words with similar frequency, valence, and syllabic length should be selected.

We recommend that:

- Translations should not be direct translations but rather based on the frequency and valence of the included words.
- Auditory presentation rather than dual auditory and visual presentation of the word list to ensure consistency of administration and thus harmonization across HICs and LMICs settings.
- For ongoing HCAP studies with longitudinal administration that previously administered both auditory and visual presentations of the word list, switching to auditory-only would improve cross-sectional comparability with other studies in the network that administered auditory-only word lists, while retaining auditory and visual presentations will maintain cross-wave comparability within their own study. This choice is up to investigators.

- In line with the standard administration of the CERAD, respondents should repeat words if auditory presentation is provided, and randomized word orders should be used for the second and third trials of immediate recall.
- Build into the interview a 5-minute delay between immediate and delayed recall trials. Preferably, there should not be additional verbal tests in between immediate and delayed recall portions of the test, but this should be balanced with the demands and logistics of each individual study. Please consult further with your team's neuropsychologist.

The tasks are part of the CERAD. See here for more information and to request permission to use the measure: <https://www.naccddata.org/request-cerad-forms/>

3-word recall (immediate and delayed)

This task assesses short-term verbal memory, including immediate registration and recall following a brief delay. If respondents do not correctly complete the immediate recall task, the interviewer can repeat the task up to five times (the maximum number of trials varies by study). We anticipate that a maximum of 3 recall trials is appropriate for most studies, although pretesting is recommended to determine an appropriate number of immediate recall trials. Respondents are tested in an immediate recall and later on a delayed recall procedure.

Scoring and detailed procedures are copyrighted as a component of the Mini-Mental State Examination (MMSE) (Folstein et al., 1975), a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. See here for more copyright information: <https://www.parinc.com/products/MMSE>. Some HCAP studies use variations of this task, with some being licensed and others not. It is up to investigators to decide which version is most appropriate for their study population.

Story recall (long story and short story; immediate, delayed, and recognition)

These tests measure immediate and delayed memory for stories of different lengths. Interviewers read a story aloud to the respondent. Then, respondents are asked to immediately recall as many details of the story as they can remember. Respondents may repeat story elements verbatim (exact recall), or make statements that express the general idea about a story element (gist or approximate recall). In many HCAP studies, a recognition trial of the long story is also included. The respondent is asked a series of yes/no questions about the long story to assess their recognition memory for the story.

It is important for each HCAP study to ensure that the story is adapted – usually in relatively small ways – to be culturally and linguistically appropriate. For example, studies commonly adapt the specific details about the stories, such as the names and locations used in the story.

Across studies, available types of scoring can vary; for example, in Wave 1 of HRS-HCAP, only exact answers were recorded for the long story, but for Wave 2 the scoring protocol was

updated to facilitate scoring of exact and gist recall. More commonly, both exact and approximate (e.g., gist) story points that respondents are able to recall are recorded. Gist scoring may be more ecologically valid in low-literacy settings, but raises challenges around variability in scoring rules.

We recommend that:

- Extensive training and practice by interviewers is needed to accurately record the responses.
- Studies should create standardized guidelines for exact and gist scoring criteria for each story element.
- Cultural adaptations should be used as necessary to adjust story elements (e.g., name, city); these adaptations do not impact comparability across studies as long as the total number of points allowed is unchanged.
- Regarding scoring for the short story, the standard administration criteria for the short story include 6 scored story elements (that is, 6 points for exact recall or 12 points for gist recall), but some HCAP studies use a 10-point system. We recommend using the 10-point system to allow for greater variability in scores. If using a 10-point rubric, a 6-point version should also be included for cross-study comparability.
- Regarding scoring for the long story, ensure that there are 25 story elements.
- If allowed by your IRB, save audio recordings to allow re-scoring if needed. Consider centralized scoring to optimize reliability.

The long story task used by HRS-HCAP comes from the Wechsler Logical Memory Scale - Fourth Edition (WMS-IV) (Wechsler, 2009), and copyrighted material information can be found here: <https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Memory-Scale-%7C-Fifth-Edition/p/P100096000>.

Constructional praxis (delayed)

This test measures delayed recall of visually presented shapes that were previously presented on paper or a tablet screen. Respondents are asked to recall the geometric shapes drawn in the test of CERAD Constructional Praxis – Copy. The constructional praxis test in CERAD is adapted from Rosen’s assessment of constructional praxis (Rosen et al., 1984), and tests the ability of the respondent to copy four geometric forms of varying difficulty (circle, diamond, two overlapping rectangles, and 3D cube).

This test assumes prior exposure to drawing, familiarity with the use of writing utensils, and familiarity with the geometric shapes included in the test. We recommend adapting this test for populations that may not have such universal prior exposure. For example, HAALSI-HCAP (in their Wave 2) and CVFS-SCAN included a matchstick copy and recall test that does not have any drawing demands or assumptions about familiarity with specific geometric figures. In KLPS, one of the shapes was replaced with a triangle since the original shape is not commonly taught in Kenyan schools.

Mode of data collection also requires attention; for example, LASI-DAD used paper and pencil for Wave 1 and a tablet for Wave 2. In LASI-DAD, respondents found the tablet easier to use than paper and pencil. They were offered the option of using either a stylus or their finger to draw on the tablet screen, according to their personal preference.

Studies also vary in the number and selection of tests to include between the immediate copy and delayed trials. The original CERAD administration specifies a 5-minute delay between immediate and delayed tests.

We recommend:

- Consider adaptations to the administration or mode as necessary. In settings with varying educational backgrounds, it is advantageous to use adaptations that have already been fielded in other HCAP studies (e.g., matchstick copy and recall) or modes of administration that have worked well in other HCAP studies (e.g., drawing with a stick, drawing with paper and pencil, or drawing with a finger on a tablet). However, additional work is required to determine whether there are systematic differences in performance by modality (i.e., using a pencil vs. a finger).

See here for more information and to request permission to use the measure:

<https://www.naccddata.org/request-cerad-forms/>

Executive Function

The following section describes various tests that measure either attention, processing speed, or other aspects of executive functioning (i.e., working memory, abstract reasoning, response inhibition, problem-solving, set-shifting). Various neurodegenerative diseases affect these cognitive abilities. For instance, frontotemporal lobe dementia pathologies can impact executive functioning early in the disease process. Similarly, cerebrovascular disease has been associated with impairments in these abilities given its effect on white matter integrity. Alzheimer's disease, traditionally speaking, may affect these cognitive abilities later on in the disease process; however, atypical presentations of Alzheimer's disease have been associated with executive dysfunction early on. As such, to have a better characterization of an individual's overall cognitive health, we recommend including measures of executive function in your HCAP battery.

