Routledge
Taylor & Francis Group

Linguistic and cultural properties of the Spanish adaptation of the CAT (SP-CAT): pilot results from neurotypical subjects

Silvia Martínez Ferreiro [ID][a]*, Yina M. Quique [ID][b]*, Viviana Alejandra Rodríguez [ID][c] and Carolina Méndez Orellana [ID][d,e]

[a]Gerontology and Geriatrics Research Group, Department of Physiotherapy, Medicine and Biomedical Sciences, University of A Coruña, Spain; [b]Shirley Ryan Ability Lab & Northwestern University, Chicago, Illinois, USA; [c]Department of Clinical Epidemiology and Biostatistics, Pontificia Universidad Javeriana. Bogotá, Colombia; [d]Speech and Language Therapy School, Health Sciences Department, Faculty of Medicine, Pontificia Universidad, Católica de Chile, Santiago, Chile; [e]Orellana, Department of Radiology, Erasmus MC -University Medical Center Rotterdam, Rotterdam, The Netherlands

ABSTRACT

Introduction: Aphasia assessments in languages other than English are scarce. In the case of Spanish, this scarcity includes a need for assessments with linguistic and cultural adaptations that consider dialectal varieties and cultural traits across Spanish-speaking populations.

Aims of the study: This study discusses the linguistic and cultural adaptation of the Comprehensive Aphasia Test (CAT) to Spanish (SP-CAT), a version that can be used in Spain and Spanish-speaking Latin American countries, and provide pilot results assessing whether performance is comparable across samples.

Methods: For the linguistic adaptation, we discuss considerations such as typological differences between English and Spanish, Spanish varieties, gender cues, spelling-sound regularities, transparency, and other syntax-related aspects. For the cultural adaptation, we discuss considerations such as culturally relevant items and images, and covering different Spanish varieties within the SP-CAT. The pre-testing of items for the SP-CAT included controlling variables such as name agreement of visual stimuli (examined in n=237 healthy participants), imageability (examined in n=244 healthy participants), and lexical frequency (from the Corpus of Reference of Current Spanish). We also conducted a pilot study of the SP-CAT with 82 healthy participants from Chile, Colombia, and Spain to assess differences in performance within tasks between the included countries; analysis of such differences was completed within a Bayesian framework.

Results: The SP-CAT provides a linguistic and cultural adaptation of the original English CAT. Item pre-testing included name agreement, lexical frequency, and imageability tests to ensure comparability with the English original version. Our pilot study confirmed that there are no clinically significant differences in performance within tasks between the included countries in healthy participants, a necessary step towards the final validation of a test for the Spanish-speaking world.

CONTACT Carolina Méndez Orellana ✉ carolinamendez@me.com ⬚ Speech and Language Therapy School, Health Sciences Department, Faculty of Medicine, Pontificia Universidad, Católica de Chile, Santiago, Chile

*Shared-first authorship

Discussion: The SP-CAT responds to a need to develop linguistically and culturally sensitive adaptations of assessments for Spanish-speaking people with aphasia to be used in clinical practice. Pilot results indicate that the adaptation meets the criteria to be used across Spanish varieties. In light of promising pilot results, the next phase of this study will assess the validity and reliability of the SP-CAT, providing normative data for its administration.

1. Introduction

Aphasia assessments in languages other than English are scarce (e.g., Beveridge & Bak, 2011; Ivanova & Hallowell, 2013), leaving a huge need to develop tests that are representative of the cultural and linguistic diversity of the world. In the case of Spanish, one of the most spoken languages worldwide, there is a scarcity of psychometrically sound aphasia assessments, with many available tools lacking normative data for the population they aim to serve (Ivanova & Hallowell, 2013). Spanish is the official language in 20 countries from Europe to Polynesia (Eberhard et al., 2019), and it is estimated that it is spoken by over 586 million people (Cervantes Institute, 2020). A total of 489 million are native speakers, mainly in Latin American countries and Spain, with prominent dialectal differences that result from geographical, cultural, and historical factors (Lipski, 1994, 1996). These differences, being linguistic or cultural, must be considered when developing aphasia assessments for the Spanish-speaking population around the globe. However, some assessments have been normed in Spain (e.g., Cuetos & Alija, 2003; Valle & Cuetos, 1995), without norms developed for Spanish-speaking Latin American populations; similarly, some assessments are normed in Latin America, without norms for Spain (e.g., Vigliecca et al., 2011). This poses difficulty when assessing speakers of different dialectal varieties of Spanish, as there is no equivalence across countries.

Beyond the scarcity of aphasia assessments with sound psychometric properties for diverse Spanish-speaking contexts, an additional level of complexity comes from *direct* translations of aphasia assessments from other languages. Specifically, some aphasia assessments for Spanish speakers are direct translations from English (e.g., the Spanish version of the Western Aphasia Battery-WAB; Kertesz et al., 1990), which overlooks the cultural and linguistic adaptations necessary to achieve validity in different contexts. Although assessments that have resulted from direct translations are commonly used in clinical settings, clinicians complain about their ecological validity, and accept items as correct even if the test does not specifically mention them. For example, in the WAB naming task, the item *ball* has multiple context-specific translations in Spanish (e.g., *balón, bola, pelota*), which should be accepted as correct, but the test does not explicitly mention them; similarly, the question *Does it snow in July?* is not relevant for Spanish-speaking countries near the equator, where snow is uncommon all year round.

Another challenge when developing aphasia assessments in Spanish is the lack of open-source comprehensive databases for linguistic properties such as lexical frequency, age of acquisition, name agreement, etc. Although there are some databases for these linguistic properties (e.g., Duchon et al., 2013; Alonso et al., 2015), issues such as the size of the database (i.e., smaller compared to English), country of origin (i.e., Spain vs. Latin

American countries), date of publication (i.e., outdated), and restricted access pose a problem for their use in the development of assessment tests for Spanish-speakers with aphasia. A promising course of action is the development of tests that could be applicable to Spanish-speaking populations worldwide; that is, tests with adaptations (as opposed to direct translations) that consider linguistic and cultural differences across Spanish-speaking countries. In this paper, we report on the linguistic and cultural adaptation of the Comprehensive Aphasia Test (CAT; Swinburn et al., 2004) for Spanish speakers.

The CAT is a theoretically anchored test originally conceived for English speakers with acquired aphasia that includes a cognitive screen, a language battery, and a concise Aphasia Impact Questionnaire (AIQ; Swinburn et al., 2019). It is supported by normative data of people with and without aphasia, it is widely used in clinical settings, and shows good reliability (Howard et al., 2010). The CAT is currently being adapted to several languages spoken worldwide as part of an international research collaboration, the Collaboration of Aphasia Trialists. Thus far, there are published adaptations of the CAT to Catalan (Salmons et al., 2021), Croatian (Swinburn et al., 2020), Danish (Swinburn et al., 2014)[1], Dutch (Visch-Brink et al., 2014), Hungarian (Zakariás & Lukács, 2022), Norwegian (Swinburn et al., 2021), and Turkish (Maviş et al., 2022). Other adaptations are currently in progress (for further details, see Martínez-Ferreiro et al., 2024).

This paper aims to report on the linguistic and cultural adaptations of the cognitive and language subtests of the Spanish CAT (SP-CAT) based on data from Spain, Colombia, and Chile and provide pilot results assessing whether performance is comparable across samples. Discussion of the AIQ, which substituted the Disability Questionnaire in the second edition of the CAT (Swinburn et al., 2023), is outside the scope of this paper. However, given the simplicity of the AIQ and the wide range of rating scales made available by the English original (n = 8), no specific modifications were required to adapt the AIQ to Spanish[2] In what follows, we describe the adaptation process and present the results of a pilot study with neurotypical individuals in these countries.

2. Spanish across the globe

As mentioned earlier, Spanish is a language spoken by millions across the globe. Although there is no consensus, according to Eddington (2022), there could be six main dialect zones, which include Spain, Southern Cone (Uruguay, Argentina), Southern Central America (Costa Rica, Panama), Caribbean (Puerto Rico, Dominican Republic), Northern Central America (Nicaragua, El Salvador, Guatemala, Honduras), and Andean South America (Bolivia, Paraguay, Chile, Peru), with a lot of variability in central and northern American countries such as Cuba, Ecuador, Mexico, Venezuela, Colombia, and the US. Despite this dispersion, mutual intelligibility and a common written standard characterize all Spanish dialects. Some variation points between dialects are intonation, pronunciation, vocabulary, and pronoun use. Far from being exhaustive, a list of some of the most prominent phenomena is briefly described below.

---

[1]The Danish adaptation of the CAT was done independently of the Collaboration of Aphasia Trialist and does not adhere to the same set of guidelines as the other adaptations mentioned here.
[2]For an extensive discussion of this update see Martínez-Ferreiro et al., 2024.