Prior psychometric work across ongoing HCAP studies has demonstrated that most of these tests can fit under one latent cognitive domain when determining the battery's cognitive factor structure through confirmatory factor analysis (Gross et al., 2023, Jones et al., 2024, Mani et al., 2025). Some neuropsychological batteries other than the HCAP may separate these tests into different independent domains, but this is largely dependent on the number of tests and study sample characteristics, among other factors. As mentioned previously, we recommend that future HCAP studies establish the cognitive factor structure for their own HCAP battery to confirm whether these tests similarly best fall under one latent cognitive domain or not.

There is substantial variability across HCAP studies with respect to which of the following tests are included for executive function, largely because these types of tests frequently rely on literacy or numeracy skills, or involve cultural assumptions. We recommend reviewing the item repository to identify which tests are included across HCAP studies and get a better sense of which tests may be most appropriate for your HCAP study population.

Backward counting

The backward counting task (Lachman et al., 2013) is a measure of speeded verbal working memory. The respondent is asked to count backward from 100 as fast as possible within a 30-second time limit. The interviewer records the number they get to in the time allowed and the number of mistakes made. The score is calculated by subtracting from 100 the final number the respondent reached and the number of errors made, if any.

This test assumes a high numeracy level, such that numeric sequencing to 100 is familiar and over-learned. The test also assumes familiarity with the concept of “backward.”

This test can be a challenge for the interviewer to score, as responses may come quickly and classification of errors can be challenging. Several studies find audio recordings are helpful for reviewing and cleaning this test. Study teams interested in including this test are encouraged to contact the HRS-HCAP team for standardized scoring instructions.

Some studies among populations with varying levels of numeracy utilize a different version of this test that requires counting backward from a lower number (e.g., Mex-Cog, HAALSI-HCAP). Responses are scored as correct if the respondent makes no errors when counting back 10 consecutive numbers. The time taken to reach a given end number is also recorded, and a derived score is created that incorporates both accuracy and speed. However, this version of the test assumes at least moderate numeracy and familiarity with the concept of “backward.” This adaptation is similar to an item from the copyrighted TICS battery; see the following link for details and licensing information: <https://www.parinc.com/products/TICS>

We recommend:

- The version of backward counting from a lower number is preferred over the version of backward counting from 100. The version of backward counting from a lower number has fewer numeracy assumptions and is easier to administer and score than the version of backward counting from 100.
- If harmonization with HRS-HCAP, ELSA-HCAP, and SHARE-HCAP is more important than ease of scoring and reduced numeracy requirements (as in backward counting from a lower number), then backward counting from 100 may be preferred.
- Consider omitting this item if a high proportion of the study population has low numeracy and/or is not familiar with the concept of “backward.”

Cancellation test (letters or symbols)

This test assesses visual sustained attention and processing speed.

For the Letter Cancellation test, respondents are given an 8x11 inch paper with a large grid of letters and are asked to scan the grid as quickly as possible and to cross out as many “P” and “W” letters as they can within a 60-second time limit. Information recorded includes the last letter an individual arrived at, the number of correctly crossed-out letters, the number of missed letters, and the number of incorrectly crossed-out letters. Respondents are told to move from left to right, top to bottom on the page and to not skip around on the page. This test is administered with paper and pencil. This test was originally developed for the 1946 British Birth Cohort study (Richards et al., 1999).

The Letter Cancellation test was developed in an English-speaking population and assumes prior exposure to reading, including speeded, left-to-right visual scanning and familiarity with letters. The test also assumes familiarity with the use of writing utensils. This test has been adapted for populations with varying levels of literacy with a Symbol Cancellation test (e.g., Mex-Cog; LASI-DAD), such that respondents are asked to circle target symbols among an array of various symbols that are not arranged in rows. The Symbol Cancellation test thus does not rely on left-to-right visual scanning that is overlearned through reading, yet does assume familiarity with drawing circles.

This test has been applied using a paper/pencil paradigm (e.g., Mex-Cog; LASI-DAD Wave 1) and on a tablet (e.g., LASI-DAD Wave 2). In LASI-DAD Wave 2, the symbols were rearranged into a series of rows, and respondents were permitted to use a stylus or their finger to circle target symbols. Also, in this version, the target symbol does not remain visible during the testing period, which results in some memory demands for this version.

For populations that do not have familiarity with or exposure to drawing circles, adaptation may be needed. For example, the CVFS-SCAN study in Nepal created a novel paradigm requiring respondents to collect target beans amongst an array of different types of beans in a pan.

We recommend:

- In literate populations that use the Latin alphabet with left-to-right written language systems, Letter Cancellation should be appropriate.
- Consider using Symbol Cancellation, especially in populations with varying literacy levels. If drawing circles around a target symbol using a tablet or paper and pencil is anticipated to be challenging (this can be examined during pre-testing), then bean cancellation as in CVFS-SCAN may be an appropriate adaptation, minding the additional testing materials and time required.

Serial Subtractions, Backward Spelling, Backward Day Naming

This group of tests is intended to measure verbal working memory. They are grouped together here because of their neuropsychological similarity. Most HCAP studies include one of the following three tests.

Serial Subtractions: For this test, respondents are asked to do five consecutive subtraction trials. This test assumes numeracy skills and is a component of the MMSE (Folstein et al., 1975). This test has been adapted for populations with varying levels of numeracy. For example, the Serial 3's test (e.g., Mex-Cog) asks respondents to subtract 3 from 20 for a total of five trials. Each subtraction is scored separately. The Serial 3's test has been further adapted to accommodate populations with low numeracy by providing a story to contextualize the task (e.g., CVFS-SCAN). Specifically, respondents are asked to perform serial subtractions within the context of a story about goats that are sold 3 at a time.

In HRS-HCAP, the task is a component of the MMSE (Folstein et al., 1975), a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. See here for more copyright information: <https://www.parinc.com/products/MMSE>

Backward Spelling: Spelling a word backward is used to measure verbal working memory in the HRS-HCAP, such that respondents are asked to spell a familiar word backward. This item assumes literacy and assumes that the respondent has high familiarity with the spelling of the word. We recommend that interviewers record the letters that respondents say back and score the test after the interview, to ensure a smooth transition to the next test and reduce the likelihood of errors due to on-the-spot scoring.

We recommend:

- If translating this item into a different language than the source material, select a similarly familiar word in the target language that is monosyllabic, the same number of letters as the English version, and which has an unusual combination of consonants together; otherwise, the item may be too easy.

In HRS-HCAP, this task is a component of the Mini-Mental State Examination (MMSE) (Folstein et al., 1975), a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. See here for more copyright information: <https://www.parinc.com/products/MMSE>

Backward Day Naming: Backward day naming was designed as an adaptation of the backward counting and spelling items for populations with varying levels of literacy and numeracy. Respondents are instructed to name the days of the week in order (forward naming; Monday → Sunday) and then in reverse order (backward naming; Sunday → Monday). The day naming item does not assume literacy or numeracy. It does assume familiarity with the sequence of the days of the week and with the concept of “backward” or reverse sequencing.

Symbols and Digits

This group of tests is designed to measure processing speed primarily, as well as sustained visual attention and visual scanning.

Symbol Digit Modalities Test (SDMT): The SDMT is administered in HRS-HCAP and ELSA-HCAP. This test is administered with the official SDMT paper form. It requires respondents to substitute a number corresponding to a series of geometric figures. A printed key is provided, which pairs the numbers 1 through 9 with a specific symbol so that each number has its own unique symbol. The respondent writes as many numbers corresponding to the geometric figures as possible in 90 seconds (see (Smith, 1982)).

This test assumes familiarity with Arabic numerals, left-to-right visual scanning (overlearned through reading), and the use of a writing utensil. We recommend omitting this test for study populations for whom these assumptions do not universally apply.

Information on ordering the official test forms can be found here:

<https://www.parinc.com/products/SDMT>

Symbols and Digits: The Symbols and Digits test is administered in Mex-Cog and Chile-Cog. This test uses different visual stimuli, but the main components are similar to the SDMT.

Both tests assume familiarity with Arabic numerals and the use of paper/pencil to draw simple geometric figures.

Trail Making Test Parts A & B

This test is designed to measure speeded visual attention and cognitive flexibility. The traditional version of this test has two parts. The first part requires respondents to quickly connect numbered circles on a worksheet in numeric order (Part A). The second part requires respondents to quickly connect circles with numbers and letters, alternating between numbers and letters in order (Part B).

See Reitan (1992) for procedure and scoring. It is important for the interviewer to be in a position where they can easily watch the respondent complete the task, as error corrections in real-time are an important component of the administration.

The traditional version of this test that uses numbers and letters assumes numeracy and high familiarity with English alphabetic order. This test has been adapted for populations with varying literacy and numeracy. In the adapted version, respondents are asked to connect differently sized circles in order of increasing size (Part A), then connect differently sized squares in order of increasing size (Part B). The interviewer is instructed to point out errors to the respondent as they occur, and have the respondent go back to the previous circle and move on to the next correct one. The mode of assessment (paper/pencil vs. tablet) is another potential difference across studies.

The number of seconds to complete the task is the primary score derived from these assessments.

We recommend:

- Using the adapted version that uses circles and squares in populations with varying literacy levels. Versions using non-English alphabets have not been implemented in HCAP studies, but may be available. For example, see (Dubois et al., 2000).
- As with other items, the choice of mode of assessment (paper/pencil vs. tablet) should depend on appropriateness for the study population, cost, or other feasibility concerns regarding administration.

Official copyrighted test forms for the numbers and letters version of this test can be purchased here: <https://neuropsych.com/product/trail-making-test-adults/>

Similarities & differences

Questions on similarities & differences measure abstract reasoning by asking what the key similarity or difference is between two words. For example, if asked about the difference between a river and a pond, the correct answer would be that a river is flowing and a pond is not flowing. The level of abstraction required to answer the question can be used to modulate the difficulty of the item (e.g., the difference between a lie and a mistake tends to be more difficult as these concepts are more abstract). Each question can be scored as correct or incorrect. Many studies (e.g., LASI-DAD) score using a middle “approximately correct” option (e.g., it is okay to make a mistake, but lying is wrong) that is less than the fully correct option (e.g., lying is deliberate, mistake is unintentional).

We recommend:

- Study teams should carefully evaluate whether the constructs measured by the words in the target language are meaningfully similar or different in the target language, and whether the constructs are culturally relevant and meaningful.

Judgement and problem solving

The judgement and problem-solving questions ask questions to discern the respondent’s ability to engage in appropriate social judgment and problem-solving for everyday situations. For example, LASI-DAD, CVFS-SCAN, and KLPS ask about what the respondent would do if a child were found on the side of the road. Adaptations of this and similar questions should consider what reasonable response options would be in a given cultural setting. In KLPS, this item was changed to “found a lost child that is looking for their parents” because children often walk around alone without adults in Kenya, making the original scenario less unusual. Each question is scored as correct or incorrect.

We recommend:

- Consider the cultural appropriateness of the judgement/problem solving questions and acceptable responses and make adaptations if necessary.

Everyday calculations

These questions include simple calculations, rooted in everyday problems. For example, in LASI-DAD, one question asks, “How many 25 paise coins will you need to make six and a half rupees?” These are similar to questions included in the Core HRS cognitive testing. Changes to ensure that the currency matches the local currency are recommended. Each question is scored as correct or incorrect.

Digit span (forward and backward)

The forward and backward digit span items measure simple verbal attention span and working memory. Respondents listen to a list of random single-digit numbers and are asked to repeat them in the same order or reverse order (for backward trials) as presented. The numbers are read aloud at the rate of one per second. Trials with different numbers of digits can be used, though some studies have used only a single trial. For example, LASI-DAD administered one forward digit span trial with 5 digits, and one backward digit span trial with 3 digits.

This item may be impacted by respondents’ numeracy, though requirements are lower than counting-related tasks. The backward trial assumes familiarity with the concept of “backward” or reverse sequencing.

Hand movement sequencing

This motor sequencing test measures the repetition of a fixed sequence of hand postures designed to assess executive functioning. Respondents are asked to repeat demonstrated hand-sequencing movements. The test is adapted from existing assessments, such as those included in the Mattis 1988 Dementia Rating Scale. The LASI-DAD team has created training videos that can be requested to help in training interviewers to complete this task.

Go/No-Go

The Go/No-Go test is a measure of motor response inhibition. It requires participants to respond differently to two different cues in one trial, but then for the second trial the rule is reversed, which requires the respondent to inhibit a prepotent response. Errors in Trial 2 are considered the best measure of response inhibition. While several versions of the Go/No-Go test exist, most HCAP studies rely on adapting a version found in the Frontal Assessment Battery (Fisher et al., 2013).

Traditionally, Trial 1 has the participant tap the table twice after an examiner taps once, or tap once after the examiner taps twice, across 10 attempts. For Trial 2, the first rule remains the same, but the second rule is changed, in which the participant is required to not tap when the examiner taps twice (inhibits previously learned response) for another 10 attempts. No material is needed beyond paper and pencil to document responses, and a surface to tap.

Some HCAPs have omitted Trial 2, such as Mex-Cog in Waves 1 and 2. Mex-Cog is adding Trial 2 in Wave 3 to improve harmonization with the network. The HRS-HCAP is newly administering this item in Wave 3 (2026), also to improve harmonization with the network.