Prominent pronunciation differences reside in the presence or absence of phenomenon such as the *seseo* (consisting of the lack of distinction between /θ/ and /s/) or the *yeísmo* (the distinction between /ʎ/ and /j/), and debuccalization of final /s/, which vary across dialects both in America and Europe. Differences in vocabulary are also found, especially among common everyday objects (e.g., food, clothes), with American varieties showing lexical influence from Native American languages. Among the grammatical features, variation is found in the use of second-person pronouns. Singular pronouns *tú* and *usted* "you-sg." are associated with informal and formal registers in areas such as Spain, whereas *vos* substitutes the informal "you" in certain American countries (e.g., see *voseo* in Argentina or Uruguay). In plural, many American dialects lack the informal *vosotros* and rely on *ustedes* both for formal and informal contexts. Other less prominent features can be found. For instance, the use of compound tenses, especially the present perfect, is also subject to variation, with some varieties in America and the northwest of Spain overusing the simple past in contexts traditionally requiring the former.

3. From the ENG-CAT to the SP-CAT: the adaptation procedure

To achieve the linguistic and cultural adaptation of the CAT to Spanish (SP-CAT), we followed the adaptation guidelines of the Working Group 2 (WG-2), Aphasia Assessment and Outcomes, of the Collaboration of Aphasia Trialists (Fyndanis et al., 2017; Martínez-Ferreiro et al., 2024). The first stage entailed extensive discussions (broad and focused) regarding the linguistic and cultural particularities relevant to the SP-CAT. Broad discussions included issues such as lack of representation in the field of aphasiology, and an evaluation of needs and complaints about finding appropriate assessment tools for Spanish-speaking clients with aphasia. Focused discussions included thorough reflections on the linguistic properties of Spanish, ideas about how to adapt specific tasks, brainstorming sessions for potential items of interest, and planning for data collection. These team discussions involved researchers from Chile, Colombia, and Spain, the Spanish adaptation team, and served as a whiteboard of ideas for linguistic and cultural traits that were then included in the next stage, item selection, described in the linguistic adaptation section.

3.1. Linguistic adaptation

The CAT was originally designed to measure oral and written language comprehension and production skills at the word, sentence, and paragraph level (see Appendix 1 for a summary of CAT subtests). In the few instances in which the original CAT items were linguistically (and culturally) equivalent to the SP-CAT, they were merely translated into Spanish. Although multiple adaptations of assessments in other fields have used forward and back translations (e.g., Lenz et al., 2017; Ronê et al., 2019; Mirza et al., 2022), the adaptation guidelines were very specific against this approach, and instead encouraged an item-by-item consideration of cultural and linguistic properties. As a consequence, most items had to be modified from the

original CAT version. Only six tasks (out of 27) remained unchanged from the English version (Tasks 1 - *Line Bisection*, 3 - *Word Fluency*, 6 - *Arithmetic*, 13 - *Repetition of Complex Words*, 15 - *Repetition of Digit Strings*, and 22 - *Reading Function Words*)[3].

To align with the CAT original rationale and maintain the degree of complexity of the tasks, various measures were undertaken. First, typological differences between English and Spanish were accommodated under a single, comparable test. Second, divergences between Spanish varieties were examined and controlled to create a functional, unique version across countries. This entailed an exhaustive examination of language-specific phenomena such as the Spanish *seseo* accent, that is, the assimilation of /s/ and /θ/ (represented by the spellings 'c' before 'e' or 'i', 'z' and 's') into /s/. This phenomenon is common in certain Spanish-speaking countries, especially in America. To control for the potential effect of *seseo*, relevant for Task 3, *Word Fluency* (aiming to elicit words starting with the letter 's'), normative data from different countries, age, and educational level groups (Olabarrieta-Landa et al., 2015) was examined. Given that the ample information available allowed for an exhaustive control of the results across different demographic variables, the decision was made to keep the letter 's' as in the original test. In what follows, we discuss the linguistic adaptations carried out by the Spanish team in the cognitive and language subtests, ordered by task. A summary of the specific changes in each task can be found in Appendix 1.

Some tasks required minimal adjustments to make them adequate for Spanish speakers with aphasia across the board. For example, Task 2 (*Semantic Memory*), and consequently, Task 4 (*Recognition Memory*) were modified to accommodate differences across Spanish varieties. An illustrative example is the case of the word *glasses*, which, depending on the country or register, can be translated as *gafas* or *anteojos* (Spain), *lentes* (Chile), and *espejuelos* (Puerto Rico). Whenever possible, items eliciting different responses in different areas were removed (e.g. *libélula/matapiojo* 'dragonfly'), and items that led to an unambiguous response were selected instead (*pulpo* 'octopus'). The selection of new unambiguous items required multiple online searches and confirmation with stakeholders (e.g., clinicians, linguists, and neurotypical speakers) across Argentina, Chile, Colombia, Spain, and Uruguay.

Another consideration when developing the linguistic adaptations of the SP-CAT was the use of compounds in the original test (e.g., Task 2, item 10: ENG - *watering can*; SP-*regadera*). Compounding is claimed to be a less productive word formation strategy in Romance languages than in English (Paillard, 2000). Hence, compound forms, which in many cases correspond to simple or derived lexemes in Spanish (e.g., ENG: *paintbrush* vs. SP: *brocha* in Task 5 (*Gesture Object Use*); and the above-mentioned case of *regadera*), were avoided.

Contrary to the minimal adjustments described above for Tasks 2 and 4, other tasks, such as those including phonological distractors, had to be completely modified. In the original Tasks 7 and 8, *Comprehension of Spoken and Written words*, items are classified according to the distance between the target item and the phonological distractor in terms of distinctive features (one, two, and three features; e.g., *pear-bear, knee-bee*, and *pine-pipe*, respectively) and to the position of the differential phoneme within the word (initial, middle & final position; e.g., *pear-bear, pine-pipe*, and *leak-leaf*, respectively). Given

---

[3]In the case of Task 13 and Task 22, the Spanish counterparts of the original words met the complexity criteria (Task 13) and the functionality criteria (Task 22) required by the test, thus allowing for a mere translation.

that the authors of the original version reported that these tasks were very easy for English speakers, the Spanish version adhered to the general consensus and included words with a maximum distance of two distinctive features. Similarly, the position within the word was also adjusted. Specifically, Spanish minimal pair differences are restricted to the initial and middle positions due to the scarcity of singular nouns ending in consonants in comparison to languages such as English (e.g., one distinctive feature, middle position: *carta* 'letter' - *carpa* 'carp'; two distinctive features, initial position: *dedal* 'thimble' - *pedal* 'pedal').

In tasks targeting sentence-level phenomena (Tasks 9 - *Comprehension of Spoken Sentences*, 10 - *Comprehension of Written Sentences*, & 16 - *Repetition of Sentences*), length and complexity had to remain unchanged per CAT guidelines, thus leading to a new set of challenges to overcome. Aspects such as differences in word order between Spanish and English were easily incorporated, as they did not affect the final outcome of the tasks. For instance, Spanish syntax generally favors adjectives placed after nouns, except in cases where the adjective wants to be emphasized or to give the phrase a literary tone, resulting in adjective-noun inversions with respect to the English version (e.g., Tasks 9 & 10: *El zapato*-N *amarillo*-ADJ *está bajo el lápiz* 'The yellow shoe is under the pencil'). However, asymmetries in the use of determiners and prepositions had to be compensated for to keep the same distribution of lexical and functional elements of the English original (e.g., Task 16: SP: *El niño pequeño subió a la colina y observó el bosque* 'The small child climbed TO the hill and watched the forest' vs. ENG: *The boy and girl climbed the hill and admired the view*). Moreover, some items in Spanish allow for the inclusion of an optional reflexive pronoun (e.g., Task 9: *El hombre (se) está comiendo una manzana* 'The man is (himself) eating an apple'). Hence, they are susceptible to including an extra functional word. Given its optional character, the decision was made to keep them the same as in the original test. Another major change had to do with gender cues. Following the guidelines in Fyndanis et al. (2017), reversible sentences in Tasks 9 and 10 were modified to include animate noun phrases of matching sex to avoid possible interpretations based on this feature (e.g., SP: *El cocinero*-MASC *llama al doctor*-MASC 'The cook calls the doctor' vs. ENG: *The butcher*-MASC *shoots the nurse*-FEM).

In Tasks 12 (*Repetition of Words*), 17 (*Naming Objects*), and 20 (*Reading Words*), word length was adapted to better capture the characteristics of Spanish vocabulary. Monosyllabic words are common in English, but rare in Spanish; thus, rendering the tasks exploring the contrast between short (monosyllabic) and long (trisyllabic) words less relevant than in the English original. By general consensus, monosyllabic words were systematically substituted by bisyllabic words, and trisyllabic words were substituted by tetrasyllabic words in the SP-CAT (e.g., ENG: *guilt* & *character* vs. SP: *pesar* 'regret' & *personaje* 'character'). Despite the crucial role of morphological complexity in highly inflected languages such as Spanish, the tasks involving repetition and reading of complex words (Tasks 13 & 21, respectively) did not need major adjustments (5/6 items remained unaltered), and the same derivational and inflectional morphemes as in the English CAT were used.