We recommend:

- Investigators seeking to ensure adequate coverage of the executive functioning domain and improve harmonization of this domain with other studies in the network should consider including the Go/No-Go task.
- If including this test, follow the administration used in LASI-DAD and HRS-HCAP.

Token test

The Token test was originally designed to assess language comprehension and, in particular, is used as a screening tool for impairments in receptive language in individuals with aphasia. It consists of a series of colored tokens in square or round forms and two different sizes, and the respondent is instructed to perform a series of commands, such as "touch the black circle" or "touch the small yellow square and the blue circle". Although originally designed as a test of language comprehension, this test empirically loads onto executive function in many HCAP studies.

This test assumes familiarity/prior exposure to names of shapes and colors. Pre-testing is recommended to ensure universal familiarity with the names used in the test, and alternate simple stimuli could be substituted for any shapes that do not have universally recognized names.

HRS number series

The test is a six-item measure of quantitative reasoning ability (Raven et al., 2000). Respondents are asked to look at a series of numbers and correctly identify the missing number in the series. All participants answer the same first set of three problems, and depending on their score on the first set, they are given a second set of three items ranging from very easy to very difficult. This is described as a block-adaptive testing scenario.

Correct scoring/data entry of the first set of three items is particularly important for this test, as it dictates the difficulty level of the next set of three items. Errors in scoring/data entry on the first set will misassign the second set of number series items.

The HRS number series test has two parallel forms, and in longitudinal studies respondents are typically assigned to alternate which form they are administered.

This test assumes high numeracy levels.

Source: Adapted from the Woodcock Johnson III Tests of Achievement: Number Series ©

Riverside Publishing. Adapted by McArdle & Woodcock (2009). * Proprietary test -- For details on the individual test items, users may see Fisher, McArdle, McCammon, Sonnega, & Weir, 2013.

Raven's standard progressive matrices

This test evaluates visual abstract reasoning. Each question presents a geometric design with a small section that appears to have been cut out. Respondents are shown a set of smaller designs that fit the missing piece and are asked to identify which is the correct one to complete the pattern. The test includes items of varying difficulty. See (Raven et al., 2000) and (Raven, 2000) for administration and measurement properties.

If administering this test with the official Raven's SPM booklets, make sure the respondent does not have a pencil/pen in hand so that they are not able to mark the booklet; rather, they should point to their response.

For HRS-HCAP, a subset of 17 questions is selected out of the 60 in the full test, including one practice question.

Material ordering and licensing information can be found here:

<https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Cognition-%26-Neuro/Raven%27s-Standard-Progressive-Matrices-%28SPM%29-and-Raven%27s-Standard-Progressive-Matrices-Plus-%28SPM-Plus%29/p/100000504>

Spatial working memory (forward and backward)

The spatial working memory test uses a stimulus of four blocks or cubes. The interviewer points at the blocks in a pre-defined sequence, and the respondent is asked to repeat the sequence, first in the forward direction and then in the reverse order (i.e., backward). The length of the sequences increases from two to five blocks over the course of both the forward and then the backward trials. The score is the number of correct sequences. This item was used in HAALSI-HCAP and was selected given its lack of reliance on literacy or numeracy.

Language

Similar to memory functioning, language abilities are affected by many neurodegenerative diseases and neurological conditions, and are impacted early on in disease processes such as Alzheimer's disease and frontotemporal lobe dementias. Most neuropsychological batteries used in cognitive aging and dementia assessment include measures of naming ability and fluency. Additionally, some assessments may include measures of basic expressive and receptive language abilities (i.e., following commands, sentence repetition). Most naming tests usually include somewhere between 15 to 60 items (i.e., brief and full versions of the Boston Naming Test, the Multilingual Naming Test, etc.). Similarly, verbal fluencies are usually measured with both semantic (i.e., animal) and phonemic/letter fluency. However, given interview time constraints and limited availability of culturally and linguistically validated tests of

language functioning, most HCAPs relied on few items largely derived from screeners and brief batteries (i.e., CSI-D, FAB, HMSE) to measure some of these abilities. As such, several of these tests demonstrate ceiling effects in community-living samples and have generally limited reliability at the higher end of language functioning (see (Arce Rentería et al., 2023) and (Vonk et al., 2022) for additional information). While this cognitive domain may be most useful in diagnostic classification of cognitive disorders (i.e., MCI, dementia), its usefulness in measuring cognitive change over time may be more limited. Efforts are currently underway to develop and/or validate additional tests to improve the measurement of this cognitive domain.

Brief naming items

These items evaluate naming ability, either by presenting auditory or visual stimuli and asking respondents to name the correct object corresponding to the stimuli. Naming visually-presented stimuli is also called confrontational naming.

One example of a visual naming item is to ask respondents to name the object that the interviewer points to, while the interviewer points to their own elbow. This is derived from the 10/66 and Community Screening Inventory Dementia (CSI-D) surveys. The MMSE also includes two visual naming items, where objects are presented, and the respondent is asked to name them. An example of an auditory naming test developed in an adaptation for LASI-DAD Wave 2 gave respondents a short description of a tree and asked respondents to produce this answer. Other auditory naming items are part of the TICS battery. In considering whether to include or adapt these items, studies should consider whether the to-be-named objects are commonly known in the context of interest, and whether there are commonly recognized words for the object (e.g., in some languages, there is not a specific widely recognized word for elbow and instead this would be generically called a joint).

Naming items from the MMSE and TICS are copyrighted. See here for licensing information for the MMSE: <https://www.parinc.com/products/MMSE> and for TICS: <https://www.parinc.com/products/TICS>

Boston Naming Test

The test is a measure of confrontation naming. Respondents are asked to provide the names of common objects (pictures). The 15-item version was included in HAALSI-HCAP, though a 30-item or 60-item version is commonly included in other studies. Other HCAP studies have considered adding similar adapted naming assessments. Items should include objects with a range of difficulties.

We recommend:

- Cultural adaptation is needed to ensure that the pictures and names are accessible to respondents within a given setting.
- Include items with a range of difficulties.

Language comprehension items

These items evaluate language comprehension. Examples of items in this group include asking respondents to describe what one would do with a hammer. Acceptable responses provided in the HRS-HCAP documents are to "pound something", "drive a nail into something", and "hit something with". This is derived from the 10/66 and CSID surveys to assess cognitive impairment and dementia. Adaptations should consider what acceptable responses should be in each context. These may vary by study; e.g., "build furniture" is an acceptable answer in Chile-Cog.

Following commands

The item evaluates language comprehension. Respondents are asked to point first to the window and then to the door. This is derived from the 10/66 and CSID surveys.

HRS-HCAP regards one of the following as correct: 1) Point to the window and then the door; 2) Point to the window only if a door is not available; 3) Point to the door only if a window is not available. If the interview is not indoors, studies may have different instructions on the objects to point to (e.g., sky and ground) and on the adaptations to commands when the points of reference are not available (e.g., LASI-DAD, KLPS, LOSHAK). These adaptations are fine.

Local market

The item evaluates expressive language and spatial orientation. Respondents are asked to describe where the local market or store is located. Phrasing and terminology can be adapted to ensure validity across settings; for example, ELSA-HCAP used "shop" whereas HRS-HCAP used "market" or "store." They are instructed to be more specific if their responses do not include a clear description or address. However, interviewers commonly do not know where the local market or store is located, and therefore, answers are not assessed for correctness but rather confidence and an appearance of correctness. This item comes from the 10/66 and CSID surveys to assess cognitive impairment and dementia.

We recommend:

- This item can be a low priority for study teams, unless harmonization with 10/66 is a priority.

Animal naming

This item evaluates semantic fluency. Respondents are asked to name as many animals as they can within a 60-second time limit. This measure is consistent with animal fluency measures commonly administered in other neuropsychological exams (e.g., the CERAD animal fluency measure; (Morris et al., 1989)).

This test requires concentration on the part of the interviewer. Interviewers need to use a stopwatch or other timing device to stop the respondent at 60 seconds. The stopwatch may be built into the interview software.

Some studies have provided interviewer guidance to probe the respondent if they pause for 10 seconds or more, to see if they can think of anything else. The test is discontinued before 60 seconds if the respondent indicates they cannot think of any additional words after the probe.

This test assumes that the word for “animals” is inclusive of all animal types. The HRS-HCAP allows for specific dog breeds, types of birds, as well as general categories of animals like dog or cat. The test also assumes prior exposure and familiarity with a large variety of animals. The instructions have been adapted accordingly. Some cultural adaptations have allowed fictional creatures where culturally appropriate (such as in the LOSHAK Pilot, TILDA).

In settings like India (LASI-DAD) and Kenya (LOSHAK Pilot), there are times when a respondent may name an animal in a language with which the interviewer is unfamiliar. Guidance in those studies is to mark these as correct, but to flag the issue in a comments box. In these studies, the interviewers communicated with each other via Whatsapp following the interviews to identify whether the word is an animal or not. This approach may help improve consistency of scoring across interviewers.

We recommend:

- Allow respondents the full 60 seconds to complete the task, unless harmonization with a specific HCAP study with different procedures is a priority.
- If feasible, audio record this test for subsequent centralized scoring and/or interviewer quality assurance.
- Accept all unique animal names, regardless of language.

Follow a multi-step command

The task is intended to measure language comprehension. The interviewers read aloud and ask the respondent to follow a simple set of instructions. Some studies (e.g., HRS-HCAP) include the standard command from the MMSE, with adaptations as appropriate for the study setting. Other studies use a different, simple 3-step command. For example, pilot testing in Nepal (CVFS-SCAN) revealed that the first step of the standard MMSE command was culturally inappropriate and a novel set of commands was utilized instead.

This item is from the MMSE and is copyrighted. See [here](#) for licensing information:

Follow a 1-step command

This item is intended to measure receptive language and is part of the MMSE (Folstein et al., 1975). The standard administration of the item requires respondents to read and follow a simple written instruction.

This item has been adapted for populations with varying levels of literacy and/or vision. In the HRS-HCAP, interviewers verbally provide the instruction for respondents who cannot read due to illiteracy or vision impairment. In LASI-DAD, the interviewer performs a simple action and asks the respondent to copy/mimic the action. In Chile-Cog, respondents are shown a picture

with an illustration of an action, and the respondent is asked to copy the action illustrated in the picture.

We recommend:

- When feasible and appropriate, use the HRS-HCAP version of this task, where the interviewer reads the instruction aloud for individuals who cannot read the instruction themselves due to vision impairment or illiteracy. We recommend this adaptation because it retains the measurement of language comprehension and also accommodates individuals with vision impairment.

The 1-step item from the MMSE is copyrighted. See here for licensing information:

<https://www.parinc.com/products/MMSE>

Language repetition

Respondents are provided with a phrase and are asked to repeat the phrase verbatim. This item is included in the MMSE. This item was developed in an English-speaking population, and the selected phrase was intended to be sensitive to non-fluent aphasia (i.e., expressive language impairment) for English-speakers.

Ganguli and colleagues (1995) provide a rich discussion of adaptation considerations for this item. Because of the linguistic and cultural nuances of this item, it is unlikely to be a high-quality linking item across HCAP studies. However, this item can be included as a measure of language repetition. Scoring details for this item are copyrighted.

We recommend:

- Study teams use a phrase that is familiar and of simple to moderate complexity. Do not use a literal translation of the English phrase. For example, the LASI-DAD team used the phrase: *Neither this nor that*. For example, in Swahili, KLPS and the LOSHAK Pilot used: *Tupe tupate tumpatie Taaka*.

The language repetition item from the MMSE is copyrighted. See here for licensing information:

<https://www.parinc.com/products/MMSE>

Write a sentence

This item is intended to measure written expressive language ability and is part of the MMSE (Folstein et al., 1975). This item assumes literacy and familiarity with basic grammar.

This item has been adapted for populations with low literacy. For example, illiterate respondents in LASI-DAD are asked to say out loud a full sentence, for example, by describing something about their house.

The expressive language item from the MMSE is copyrighted. See here for licensing information: <https://www.parinc.com/products/MMSE>

Phonemic fluency

This is a test of language and executive function. Respondents are asked to list as many different words as they can that start with a specific letter as fast as they can within 60 seconds. Semantic fluency for animals is a related test, although semantic and phonemic fluency are somewhat different. For example, phonemic fluency tests have higher demands on executive functioning compared to semantic fluency tasks. This test was added to the HRS-HCAP in 2026, where respondents list words starting with C then F if doing the interview in English, and letters starting with letters P then M if completing the interview in Spanish (Benton, 1969). An adaptation for low-literacy populations tested in HAALSI asked respondents to list words that begin with a specific sound (e.g., ssssss). Despite this adaptation, this test was dropped from the HAALSI battery due to frequently observed scores of 0. In addition, this adaptation is designed to accommodate illiterate respondents but may disadvantage literate respondents (e.g., psychology would be acceptable; ship would not). Further work is needed to better understand the performance of this adaptation in settings with varying literacy.

Visuospatial

Tests of visuospatial functioning measure different aspects of how individuals process visual and spatial information. Because we spatially navigate and interact with our daily environment primarily through vision, visuospatial abilities are fundamental to everyday functioning in later life. Impairment in visuospatial functioning is observed in several forms of AD/ADRD, and is impaired relatively early in the disease course for several dementia types (e.g., Lewy body dementia; posterior cortical atrophy). The relative importance of this domain may vary across cultural and environmental contexts.