In the original CAT, spelling-sound regularity played an important role in Tasks 14 (*Repetition of Non Words*) and 23 (*Reading Non Words*). However, Spanish is an orthographically transparent language. Thus, the regularity of the correspondence between graphemes and phonemes limits the number of manipulations that can be exploited

for assessment purposes (Fyndanis et al., 2017). Accordingly, the complexity entailed by irregular orthographies was substituted by syllabic complexity (i.e., by the use of less frequent syllables such as those including vocalic and consonantal clusters). According to Guerra (1983), the most frequent syllable structures in Spanish are CV sequences (51.35%; e.g., *ca-sa* 'house'), followed by CVC (18.03%; e.g., *cam-po* 'field'), V (10.75%; e.g., *a-gui-la* 'eagle') and VC (8.60%, e.g., *an-cla* 'anchor'). Syllables, including vocalic or consonantal clusters, are rather infrequent (CVV: 3.37%, e.g., *sue-lo* 'floor'; CVVC: 3.31%, e.g., *pies* 'feet'; CCV: 2.96%, e.g., *tra-mo* 'stretch'; CCVC: 0.88%, e.g., *trac-tor* 'tractor'). Based on this classification, to adapt Tasks 14 and 23, we identified the less frequent consonant clusters in Spanish and included words adjusting to this criterion (e.g., *trimpo*).

Transparency is also key in the CAT written tasks (Tasks 24 - *Writing: Copying*, 25 - *Writing Picture Names*, & 26 - *Writing to Dictation*). However, in these cases, the regular/ irregular contrasts were resolved using two different strategies: a) employing items with a lower degree of transparency, including a silent 'h' (e.g., *hada* [a ða] 'fairy') and phonemes with two potential orthographic realizations (e.g., /x/: *jirafa* [xi 'ra fa] 'giraffe', *gerente* [xe 'ren te] 'manager'; /b/: *baca* ['ba ka] 'roof rack', *vaca* ['ba ka] 'cow'); and, b) including an item that in certain varieties of Spanish is susceptible to a mismatch between orthography and pronunciation (e.g., *taza* "cup" - pronunciation of /θ/ as /s/ as a consequence of *seseo*).

A final consideration for the SP-CAT linguistic adaptation was the selection of the verbs included in Task 18 (*Naming Actions*). The Spanish version depicts verbs belonging to the 1st and 2nd conjugation (verbs ending in -ar and -er, respectively), given that these are the most frequent and regular. Also, whereas in English the use of the gerund is expected in response to *What is the person doing?* (e.g., *eating*), in Spanish, the infinitive (e.g., *comer*), or the 3rd person singular present tense (e.g., *come*) are the most common responses.

3.2. Cultural adaptations of the SP-CAT

No matter how technically appropriate linguistic adjustments are, no test can be considered adapted without an exhaustive cultural review. One of the major challenges of the SP-CAT was to create an adaptation culturally suitable and engaging for a wide variety of Spanish-speaking countries. To clarify, we understand cultural adaptations as the development of items that are relevant to a given culture (in our case, some Latin American countries and Spain), respecting the underlying concepts that the target questions intend to assess (e.g., in line with Mirza et al., 2022). To do so, as in the case of the linguistic adaptation, we followed the WG-2 guidelines (Fyndanis et al., 2017; Martínez-Ferreiro et al., 2024), and DuBay and Watson's (2019) steps for cultural adaptations. For linguistic adaptations, these include: 1) reproducing the source instrument into the new target language, 2) pre-testing with members of the target population, and 3) carrying out a psychometric analysis of the new version. Given the diversity we had to cover, we began by considering aspects relevant to three geographically distant countries: Chile, Colombia, and Spain.

The completion of the first step towards the adaptation of the CAT involved multiple stages. Even the instructions and descriptions of each task, which may apparently look like mere translations of the original, had to be re-examined. For example, in Spanish, the pronoun *usted* is the courtesy form for the second person singular. This form, which

stands in opposition to *tú* 'you', requires 3rd person singular verb agreement (e.g., ENG: *Say*-2nd.sg *out loud* vs. SP: *Diga*-3rd.sg *en voz alta*). Hence, *you* was systematically substituted by *usted* in the instructions. However, since the use of the courtesy form has significantly decreased in certain Spanish-speaking territories in the past decades (e.g., Spain), speech and language therapists (SLTs) are encouraged to use the participants' preferred form (*tú* vs. *usted*) to promote engagement.

A critical stage to keep the test relevant for such a wide audience was the selection of culturally-relevant items for Spanish-speaking populations. This was especially visible in the case of individual items, which now include references to Latin American animals such as the *llama*. In Task 11 (*Comprehension of Spoken Paragraphs*), Spanish proper names of wide geographical distribution (*Sandra* and *Pablo*) were selected. Similarly, the original cities mentioned in this task were changed to *Santiago* and *Córdoba*, as there are cities with these names in several Spanish-speaking countries. Cities named Santiago can be found in Spain, Chile, Argentina, Honduras, Guatemala, Cuba, Colombia, Ecuador, Mexico, and Uruguay, among many others (including non-Spanish-speaking countries). *Miles* and *pounds* were also converted. Whereas *miles* were substituted by *kilometers*, multiple currencies are used across Spanish-speaking countries. Thus, reference to a specific monetary unit and amount was substituted by the expression *extensive material damage* (instead of "*over a million pounds*"). These changes rendered the paragraph equally applicable across Spanish-speaking countries. Also, the general consensus for all adaptations was that violent references should be consistently removed. Accordingly, we changed items depicting this type of actions (e.g., Tasks 9: from the ENG: *The butcher shoots the nurse*, to the SP: *El cocinero llama al doctor* 'The cook calls the doctor').

A major challenge in selecting culturally-relevant items was the difference found in vocabulary and, in some cases, pronunciation, across Spanish-speaking countries. In section 2.1, we mentioned the attempt to select unambiguous items. However, with such a huge portion of the globe to cover, this was not always possible. In an attempt to standardize the acceptable responses and ease the task of the test administrators, one distinctive feature of the SP-CAT was introduced: the PANs (from Panhispanic). When an item was susceptible to being produced differently in one of the Spanish variants, this was marked as PAN (e.g. *cerilla* [PAN01] 'match'). A list of the accepted PANs is provided at the end of each task to homogenize the scoring (e.g. [PAN01] *cerillo, fósforo* 'match'). The list was developed based on the responses obtained during the verification of name agreement for pictorial stimuli, which we expand on in Section 3 below. Regarding their presence in the test, most PANs are to be found among distractors. However, due to the characteristics of these words (generally, highly frequent, highly imaginable words), certain PANs have been included as targets (except for naming tasks). Overall, PANs represent 7% of the total target items at the word level. As clearly specified in the manual, PANs are accepted as correct answers. It is expected that more PANs will be added as a result of both the natural evolution of vocabulary and the increase of users over time.

Finally, in addition to the endeavor of developing culturally-relevant items, all the final items were drawn by a Chilean artist familiarized with the language and culture and with experience in creating stimuli for Spanish-speaking populations with aphasia.

3.3. *Scoring*

The scoring system of the SP-CAT comes from the original English version (Swinburn et al., 2004), with the exception of Tasks 19 and 27 (*Spoken* and *Written Picture Description*). For these two tasks, and following the general consensus of the Collaboration of Aphasia Trialists, a modified version of the Dutch model (Visch-Brink et al., 2014) was adopted (see Martínez Ferreiro et al., 2024 for an exhaustive discussion). The new scoring system focuses on two main aspects: a) content, that is, the presence and the exhaustiveness of the description of the 4 main elements of the picture (0-8 points), and b) the form in which the oral and written productions are made (0-9 points), including fluency (smoothness and speed of the speech output), grammatical complexity (length, extent of clausal embedding, degree of interrelationship across sentences), and grammaticality (well-formedness). As in the Norwegian version (Swinburn et al., 2021), fluency/tempo is only evaluated in Task 19 (*Spoken Picture Description*), as fluent writing may vary due to extra-linguistic reasons such as the availability of the dominant hand to complete the task (hemiplegia) or the educational and usage level of the speakers. This adaptation aimed at simplifying the analysis to reduce inter-rater variability and evaluation time, thus making the administration of the CAT easier.

4. Pre-testing items for the SP-CAT

Resuming the WG-2 guidelines, and in line with DuBay and Watson (2019), the second step towards adapting the CAT to Spanish was pre-testing the selected items with members of the target population. First, we tallied the psycholinguistic properties of the items, including name agreement of visual stimuli, imageability, and frequency rates.

4.1. *Name agreement of visual stimuli*

The CAT contains multiple black-and-white visual stimuli. As the Spanish adaptation required the creation of a significant number of new drawings, the decision was made to have them all redrawn, respecting the original style. Consequently, name agreement ratings had to be calculated for all pictures. In agreement with the guidelines established for the cross-linguistic adaptations of the CAT (Fyndanis et al., 2017), the threshold for name agreement was set at a mean of 90% correct identification across Spanish varieties (with a minimum of 85% for each individual dataset).

A total of 119 visual stimuli were evaluated by 237 native speakers of three distinct varieties of Spanish (n = 112 Chilean, n = 75 Colombian, and n = 50 European Spanish) ranging from 19 to 82 years of age (mean = 50.4, SD = 13.9). After giving consent, participants, recruited via snowball sampling, were asked to name a set of pictures presented by means of a questionnaire available both in-person and online. Pictures were tested in 4 rounds. A total of 75 respondents participated in round 1, which included 77 pictures. Visual materials not reaching the acceptability threshold were disregarded, and new visual materials were created. The same procedure was applied in round 2 (56

participants, 25 pictures), round 3 (63 participants, 16 pictures), and round 4 (43 participants, 1 picture). Participants did not receive any compensation for their participation. The final list of selected target items (n = 74) and their name agreement rates have been included in Appendix 2.