Most of the visuospatial tests used across various HCAPs rely on asking the participant to copy or produce different geometric figures. As such, the tests included for this domain require that the participant have some familiarity with geometric shapes and utilization of writing utensils, which could impact their validity and reliability among populations with varying literacy levels.

It is important to note that while we advocate for the inclusion of tests to measure five cognitive domains (including visuospatial functioning) in an HCAP, ongoing harmonization efforts do not include harmonized visuospatial domain factors (see (3)) because most existing HCAPs do not contain a sufficient number of test items measuring this domain. In some HCAP dementia classification algorithms, constructional praxis alone is used to represent the visuospatial domain. Given the small number of available items, it may not be possible to estimate a measurement model for this domain in factor analyses or to complete statistical harmonization of the domain. However, given its relevance to later-life cognitive functioning and its sensitivity to AD/ADRD, we still encourage investigators to include tests of visuospatial functioning or consider supporting the development and adaptation of tests to be included in future HCAP studies.

Tasks in this domain depend on a level of visual acuity and contrast sensitivity that is sufficiently adequate to perceive visually-presented stimuli. In populations with a high prevalence of

uncorrected vision impairment, performance on visuospatial tests may reflect sensory limitations rather than cognitive ability per se, complicating interpretation. Adequate interpretation of cognitive scores is therefore reliant on the documentation of vision issues by interviewers or, ideally, objective measures of vision impairment.

Constructional praxis (copying)

This test involves visuospatial construction and visual-motor integration to perceive and reproduce a series of geometric figures. Respondents are asked to draw four geometric shapes of varying complexity. Each shape is shown one at a time on a worksheet. Respondents are asked to copy the shape below it on the same sheet of paper. The tasks are part of the CERAD.

See here for more information and to request permission to use the measure:

<https://www.naccddata.org/request-cerad-forms/>

This test is not feasible in all settings due to requirements with respect to drawing shapes that are often taught in school. For example, Chile-Cog and Mex-Cog instruct respondents to draw a rhombus instead of one of the CERAD shapes, given greater familiarity among those with low formal education. In KLPS (Kenya), one CERAD shape was replaced with a triangle since the original shape is not commonly taught in Kenyan schools.

Further, some studies with populations with varying levels of formal school exposure have substituted this test with a matchstick copy test (e.g., HAALSI, CVFS-SCAN; (Baiyewu et al., 2005). Respondents are shown a series of designs made with matchsticks and are asked to use matchsticks to reproduce the design. The matchstick copy test does not require drawing.

We recommend:

- Investigators consider whether respondents are able to perform copying tasks with a writing utensil, and whether the use of different shapes or substitution with an alternative material, such as the matchstick copy test, is needed.

Interlocking Shapes

The task requires spatial perception and visual-motor integration and measures visuospatial constructional ability. Respondents are presented with a picture of two overlapping shapes and are asked to recreate it. Scoring criteria are copyrighted as part of the MMSE (6) and may differ between studies. This test assumes familiarity and prior exposure to drawing geometric figures and the use of writing utensils, which are experiences that are typically provided through school exposure.

The task is part of the Mini-Mental State Examination (MMSE) (Folstein et al., 1975). See here for more copyright information: <https://www.parinc.com/products/MMSE>

Considering high illiteracy rates and potential challenges with drawing complex shapes, Chile-Cog additionally incorporates the task of drawing two overlapping circles (Quiroga et al., 2004).

CVFS-SCAN created a novel matchstick item, using the same framework as the matchstick copy test as a substitute for the interlocking shapes item.

Clock drawing

The test involves visuospatial constructional and executive abilities. Respondents are asked to draw a clock showing a pre-specified time. Scoring is based on three criteria: whether drawing a closed circle or square, correctly placing clock numbers on the circle/square, and showing two clock hands. The clock drawing test assumes familiarity with drawing, writing numbers, and analog (i.e., non-digital) clocks.

We recommend:

- Pre-testing the appropriateness of this item in settings with varied educational attainment or where the use of analog clocks is not common. For example, this item performed well in the LASI-DAD study in India, but not in the LOSHAK pilot study in Kenya or CVFS-SCAN in Nepal.

Orientation

Being oriented to time, place, and self are fundamental cognitive abilities included in most cognitive screeners (i.e., MMSE, MoCA) and featured in most routine clinical evaluations. As such, all HCAP studies have included several items for orientation to place and to time. While these items, and any summary score for overall orientation (either a sum score or a factor derived from CFA), generally have a ceiling effect in most community-based samples, problems with orientation can indicate cognitive impairment.

Place

These items assess ability to orient to the current physical environment. Studies usually administer five questions within this category. The following items are examples of what respondents are asked. Copyrighted items from the MMSE are redacted.

1. What is the name of the building we are in now?
2. Name two roads nearby.
3. What is the name of this place (or room) you are in now?
4. What is this room used for?

Scoring criteria can vary across studies depending on what is appropriate in different populations.

Care is warranted when selecting broader geographically-based items, as politically defined geographic units such as province or county may have limited salience for many older adults. Or, worse, familiarity with such boundaries may tap into a domain different from cognitive orientation ability: for example, consider whether those who travel frequently across major geographic borders such as provinces may be more attuned to these distinctions (e.g., people

with higher socioeconomic backgrounds). As a result, differential exposure rather than underlying cognitive ability could influence item performance, potentially introducing spurious socioeconomic gradients in scores.

Several questions about orientation to place are part of the Mini-Mental State Examination (MMSE) or Folstein test (Folstein et al., 1975). See here for more copyright information: <https://www.parinc.com/products/MMSE>

We recommend:

- Investigators decide on and enumerate appropriate orientation to place items and accompanying correct responses prior to study implementation.

Time

These items assess awareness of current time context. Studies usually administer five questions within this category. Respondents are asked separate questions, such as:

1. What time of day is it?
2. How long ago was {last Christmas} {a universally memorable event}?

Across studies, there are slight differences in the wording of some of the MMSE orientation-to-time items, and acceptable responses vary by study and should be evaluated during pre-testing and decided before the HCAP is fielded.

Several questions about orientation to time are part of the MMSE or Folstein test (Folstein et al., 1975): <https://www.parinc.com/products/MMSE>

We recommend:

- Engage adaptation teams and perform pre-testing to determine appropriate orientation items and corresponding acceptable correct responses.
- Decide, based on cultural context and pre-testing, what is a correct response before the HCAP is fielded.