4.2. Lexical Frequency

Another psycholinguistic variable that needed to be controlled for during the adaptation of the CAT was frequency. Lexical frequency plays an important role in repetition, naming, reading, and writing tasks (Tasks 12, 17, 20, 25, and 26). The cutoff between low- and high-frequency items in the SP-CAT was set based on the *Corpus de Referencia del Español Actual* 'Corpus of Reference of Current Spanish' (CREA) developed by the Royal Spanish Academy (RAE). The CREA includes data from a wide variety of written and oral texts from several Spanish-speaking countries and provides different frequency lists (1,000, 5,000, 10,000 most frequent forms, and a total frequency list). The 10,000 most frequent word list was selected to establish the boundary between high and low frequency. Words included in this list were judged as highly frequent, whereas those not included were considered low-frequency words (See Appendix 3).

4.3. Imageability

Imageability ratings were also relevant for some CAT tasks (e.g., Tasks 12 & 17). The calculation of the imageability rates for Spanish items took place in two rounds and included both consultation of existing corpora and the collection of new data. First, a preliminary group of 20 participants from Spain (4 male, mean age = 22.6, SD = 4.42) were recruited and asked to rank the degree of imageability of a preliminary list of 256 items on a 5-point scale. Given that most items had to be depictable, the cutoff point for high vs. low imaginable words was set at more than 4.5 on a 5-point Likert scale. The preliminary item selection of the SP-CAT was made based on these results (see Rofes et al., 2018). However, for the final item selection, following the same procedure as in the name agreement phase, we expanded the sample by running a second round of questionnaires addressed to speakers of three distinct varieties of Spanish (Chilean, Colombian, and European Spanish). Overall, 244 respondents (124 participants from Chile, 67 from Colombia and 53 from Spain) ranging from 19 to 82 years of age (mean = 49.6, SD = 14.9) participated in the four rounds of imageability judgments, which included 89 words (Round 1: 75 participants, 33 words; Round 2: 64 participants, 28 words; Round 3: 71 participants, 26 words; Round 4: 34 participants, 2 words). Given that items had to be depictable, and due to the demanding constraints imposed by name agreement requirements, the cutoff point for high vs. low imaginable words was again set at more than 4.5 on a 5-point Likert scale. The final list of target items included in the Spanish CAT that had to be controlled for imageability (n = 73) and their characteristics have been included in Appendix 3.

   As a further control measure and to compensate for the lack of our own imageability data for a few late incorporation items (14/73 items), our results were completed with those from an existing database containing norms for affective and lexico-semantic variables for 1,400 Spanish words (Guasch, Ferré & Fraga, 2016). Contrary to our own

data set, Guasch and collaborators measured imageability on a 7-point Likert scale. For this study, the cutoff point for high and low imageability was set at more than 5.5 for the data points retrieved from this list.

5. The SP-CAT Pilot Study across Spanish varieties

Contrary to most cross-linguistic adaptations, an additional usability pretest was necessary in the case of the SP-CAT to rule out differences across Spanish-speaking countries. In order to assess the adequacy of the adapted version, we conducted a pilot study with a total of 82 neurotypical representative users who consented to participate in the study online or in person, 44 (53.7%) from Chile, 28 (34.1%) from Colombia, and 10 (12.2%) from Spain. As in previous phases, recruitment took place via snowball sampling, and the participants did not receive compensation. Participation in the name agreement and imageability phases was a criterion for exclusion, along with speech or language related conditions. Descriptive statistics of participants' sociodemographic characteristics and test results were calculated for the total sample and by country. The participants' mean age was 53.9 years (SD = 8.0), and 57.3% were female. On average, the sample reported 13.2 years of formal education (SD = 3.9). Comparisons to identify possible differences across countries were conducted in a Bayesian framework through JASP (2023), reporting the Bayesian factor (BF). Notably, the Bayes Factor (BF) measures how likely the data are to be observed under the assumption of no differences between the compared groups. Unlike p-values, there are no specific cutoff values for BF. Smaller BF values generally suggest differences between groups, while values greater than 1 indicate support for similarity in the values of the compared groups. For example, if the BF equals 4, the evidence in the observed data is four times stronger in favor of similar groups compared to differences between them (Hoijtink et al., 2019). Nonetheless, deviating from traditional statistical methods, the Bayes Factor does not provide a straightforward binary decision (reject or not reject H0); instead, it assesses the extent of support within the data for the hypotheses related to differences between groups (countries in this study).

Overall, participants from Spain were older than participants from Chile and Colombia. Also, participants from Colombia reported more extensive formal education than participants from Chile and Spain. A description of sociodemographic characteristics is shown in Table 1.

Table 1. Sociodemographic characteristics of the sample.

Variable	Total (N=82)	Chile (N=44)	Colombia (N=28)	Spain (N=10)	BF
Age					
Mean (SD)	53.9 (8.0)	54.3 (7.0)	50.6 (8.4)	61.3 (6.1)	0.033
Median [Min, Max]	54.0 [40.0, 71.0]	54.5 [40.0, 71.0]	49.5 [40.0, 71.0]	60.5 [51.0, 71.0]	
Sex					
Female	47 (57.3 %)	23 (52.3 %)	19 (67.9 %)	5 (50.0 %)	0.226
Male	35 (42.7 %)	21 (47.7 %)	9 (32.1 %)	5 (50.0 %)	
Education (years)					
Mean (SD)	13.2 (3.9)	11.9 (3.7)	15.3 (3.5)	12.7 (3.8)	0.020
Median [Min, Max]	13.0 [4.00, 22.0]	12.0 [4.00, 19.0]	16.0 [5.00, 22.0]	12.5 [7.0, 21.0]	

Note. BF represents the Bayes factor supporting no differences between countries (null model). Therefore, small values support the hypothesis of differences between countries.

Table 2 summarizes the statistical differences found across tasks (see Figure 1 for density distribution of the performance by task and country). The scores for all tasks by country are included in Appendix 4.

Table 2. Summary of tasks with low Bayes Factor (BF) values.

Task	Total (N=82)	Chile (N=44)	Colombia (N=28)	Spain (N=10)	BF
Gesture object use					
Mean (SD)	10.9 (0.9)	11.0 (0.7)	10.4 (1.1)	11.8 (0.4)	0.006
Median [Min, Max]	11.0 (8.0, 12.0)	11.0 (9.0, 12.0)	10.0 (8.0, 12.0)	12.0 (11.0, 12.0)	
Comprehension spoken sentences					
Mean (SD)	29.0 (3.1)	27.8 (3.3)	30.8 (1.5)	29.7 (2.6)	0.003
Median [Min, Max]	30.0 (17.0, 32.0)	28.0 (17.0, 32.0)	31.5 (28.0, 32.0)	30.5 (24.0, 32.0)	
Repetition of words					
Mean (SD)	31.4 (1.3)	30.8 (1.7)	32.0 (0.2)	31.9 (0.3)	0.015
Median [Min, Max]	32.0 (26.0, 32.0)	32.0 (26.0, 32.0)	32.0 (31.0, 32.0)	32.0 (31.0, 32.0)	
Spoken picture description					
Mean (SD)	15.0 (1.7)	15.4 (1.2)	13.9 (2.1)	16.4 (1.1)	0.001
Median [Min, Max]	15.0 (11.0, 17.0)	15.0 (13.0, 17.0)	13.5 (11.0, 17.0)	17.0 (14.0, 17.0)	
Reading aloud words					
Mean (SD)	47.6 (0.9)	47.3 (1.1)	48.0 (0.0)	48.0 (0.0)	0.041
Median [Min, Max]	48.0 (44.0, 48.0)	48.0 (44.0, 48.0)	48.0 (48.0, 48.0)	48.0 (48.0, 48.0)	
Reading nonwords					
Mean (SD)	9.7 (0.9)	9.4 (1.1)	10.0 (0.0)	10.0 (0.0)	0.071
Median [Min, Max]	10.0 (6.0, 10.0)	10.0 (6.0, 10.0)	10.0 (10.0, 10.0)	10.0 (10.0, 10.0)	