Name political leader

This task evaluates orientation to time and situation and verbal recall. This item is commonly assigned to either the language or orientation domain, though we include it in orientation here as it is often clinically used to assess orientation. Respondents are instructed to name the current political leader, such as the prime minister. The question and acceptable responses about the political leader are culturally adapted as necessary, such as the Prime Minister of the United Kingdom (ELSA-HCAP), Prime Minister of India (LASI-DAD), and President and Deputy President of South Africa (two separate questions in HAALSI-HCAP). We note several key challenges with this item: (1) political changes and election seasons can impact comparability over time, (2) differences in political systems hinder international comparisons, and the item may therefore not be considered equivalent across settings even despite appropriate

adaptation, and (3) political conflict can lead to protest answers, which are difficult to score. For scoring responses to this item when the field period overlaps with a change in political leadership, either the old or new leader can be accepted within a pre-specified window.

The HRS-HCAP political leader question comes from the copyrighted TICS task; information on licensing can be found here: <https://www.parinc.com/products/TICS>

Missing data coding

To ensure cross-study comparability within the HCAP, standardization and consistency in test administration, scoring procedures, and missing data coding are crucial. As stressed throughout this guidance document, consistent administration procedures are essential to ensure that comparable data are collected across study settings. Proper documentation on missing codes is also instrumental for secondary data analyses, including imputation strategies and sensitivity assessments.

HCAP studies are encouraged to implement a standardized set of missingness categories for every variable. These categories may include, but are not limited to: “don’t know,” “refusal,” “skipped,” “not assessed,” “administration error,” “cannot read,” “cannot write,” “cannot count,” and “physical limitation.”

We also recommend including an indicator for the type of respondent in an HCAP dataset, namely whether the respondent has 1) both a respondent and informant interview, 2) a respondent interview only, or 3) an informant interview only.

For some cognitive test items, particularly delayed recall tasks and tests involving numerical counts, it is critical, at the level of data collection, to distinguish true scores of zero from various forms of missing data. A score of zero should reflect the difficulty level of cognitive tests and participant cognitive performance, while missing codes record true missingness due to non-cognitive reasons, such as physical, sensory, or language barriers, or refusals. Accurate item-level documentation of missingness, supported by interviewer training on appropriate code selection, is required to inform analytic decisions and imputation strategies. Responses reflecting cognitive difficulty (e.g., attempted but cannot answer) may be coded as “don’t know” but scored as zero analytically, whereas items not completed for non-cognitive reasons retain their distinct missing codes. Clear guidance on these distinctions is essential to preserve data integrity and facilitate appropriate cross-national harmonization and analysis.

More information on missing data coding in the HRS-HCAP is available on the [HRS website](#). Further documentation of observed missingness and an approach for imputations across several HCAP studies is available in the following work:

Meijer E, Liu Y. **Missing data and imputation in the international studies following the Harmonized Cognitive Assessment Protocol**. *J Gerontol B Psychol Sci Soc Sci*. 2026;81(Supplement_1):S55-S63.

Recommended further reading

Briceño EM, Bassil DT, Khobragade P, Ngugi A, El Bejjani M, Ochieng E, Mejia-Arango S, Douhou S, Kulisewa K, Arce Rentería M, Gross AL, Jones RN, Lee J, Langa KM, Kobayashi LC. **Recommended best practices for construct-centered adaptation of the Harmonized Cognitive Assessment Protocol.** *J Gerontol B Psychol Sci Soc Sci.* 2025 Oct 25:gbaf201. doi: 10.1093/geronb/gbaf201. Epub ahead of print.

Briceño EM, Dhakal U, Sharma U, Adhikari N, Pradhan MS, Shrestha L, Jalan P, Rai J, Langa KM, Lee J, Ghimire D, Mendes de Leon CF. **Neuropsychological Assessment of Older Adults in Nepal for Population-Based Dementia Ascertainment: Needs, Challenges, and Opportunities.** *J Alzheimers Dis.* 2023;96(4):1339-1352. doi: 10.3233/JAD-230906.

Briceño EM, Arce Rentería M, Gross AL, Jones RN, Gonzalez C, Wong R, Weir DR, Langa KM, Manly JJ. **A cultural neuropsychological approach to harmonization of cognitive data across culturally and linguistically diverse older adult populations.** *Neuropsychology.* 2023;37(3):247–257. doi:10.1037/neu0000816

Gross AL, Mani SS, Weir D, Briceño EM, Rentería MA, Nichols E, Jones RN, Manly JM, Langa KM, Ikanga J, Lee J, Kobayashi LC. **Harmonized Cognitive Assessment Protocol (HCAP): What Makes an HCAP an HCAP? An Adaptive Architecture for Measuring Cognitive Aging Around the World.** *J Gerontol B Psychol Sci Soc Sci.* 2025 Oct 23:gbaf202. doi: 10.1093/geronb/gbaf202. Epub ahead of print.

Kobayashi LC, Jones RN, Briceño EM, Rentería MA, Zhang Y, Meijer E, Langa KM, Lee J, Gross AL. **Cross-national comparisons of later-life cognitive function using data from the Harmonized Cognitive Assessment Protocol (HCAP): Considerations and recommended best practices.** *Alzheimers Dement.* 2024;20(3):2273-2281. doi: 10.1002/alz.13694.

Meijer E, Liu Y. **Missing data and imputation in the international studies following the Harmonized Cognitive Assessment Protocol.** *J Gerontol B Psychol Sci Soc Sci.* 2026;81(Supplement_1):S55-S63.

Nichols E, Gross AL, Kobayashi LC, Langa KM, Lee J. **The measurement of cognition in large-scale cross-national surveys: Lessons from the Health and Retirement International Network of Studies and the Harmonized Cognitive Assessment Protocol.** *J Gerontol B Psychol Sci Soc Sci.* 2025 Oct 21;gbaf200. Epub ahead of print.

Cohort profile papers

Bassil DT, Farrell MT, Wagner RG, Brickman AM, Glymour MM, Langa KM, Manly JJ, Salinas J, Tipping B, Tollman S, Berkman LF. **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.* 2021;51(4):e217–e226.

Cadar D, Abell J, Matthews FE, Brayne C, Batty GD, Llewellyn DJ, Steptoe A. **Cohort profile update: the harmonised cognitive assessment protocol sub-study of the English Longitudinal Study of Ageing (ELSA-HCAP).** *Int J Epidemiol.* 2021;50(3):725–726i.

Feeney J, Monaghan A, McLoughlin S, De Looze C, Oto G, Lawlor B, Weir DR, Kenny RA, McGarrigle CA. **Cohort profile update: the harmonised cognitive assessment protocol sub-study of The Irish Longitudinal Study on Ageing (TILDA-HCAP).** *Int J Epidemiol.* 2025;54(1):dyaf008.

Khobragade PY, Petrosyan S, Dey S, Dey AB, Lee J. **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.* 2025;73(3):685–96.