Note. BF represents the Bayes factor supporting no differences between countries (null model). Therefore, small values support the hypothesis of differences between countries.
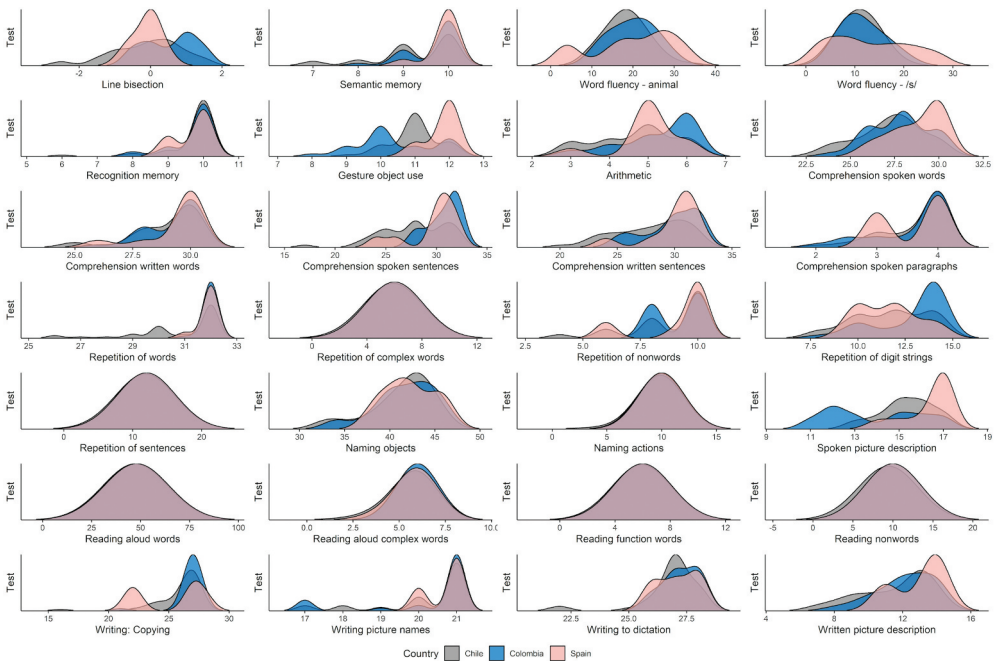

Figure 1. Density distribution of the performance by task and country. *Note.* This figure presents a smooth curve that shows the probability density function, not the observed values. The x-axis shows scores in each test.

Out of the 27 tasks, 5 showed statistical differences. Upon close scrutiny, these differences turned out not to be clinically significant, as we elaborate on below. The results of task 5 (Gesture object use), task 12 (Repetition of words), task 20 (Reading aloud words), and task 23 (Reading words) differ in performance in less than 1 item. These differences are thus due to zero variability in some task scores in Colombia and Spain. For instance, the maximum score in task 23 is 10 points (two points per item). Whereas Colombian and Spanish participants all scored 10, Chilean participants had an average score of 9.4 due to a few participants scoring less than 10, thus leading to statistical differences but without a clinically meaningful effect on the test.

Similarly, in task 9 (Comprehension of spoken sentences), there is a difference of 1.5 items between the Chilean and the Colombian samples, with results from Spain falling in the middle. For this task, the score for people without aphasia in the English version ranges from 26-32 points. Thus, we do not consider the difference of 1.5 items to be clinically significant. Lastly, as in other adaptations of the CAT, differences were found in task 19 (Spoken picture description). Colombian participants showed lower scores than Chilean and Spanish participants (13.9 out of a maximum of 17 points). Performance in this task is classified into 4 categories: absent, poor, moderate, or good. A score above 12.45 would fall within the category of 'good'. In other words, a difference of 2.5 is not clinically significant as it puts all participants within the same category.

6. Discussion

In the past decade, the field of aphasiology has begun to show an increased awareness of the importance of having appropriate linguistically and culturally adapted assessment tools designed for people with aphasia of diverse backgrounds (e.g., Tsapkini et al., 2010; Khatoonabadi et al., 2015; Muò et al., 2021). Although there is still a lot of terrain to cover, initiatives such as the one from the WG-2 of the Collaboration of Aphasia Trialists, with the adaptation of the CAT to multiple cultures and languages, have a significant impact on SLTs' practice and, more importantly, in the access to adequate assessments for people with aphasia who speak the diverse languages of the world. In what follows, we focus on three challenges that deserve particular: attention: time management, financial support, and multilingual societies.

Challenge 1: Time management. Completing the multiple phases required to thoroughly adapt an assessment tool was an overwhelming task. This was particularly challenging given that languages such as Spanish have historical and geographical factors that have led to rich dialectal varieties. To provide a sense of complexity and resource-intensive work, the research team began to meet regarding the SP-CAT adaptations in 2015. Since then, the team extensively discussed the necessary linguistic and cultural adaptations of aphasia assessments for Spanish-speaking populations. Within the linguistic adaptations, we began discussing typological differences between English and Spanish, differences between Spanish varieties, Spanish *seseo*, unambiguous items, compounds, gender cues, spelling-sound regularities, transparency, and multiple other syntax-related considerations. The use of the formal second person pronouns generated interesting discussions showing the different evolution routes taken by American and European dialects. Similarly, in 2015, the research team began discussing aspects of the cultural adaptation, such as selecting culturally-relevant items and images (e.g., shared

place names, images depicting familiar physical traits, neutral clothing, and animals or objects from both sides of the ocean to avoid geographical biases), and using PANs to cover different Spanish varieties.

Gathering data for name agreement and imageability started in 2016 and was finalized in 2019 when the final selection of items passed the desired linguistic properties. As soon as this step was completed and the test was assembled for its administration, we began planning on data collection for the pilot study, which started in 2020, with the beginning of the global pandemic. Data collection was slow while the teams in the different countries adopted biosecurity protocols; still, our pilot results show that performance is consistent across geographically distant varieties of Spanish, with differences restricted to effects without clinical significance. Specifically, whereas 21/27 tasks showed no statistical differences, in 4/27 tasks, differences were equal to less than 1 item. The remaining two tasks also fall within the expected ranges for subjects without language disorders. This implies that the exhaustive examination of the different varieties of a language allows the creation of linguistically and culturally adequate wide range tools suitable to respond to the needs of a greater number of individuals.

Challenge 2: Financial support. Besides an extensive timeline, funding was another challenge for the SP-CAT adaptation. No specific grant supported the SP-CAT adaptation, with the exception of specific funding for the creation of the visual stimuli. Thus, researchers used their limited time and invested out of their finances to fund every project stage. Our experience seems not to differ from other researchers whose English is not their primary language, who have shown an increased effort (in time and finances) in conducting scientific activities, from reading papers to disseminating study results (Amano et al., 2023). Despite the challenges, it is necessary to continue the efforts of developing aphasia assessments (and treatments; see Forero et al., 2023) for Spanish speakers, to meet the needs of this population around the world.

Challenge 3: Multilingual societies. Speech-language therapists will continue to experience increasing caseloads of people with aphasia who speak a different language than them, especially in the current times, characterized by significant migration waves. However, SLTs report a lack of confidence when working with people with aphasia who do not share their language (Larkman et al., 2022). They also report a lack of assessment tools appropriate for their language and contexts (e.g., Badar et al., 2021). This issue remains even when using interpreter services, given documented issues such as miscategorization of language errors, omission of information for aphasia diagnosis, and overall compromised validity of aphasia assessments (e.g., Babbit et al., 2022; Roger & Code, 2011, 2020; Kambanaros & van Steenbrugge, 2004). Thus, initiatives such as the one reported in this paper are critical to meet the needs of *clinicians* working with Spanish speakers (and speakers of multiple other languages) around the world.

Communication is a human right (see Hersh, 2018), which is withheld when people with aphasia are not assessed in their primary language. The linguistic and cultural adaptation of the Comprehensive Aphasia Test to Spanish (SP-CAT) provides a step forward in ensuring that Spanish speakers with aphasia are assessed in their primary language with a tool that is not only comprehensive, but also informative in terms of classification of correct/incorrect responses and rating for their contexts. The introduction of the PANs contributes to making the test more inclusive across socio-demographic groups and favors consistency during the assessment process. Although more PANs will need to be added when

considering other Spanish-speaking countries, these guarantee that Spanish variety and register will not be penalized. We are currently working on the SP-CAT normative data taking all these aspects into account. Our next step is to report on psychometrics, including aspects such as validity, sensitivity, and reliability, focusing on a larger neurotypical sample as well as a group of people with aphasia from the same geographical areas.

7. Conclusion

This contribution is an invitation to the aphasia clinical community, and the greater scientific community in communication disorders, to reflect upon the meaning of equitable assessment measures and the critical role they play in adequately diagnosing people with aphasia. We have exhaustively reported the procedure followed for the linguistic and cultural adaptation of the cognitive and language batteries of the CAT to Spanish across a variety of geographical and dialectal borders. The SP-CAT has been shown to be linguistically and culturally adequate for different varieties of Spanish in Europe and Latin America, throwing comparable results across neurotypical speakers of Chile, Colombia, and Spain. In doing so, we have also contributed new imageability data for Spanish words, which can be used by a broader research community. The next phase will define the psychometric properties of the SP-CAT, including its validity, sensitivity, specificity, accuracy, test-retest reliability, internal consistency, and inter-rater reliability, to fulfill our aim to provide Spanish speakers with additional appropriate testing materials and paving the way towards equitable care.

Disclosure statement

Funding

ORCID

Silvia Martínez Ferreiro ⓘ http://orcid.org/0000-0003-2393-1214
Yina M. Quique ⓘ http://orcid.org/0000-0001-9129-4742
Viviana Alejandra Rodríguez ⓘ http://orcid.org/0000-0001-8487-0676
Carolina Méndez Orellana ⓘ http://orcid.org/0000-0001-5217-9725

References

Alonso, M. A., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior research methods*, *47*(1), 268–274. https://doi.org/10.3758/s13428-014-0454-2

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J. D., Montaño-Centellas, F., Paudel, K., White, R. L., & Veríssimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLoS biology*, 21(7), e3002184. https://doi.org/10.1371/journal.pbio.3002184

Babbitt, E. M., Ginsberg-Jaeckle, M., Larkin, E., Escarcega, S., & Cherney, L. R. (2022). Classifying Interpreter Behaviors During Aphasia Assessments: Survey Results and Checklist Development. *American journal of speech-language pathology*, 31(5S), 2329–2347. https://doi.org/10.1044/2022_AJSLP-21-00306

Badar, F., Naz, S., Mumtaz, N., Babur, M. N., & Saqulain, G. (2021). Aphasia-Diagnostic Challenges and Trends: Speech-Language Pathologist's Perspective. *Pakistan Journal of Medical Sciences*, 37(5). https://doi.org/10.12669/pjms.37.5.2314

Beveridge, M. E. L., & Bak, T. H. (2011). The languages of aphasia research: Bias and diversity, Aphasiology, 25(12), 1451–1468. https://doi.org/10.1080/02687038.2011.624165

Cuetos, F., & Alija, M. (2003). Normative data and naming times for action pictures. *Behavior Research Methods, Instruments & Computers*, *35*(1), 168–177. https://doi.org/10.3758/BF03195508

DuBay, M., & Watson, L. R. (2019). Translation and cultural adaptation of parent-report developmental assessments: Improving rigor in methodology. *Research in Autism Spectrum Disorders*, *62*, 55–65. https://doi.org/10.1016/j.rasd.2019.02.005

Duchon, A., Perea, M., Sebastián-Gallés, N. *et al*. (2013). EsPal: One-stop shopping for Spanish word properties. *Behav Res* 45, 1246–1258. https://doi.org/10.3758/s13428-013-0326-1

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2019). Summary by language size. *Ethnologue*. SIL International.

Eddington, D. E. (2022). Grouping Spanish-speaking countries by dialect: An exploratory corpus dialectometric approach. Isogloss. *Open Journal of Romance Linguistics*, *8*(1/*9*), 1–30. https://doi.org/10.5565/rev/isogloss.207

Forero, L., Bernal, M., Aguilar, O., Quique, Y. (2023). Aphasia Treatments for Spanish Speakers With Aphasia: A Scoping Review. *Revista de Investigación en Logopedia*, 13(1), e81535. (available in Spanish)

Fyndanis, V., Lind, M., Varlokosta, S., Gram Simonsen, H., Kambanaros, M., Ceder, K., Rofes, A., Soroli, E., Bjekic, J., Gavarró, A., Grohmann, K., Kuvac, J., Martinez Ferreiro, S., Munarriz, A., Vuksanovic, J., Zakarias, L., & Howard, D. (2017). Cross-linguistic adaptations of The Comprehensive Aphasia Test: Challenges and solutions. *Clinical linguistics & phonetics*, *31*(7–9), 697–710. https://doi.org/10.1080/02699206.2017.1310299

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behav Res Methods*, *48*(4), 1358–1369. https://doi.org/10.3758/s13428-015-0684-y

Guerra, R. (1983). Estudio estadístico de la sílaba en español. En M. Esgueva y M. Cantarero (Eds.), *Estudios de fonética I*. (pp. 9–112). Consejo Superior de Investigaciones Científicas.

Hersh D. (2018). From individual to global: Human rights and aphasia. International journal of speech-language pathology, 20(1), 39–43. https://doi.org/10.1080/17549507.2018.1397749

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychol Methods.*, *24*(5), 539–556. https://doi.org/10.1037/met0000201

Howard, D., Swinburn, K., & Porter, G. (2010). Putting the CAT out: What the Comprehensive Aphasia Test has to offer. Aphasiology, 24(1), 56–74. https://doi.org/10.1080/02687030802453202

Instituto Cervantes. (2020). El español: una lengua viva – Informe 2020 (Report). Retrieved 14 December 2020.

Ivanova, M. V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. *Aphasiology*, *27*(8), 891–920.

JASP Team (2023). JASP (Version 0.17.2) [Computer software].

Kambanaros, M., & van Steenbrugge, W. (2004). Interpreters and language assessment: Confrontation naming and interpreting. *Advances in Speech-Language Pathology*, 6(4), 247–252. https://doi.org/10.1080/14417040400010009

Kertesz, A., Pascual-Leone, P., & Pascual-Leone, G. (1990). *Western Aphasia battery en versión y adaptación castellana*. Valencia: Nau Libres.

Khatoonabadi, A. R., Nakhostin-Ansari, N., Piran, A., & Tahmasian, H. (2015). Development, cross-cultural adaptation, and validation of the Persian Mississippi Aphasia Screening Test in patients with post-stroke aphasia. *Iran J Neurol*, *14*(2), 101–107.

Larkman, C. S., Mellahn, K., Han, W., & Rose, M. L. (2022). Aphasia rehabilitation when speech pathologists and clients do not share the same language: A scoping review. *Aphasiology*, 1–23. https://doi.org/10.1080/02687038.2022.2035672

Lenz, A. S., Gómez Soler, I., Dell'Aquila, J., & Uribe, P. M. (2017). Translation and Cross-Cultural Adaptation of Assessments for Use in Counseling Research. *Measurement and Evaluation in Counseling and Development*, *50*(4), 224–231. https://doi.org/10.1080/07481756.2017.1320947

Lipski, J. M. (1994). *Latin American Spanish*. Longman.

Lipski, J. M. (1996). *El español de América*. Cátedra.

Martínez-Ferreiro, S., Arslan, S., Fyndanis, V., Howard, D., Kuvač Kraljević, J., Matić Škorić, A., Munarriz-Ibarrola, A., Norvik, M., Peñaloza, C., Pourquie, M., Simonsen, H. G., Swinburn, K., Varlokosta, S. & Soroli, E. (2024, preprint). Guidelines and recommendations for cross-linguistic aphasia assessment: A review of 10 years of Comprehensive Aphasia Test adaptations. https://osf.io/t6gbj/

Maviş, İ, Müge Tunçer, A., Selvi-Balo, S., Dilara Tokaç, S. & Özdemir, S. (2022). The adaptation process of the Comprehensive Aphasia Test into CAT-Turkish: psycholinguistic and clinical considerations. *Aphasiology*, *36*(4), 493–51.

Mirza, N., Waheed, M. W., & Waheed, W. (2022). A new tool for assessing the cultural adaptation of cognitive tests: demonstrating the utility of the Manchester Translation Evaluation Checklist (MTEC) through the Mini-Mental State Examination Urdu. *BJPsych Open*, *9*(1), e5. https://doi.org/10.1192/bjo.2022.620

Muò, R., Raimondo, S., Martufi, F., Cavagna, N., Bassi, M., & Schindler, A. (2021). Cross-cultural adaptation and validation of the Italian Aachener Aphasie Bedside Test (I-AABT), a tool for Aphasia assessment in the acute phase. *Aphasiology*, *35*(9), 1238–1261. https://doi.org/10.1080/02687038.2020.1819953

Olabarrieta-Landa, L., Rivera, D., Galarza-Del-Angel, J., Garza, M. T., Saracho, C. P., Rodríguez, W., Chávez-Oliveros, M., Rábago, B., Leibach, G., Schebela, S., Martínez, C., Luna, M., Longoni, M., Ocampo-Barba, N., Rodríguez, G., Aliaga, A., Esenarro, L., García de la Cadena, C., Perrin, B. P., & Arango-Lasprilla, J. C. (2015). Verbal fluency tests: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, *37*(4), 515–561. https://doi.org/10.3233/NRE-151279

Paillard, M. (2000). *Lexicologie contrastive français-anglais. Formation des mots et construction du sens*. Gap: Ophrys.

Real Academia Española. Banco de datos (CREA) [on line]. Corpus de referencia del español actual. [Retrieved October 2017, October 2020].

Rofes, A., Zakariás, L., Ceder, K., Lind, M., Blom Johansson, M., De Aguiar, V., Bjekić, J., Fyndanis, V., Gavarró, A., Gram Simonsen, H., Hernández Sacristán, C., Kambanaros, M., Kuvač Kraljević, J.,

Martínez-Ferreiro, S., Mavis, I., Méndez Orellana, C., Salmons, I., Sör, I., Lukács, A., Tunçer, M., Vuksanovic, J., Munarritz Ibarrola, A., Pourquie, M., Varlokosta, S., & Howard, D. (2018). Imageability ratings across languages. *Behavior Research Methods*, *50*(3), 1187–1197. https://doi.org/10.3758/s13428-017-0936-0

Roger, P., & Code, C. (2011). Lost in translation? Issues of content validity in interpreter-mediated aphasia assessments. I*nternational journal of speech-language pathology*, 13(1), 61–73. https://doi.org/10.3109/17549507.2011.549241

Roger, P., & Code, C. (2020). Interpreter-mediated aphasia assessments: Mismatches in frames and professional orientations. Communication & medicine, 15(2), 233–244. https://doi.org/10.1558/cam.38680

Ronê, P., Triguero Veloz Teixeira, M. C., Cantiere, C., Efstratopoulou, M., & Carreiro, L. R. (2019). Translation and cross-cultural adaptation of the Motor Behavior Checklist (MBC) into Brazilian Portuguese. Trends in Psychiatry and Psychotherapy, 41. https://doi.org/10.1590/2237-6089-2017-0104

Salmons, I., Rofes, A., & Gavarró, A. (2021). Prova integral d'afàsia. Llibre d' ítems. Bellaterra: Servei de Publicacions de la UAB.