Langa KM, Ryan LH, McCammon RJ, Jones RN, Manly JJ, Levine DA, Sonnega A, Farron M, Weir DR. **The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: study design and methods.** *Neuroepidemiology.* 2020;54(1):64–74. doi: 10.1159/000503004.

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;9(7):e030300.

Liu MM, Llibre-Guerra J, Soria C, Li J, Llerena TZ, Rodriguez G, Acosta D, Velázquez IJ, Llibre-Rodriguez JJ, Dow WH. **The Caribbean American Dementia and Aging Study: protocol for a population-based study of older adult health and dementia in Cuba, the Dominican Republic, and Puerto Rico.** *BMC Geriatr.* 2025;25:669.

Mejia-Arango S, Nevarez R, Michaels-Obregon A, Trejo-Valdivia B, Mendoza-Alvarado LR, Sosa-Ortiz AL, Martinez-Ruiz A, Wong R. **The Mexican Cognitive Aging Ancillary Study (Mex-Cog): study design and methods.** *Arch Gerontol Geriatr.* 2020;91:104210.

Riang'a RM, Mwangi EM, Nagarajan N, Agoi F, Mwangala PN, Gross AL, Ikanga J, Langa KM, Miguel E, Gichu M, Ehrlich JR, Ngugi AK. **Contextualization of Harmonized Cognitive Assessment Protocol (HCAP) in an aging population in rural low-resource settings in Africa: Experiences and strategies adopted to optimize effective adaption of cognitive tests in Kenya.** *Alzheimers Dement.* 2025;21(8):e70552. doi: 10.1002/alz.70552.

Citations

1. Manly JJ, Jones RN, Langa KM, 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;79(12):1242-1249. doi:10.1001/jamaneurol.2022.3543
2. Mejia-Arango S, Arce Rentería M, Saenz JL, Wong R. Prevalence of dementia and mild cognitive impairment in the Mexican population: evidence from Mex-Cog, a harmonized cognitive assessment protocol study. Muñoz E, ed. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2026;81(3):gbaf268. doi:10.1093/geronb/gbaf268
3. Gross A, Kobayashi L, Li C, et al. Harmonization of Later-Life Cognitive Function Across National Contexts: Results from the Harmonized Cognitive Assessment Protocols (HCAPs). *medRxiv*. Published online 2023:2023.06.09.23291217. doi:10.1101/2023.06.09.23291217
4. Gross AL, Mani SS, Weir D, et al. Harmonized Cognitive Assessment Protocol: What makes an HCAP an HCAP? An adaptive architecture for measuring cognitive aging around the world. Behrman JR, ed. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2026;81(Supplement_1):S31-S43. doi:10.1093/geronb/gbaf202
5. Briceño EM, Bassil DT, Khobragade P, et al. Recommended best practices for construct-centered adaptation of the Harmonized Cognitive Assessment Protocol. Behrman JR, ed. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2026;81(Supplement_1):S44-S54. doi:10.1093/geronb/gbaf201
6. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." *Journal of Psychiatric Research*. 1975;12(3):189-198. doi:10.1016/0022-3956(75)90026-6
7. Wechsler D. *Wechsler Memory Scale*. 4th ed.
8. Rosen W, Mohs R, Davis K. A new rating scale for Alzheimer's disease. *AJP*. 1984;141(11):1356-1364. doi:10.1176/ajp.141.11.1356
9. 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*. 2024;30(1):47-55. doi:10.1017/S135561772300019X
10. Mani SS, Rianga RM, Nagarajan N, et al. Validation of cognitive and psychosocial tools in Kenya: findings from the LOSHAK feasibility pilot. *BMC Public Health*. 2025;25(1):4008. doi:10.1186/s12889-025-24918-z
11. Lachman ME, Agrigoroaei S, Tun PA, Weaver SL. Monitoring Cognitive Functioning: Psychometric Properties of the Brief Test of Adult Cognition by Telephone. *Assessment*. 2013;21(4):404-417. doi:10.1177/1073191113508807
12. Richards M, Kuh D, Hardy R, Wadsworth M. Lifetime cognitive function and timing of the natural menopause. *Neurology*. 1999;53(2):308-308. doi:10.1212/WNL.53.2.308
13. Smith A. *Symbol Digit Modalities Test: Manual*.

14. Reitan R. *Trail Making Test: Manual for Administration and Scoring*.
15. Lu L, Bigler ED. Performance on Original and a Chinese Version of Trail Making Test Part B: A Normative Bilingual Sample. *Applied Neuropsychology*. 2000;7(4):243-246. doi:10.1207/S15324826AN0704_6
16. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: A frontal assessment battery at bedside. *Neurology*. 2000;55(11):1621-1626. doi:10.1212/WNL.55.11.1621
17. Fisher G, McArdle J, McCammon R, Sonnega A, Weir D. *New Measures of Fluid Intelligence in the HRS*. 2013. <http://hrsonline.isr.umich.edu/sitedocs/userg/dr-027.pdf>
18. Raven J, Raven J, Court J. *Manual for Raven's Progressive Matrices and Vocabulary Scales: Research Supplement No. 3. A Compendium of International and North American Normative and Validity Studies Together with a Review of the Use of the RPM in Neuropsychological Assessment*. 2nd/3rd. The Psychological Corporation; 2000.
19. Raven J. The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*. 2000;41(1):1-48. doi:10.1006/cogp.1999.0735
20. Arce Rentería M, Briceño EM, Chen D, et al. Memory and language cognitive data harmonization across the United States and Mexico. *Alzheimers Dement (Amst)*. 2023;15(3):e12478. doi:10.1002/dad2.12478
21. Vonk MJM, Gross AL, Zammit AR, et al. Cross-national harmonization of cognitive measures across HRS HCAP (USA) and LASI-DAD (India). *PLOS ONE*. 2022;17(2):e0264166. doi:10.1371/journal.pone.0264166
22. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assesment of Alzheimer's disease. *Neurology*. 1989;39(9):1159-1159. doi:10.1212/WNL.39.9.1159
23. Ganguli M, Ratcliff G, Chandra V, et al. A hindi version of the MMSE: The development of a cognitive screening instrument for a largely illiterate rural elderly population in india. *Int J Geriat Psychiatry*. 1995;10(5):367-377. doi:10.1002/gps.930100505
24. Benton AL. Development of a multilingual aphasia battery. *Journal of the Neurological Sciences*. 1969;9(1):39-48. doi:10.1016/0022-510X(69)90057-4
25. Baiyewu O, Unverzagt FW, Lane KA, et al. The Stick Design test: A new measure of visuoconstructional ability. *J Int Neuropsychol Soc*. 2005;11(5):598-605. doi:10.1017/S135561770505071X
26. Quiroga RQ, Nadasdy Z, Ben-Shaul Y. Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. *Neural Computation*. 2004;16(8):1661-1687. doi:10.1162/089976604774201631