Swinburn, K., Best, W., Beeke, S., Cruice, M., Smith, L., Pearce Willis, E., Ledingham, K., Sweeney, J. & McVicker, S. J. (2019). A concise patient reported outcome measure for people with aphasia: the aphasia impact questionnaire 21. *Aphasiology*, *33*(9), 1035–1060. https://doi.org/10.1080/02687038.2018.1517406

Swinburn, K., Porter, G., & Howard, D. (2004). *The Comprehensive Aphasia Test*. Hove, UK: Psychology Press.

Swinburn, K., Porter, G., & Howard, D. (2014). *Comprehensive Aphasia Test* (L. Haaber Hansen & M. Kaae Frederiksen, adaptation). Copenhagen, Denmark: Dansk Psykologisk Forlag.

Swinburn, K., Porter, G., & Howard, D. (2023). *The Comprehensive Aphasia Test (2nd edition)*. Routledge.

Swinburn, K., Porter, G., Howard, D., Høeg, N., Norvik, M., Røste, I., & Simonsen, H. G. (2021). *CAT-N – Comprehensive Aphasia Test*. Novus Forlag.

Swinburn, K., Porter, G., Howard, D., Kuvač Kraljević, J., Lice, K., & Matić, A. (2020). *Sveobuhvatni test za procjenu afazije CAT-HR*. [Comprehensive Aphasia Test - Croatian version]. Naklada Slap.

Tsapkini, K., Vlahou, C. H., & Potagas, C. (2010). Adaptation and validation of standardized aphasia tests in different languages: Lessons from the Boston Diagnostic Aphasia Examination - Short Form in Greek. Behav Neurol, 22(3–4), 111–119. https://doi.org/10.3233/ben-2009-0256

Valle, F., & Cuetos F. (1995). Evaluación del Procesamiento Lingüístico en la Afasia (EPLA). Hove, UK: Lawrence Eribaum Associates. (Adaptación española de Kay, J., Lesser, R. y Coltheart, M. [1992]. Psycholinguistic Assessments of Language Processing in Aphasia [PALPA]. Psychology Press).

Vigliecca, N. S., Peñalva, M. C., Castillo, Molina, S. C., Voos, J. A., Ortiz, M. et al. (2011). Brief aphasia evaluation (minimum verbal performance): concurrent and conceptual validity study in patients with unilateral cerebral lesions. *Brain Injury*, *25*(4), 394–400. https://doi.org/10.3109/02699052.2011.556106

Visch-Brink, E., Vandenborre, D., de Smet, H. J., & Mariën, P. (2014). *Comprehensive Aphasia Test - Nederlandse bewerking - Handleiding*. Amsterdam, The Netherlands: Pearson.

Zakariás, L. & Lukács, A. (2022). The Comprehensive Aphasia Test–Hungarian: adaptation and psychometric properties. *Aphasiology*, *36*(9), 1127–1145, https://doi.org/10.1080/02687038.2021.1937921

Appendix 1: From the English CAT to the SP-CAT

COGNITIVE BATTERY

Task	Adaptation
1. Line bisection	Same as original
2. Semantic memory	6/10 target items preserved
3. Word fluency	Same as original
4. Recognition memory	7/10 target items preserved
5. Gesture object use	Same as original
6. Arithmetic	Same as original

LANGUAGE BATTERY

a. Language comprehension

- Comprehension of spoken language:

7. Comprehension of spoken words	N° of features & position changes
9. Comprehension of spoken sentences	Gender & word order
11. Comprehension of spoken paragraphs	Cultural adaptation

- Comprehension of written language:

8. Comprehension of written words	N° of features & position changes
10. Comprehension of written sentences	Gender & word order

b. Expressive language:

- Repetition

12. Repetition of words	Word length
13. Repetition of complex words	Same as original
14. Repetition of non-words	Consonant clusters
15. Repetition of digit strings	Same as original
16. Repetition of sentences	6/8 target items preserved

- Spoken language production

17. Naming objects	Word length
18. Naming actions	Infinitives

- Spoken picture production

19. Spoken picture description	Picture + Scoring system

- Reading aloud

20. Reading words	Word length
21. Reading complex words	2/3 target items preserved
22. Reading function words	Same as original
23. Reading non-words	Consonant clusters

- Writing

24. Writing: Copying	Orthographic transparency
25. Writing picture names	Orthographic transparency
26. Writing to dictation	Orthographic transparency

- Written picture description

27. Written picture description	Picture + Scoring system

Appendix 2: Name agreement SP-CAT

Word	Task	Item	Chile	Colombia	Spain	Name agreement average	Total N
silbato	2	1	0.97	0.91	1	0.96	75
mano	2	2	1	1	1	1	75
cerillas	2	3	0.97	0.96	1	0.98	75
almohada	2	4	0.97	1	1	0.99	74
reloj	2	6	1	0.96	1	0.99	75
monja	2	7	0.95	0.91	1	0.95	74
tienda	2	8	1	0.95	1	0.98	74
botella	2	9	1	0.96	1	0.99	75
flor	2	10	0.92	0.87	1	0.93	75
casa	7	1	1	1	1	1	74
cuerno	7	3	0.97	1	0.92	0.97	75
bota	7	5	1	1	0.92	0.97	75
copa	7	6	0.97	1	1	0.99	75
foca	7	8	1	0.91	1	0.97	70
gato	7	9	1	1	1	1	75
oveja	7	10	0.97	0.96	1	0.98	74
rata	7	13	1	1	1	1	75
mesa	8	3	1	1	1	1	75
carta	8	4	1	0.96	1	0.99	74
vela	8	7	0.97	0.96	1	0.98	75
rama	8	8	1	0.91	0.92	0.95	75
toro	8	9	0.97	1	0.92	0.97	75
hueso	8	10	1	0.91	1	0.97	73
lima	8	11	0.9	0.91	0.92	0.91	75
cama	8	13	0.95	1	1	0.98	75
fuente	8	14	1	0.95	1	0.98	74
barco	8	15	0.92	0.91	1	0.95	75
teléfono	17	1	1	0.96	1	0.99	75
reloj	17	2	0.97	0.91	1	0.96	75
microscopio	17	6	0.92	0.87	1	0.93	74
lupa	17	7	1	1	1	1	75
pera	17	8	1	1	1	1	75
bicicleta	17	11	1	0.91	1	0.97	75
elefante	17	13	1	1	1	1	75
barril	17	18	1	1	1	1	75
escalera	17	19	0.95	1	1	0.98	75
llama	17	21	0.95	0.91	0.92	0.93	75
murciélago	17	22	1	0.96	1	0.99	75
leer	18	1	1	0.87	1	0.96	75
barrer	18	2	1	0.91	1	0.97	75
pintar	18	3	0.97	0.87	0.92	0.92	75
recortar	18	5	1	1	1	1	75
rey	25	1	0.97	1	1	0.99	75
oso	25	2	1	1	1	1	75
ancla	25	4	1	0.96	1	0.99	75
jirafa	25	5	1	0.96	1	0.99	75
egipcio	2	5	0.85	1	0.92	0.92	56
pala	7	2	1	1	1	1	56
sierra	7	4	1	1	1	1	56
cuna	7	11	1	1	1	1	56

(Continued).

Word	Task	Item	Chile	Colombia	Spain	Name agreement average	Total N
maleta	7	12	1	1	1	1	56
juego	7	15	1	1	1	1	55
dedal	8	1	0.88	1	1	0.96	55
sable	8	2	1	1	1	1	53
cocina	8	5	0.96	1	1	0.99	56
pata	8	12	1	1	1	1	56
submarino	17	3	0.88	1	1	0.96	56
semáforo	17	4	1	1	1	1	56
pulpo	17	5	1	1	1	1	55
raíz	17	14	1	1	1	1	56
calendario	17	16	0.96	0.94	1	0.97	56
termómetro	17	23	1	1	1	1	55
túnel	17	24	1	1	1	1	56
nadar	18	4	1	1	1	1	55
muro	7	14	0.92	1	1	0.97	63
espátula	17	17	1	1	0.93	0.98	62
dinosaurio	17	9	1	0.91	1	0.97	61
timón	17	10	0.97	1	1	0.99	62
búho	17	20	0.97	1	0.93	0.97	63
pavo	25	3	0.97	1	0.93	0.97	63
cisne	17	15	0.95	1	1	0.98	63
rana	7	7	1	1	1	1	31
ostra	17	12	0.89	0.74	0.93	0.85	75
espina	8	6	0.97	0.94	1	0.97	68

Note: The values on this table range from 0 to 1, with 1 representing 100% agreement. The columns Chile, Colombia, and Spain reflect the name agreement found in each country. The name agreement average column represents the average name agreement across countries. The total N column represents the number of respondents across countries for each item.

Appendix 3: Imagebaility & Frequency SP-CAT

Word	Task	Item	Imageability Value	Imageability Source	Imageability for the SP-CAT	Lexical Frequency for the SP-CAT	RAE Lexical frequency
patín	12	1	4.6	pre-testing	H	L	NA
presidente	12	2	3.65	pre-testing	L	H	598.78
desdén	12	3	1.65	pre-testing	L	L	NA
radio	12	4	4.9	pre-testing	H	H	126.31
pesar	12	5	3.42	pre-testing	L	H	206.05
ministerio	12	6	2.6	pre-testing	L	H	155.88
planta	12	8	4.6	pre-testing	H	H	69.4
personaje	12	10	3.5	pre-testing	L	H	68.03
marinero	12	12	4.15	pre-testing	L	L	NA
evidencia	12	13	1.6	pre-testing	L	H	37.63
sacacorchos	12	14	4.9	pre-testing	H	L	NA
cara	12	15	4.8	pre-testing	H	H	181.93
castor	12	16	4.6	pre-testing	H	L	NA
servilleta	12	7	6.9	Guasch	H	L	NA
rectángulo	12	9	6.86	Guasch	H	L	NA
quietud	12	11	3.04	Guasch	L	L	NA
reloj	17	2	4.9	pre-testing	H	H	34.07
pera	17	8	4.85	pre-testing	H	L	NA
elefante	17	13	4.85	pre-testing	H	L	NA
bicicleta	17	11	4.95	pre-testing	H	H	14.7
teléfono	17	1	4.9	pre-testing	H	H	82
lupa	17	7	4.59	pre-testing	H	L	NA
barril	17	18	4.6	pre-testing	H	L	NA
llama	17	21	4.28	pre-testing	L	L	97.68
murciélago	17	22	4.6	pre-testing	H	L	NA
submarino	17	3	4.23	pre-testing	L	L	NA
semáforo	17	4	4.66	pre-testing	H	H	NA
pulpo	17	5	4.52	pre-testing	H	L	NA
raíz	17	14	4.36	pre-testing	L	H	45.28
calendario	17	16	4.5	pre-testing	L	H	21.92
espátula	17	17	4.33	pre-testing	L	L	NA
termómetro	17	23	4.5	pre-testing	L	L	NA
túnel	17	24	4.47	pre-testing	L	H	97.68
dinosaurio	17	9	4.18	pre-testing	L	L	NA
timón	17	10	3.96	pre-testing	L	L	NA
búho	17	20	4.25	pre-testing	L	L	NA
ostra	17	12	4.14	pre-testing	L	L	NA
microscopio	17	6	6.82	Guasch	H	L	NA
cisne	17	15	6.68	Guasch	H	L	NA
escalera	17	19	6.82	Guasch	H	H	26.34
contenedor	20	1	4.75	pre-testing	H	L	NA
bailarina	20	16	4.8	pre-testing	H	L	NA
brillar	20	21	3.95	pre-testing	L	L	NA
chocolate	20	24	4.9	pre-testing	H	H	16.68
culpa	20	13	2.55	pre-testing	L	H	19.69
diccionario	20	20	4.74	pre-testing	H	H	13.92
educación	20	14	3.37	pre-testing	L	H	172.57
estropajo	20	10	4.75	pre-testing	H	L	NA
pasaporte	20	6	4.6	pre-testing	H	H	10.21

(Continued).

Word	Task	Item	Imageability Value	Imageability Source	Imageability for the SP-CAT	Lexical Frequency for the SP-CAT	RAE Lexical frequency
pasar	20	15	2.9	pre-testing	L	H	157.63
pierna	20	4	4.8	pre-testing	H	H	25.31
porcelana	20	22	3.55	pre-testing	L	L	NA
siglo	20	18	2.15	pre-testing	L	H	274.58
sonajero	20	7	4.65	pre-testing	H	L	NA
televisión	20	19	4.85	pre-testing	H	H	156.91
tomar	20	9	3.6	pre-testing	L	H	134.32
zorro	20	8	4.55	pre-testing	H	L	NA
causa	20	11	2.25	pre-testing	L	H	152.61
goce	20	2	2.77	Guasch	L	L	NA
micrófono	20	3	6.91	Guasch	H	L	NA
trecho	20	5	3	Guasch	L	L	NA
dormitorio	20	12	6.87	Guasch	H	H	19.69
hurgar	20	17	4.25	Guasch	L	L	NA
pecar	20	23	2.04	Guasch	L	L	NA
jirafa	25	5	4.9	pre-testing	H	L	NA
rey	25	1	4.57	pre-testing	H	H	143.59
ancla	25	4	4.61	pre-testing	H	L	NA
pavo	25	3	4.51	pre-testing	H	L	NA
oso	25	2	4.58	pre-testing	H	H	9.01
pez	26	1	4.85	pre-testing	H	H	10.81
pulpa	26	3	3.15	pre-testing	L	L	NA
hada	26	2	5.59	Guasch	H	L	NA
incomodidad	26	4	2.59	Guasch	L	L	NA

Note: The numbers on the imageability value column corresponds to imaginability scores obtained in our pre-testing of items or derived from Guasch, Ferré & Fraga (2016). In this column, the cutoff point from our item pre-testing for high vs. low imaginable words was set at more than 4.5 on a 5-point Likert scale. Similarly, the cutoff point for high vs. low imageability from the Guasch et al., 2016 was set at more than 5.5. The imageability source column reflects whether the imageability value comes from our pre-testing or from Guasch et al., 2016. H = high. L = low. The RAE Lexical frequency column reflects the values of lexical frequency reported in the 'Corpus of Reference of Current Spanish' (CREA) developed by the Royal Spanish Academy (RAE). Words included in the 10,000 most frequent word were judged as highly frequent, whereas those not included were considered low-frequency words.

Appendix 4: Summary of results by task and country

Task	Total (N=82)	Chile (N=44)	Colombia (N=28)	Spain (N=10)	BF
Line bisection					
Mean (SD)	0.1 (0.9)	-0.1 (1.0)	0.6 (0.7)	-0.1 (0.4)	0.111
Median [Min, Max]	0.0 (-2.5, 1.5)	0.0 (-2.5, 1.5)	0.8 (-0.5, 1.5)	0.0 (-0.8, 0.5)	
Semantic memory					
Mean (SD)	9.5 (0.8)	9.3 (0.9)	9.7 (0.5)	9.9 (0.3)	0.299
Median [Min, Max]	10.0 (7.0, 10.0)	9.5 (7.0, 10.0)	10.0 (8.0, 10.0)	10.0 (9.0, 10.0)	
Word fluency - animal					
Mean (SD)	19.1 (6.1)	18.5 (5.2)	19.9 (5.5)	20.0 (10.2)	4.842
Median [Min, Max]	19.0 (4.0, 34.0)	18.0 (9.0, 34.0)	21.0 (11.0, 30.0)	21.5 (4.0, 33.0)	
Word fluency - /s/					
Mean (SD)	11.4 (4.8)	11.0 (3.8)	11.7 (4.4)	13.2 (8.4)	3.75
Median [Min, Max]	11.0 (2.0, 27.0)	11.0 (3.0, 20.0)	10.5 (5.0, 21.0)	12.0 (2.0, 27.0)	
Recognition memory					
Mean (SD)	9.7 (0.7)	9.7 (0.8)	9.7 (0.6)	9.7 (0.5)	7.872
Median [Min, Max]	10.0 (6.0, 10.0)	10.0 (6.0, 10.0)	10.0 (8.0, 10.0)	10.0 (9.0, 10.0)	
Gesture object use					
Mean (SD)	10.9 (0.9)	11.0 (0.7)	10.4 (1.1)	11.8 (0.4)	0.006
Median [Min, Max]	11.0 (8.0, 12.0)	11.0 (9.0, 12.0)	10.0 (8.0, 12.0)	12.0 (11.0, 12.0)	
Arithmetic					
Mean (SD)	4.9 (1.0)	4.7 (1.0)	5.3 (1.0)	5.0 (0.8)	0.637
Median [Min, Max]	5.0 (3.0, 6.0)	5.0 (3.0, 6.0)	6.0 (3.0, 6.0)	5.0 (3.0, 6.0)	
Comprehension spoken words					
Mean (SD)	27.6 (1.8)	27.3 (1.9)	27.8 (1.7)	28.8 (1.5)	0.71
Median [Min, Max]	28.0 (23.0, 30.0)	27.5 (23.0, 30.0)	28.0 (24.0, 30.0)	29.5 (26.0, 30.0)	
Comprehension spoken sentences					
Mean (SD)	29.0 (3.1)	27.8 (3.3)	30.8 (1.5)	29.7 (2.6)	0.003
Median [Min, Max]	30.0 (17.0, 32.0)	28.0 (17.0, 32.0)	31.5 (28.0, 32.0)	30.5 (24.0, 32.0)	
Comprehension spoken paragraphs					
Mean (SD)	3.6 (0.6)	3.7 (0.5)	3.6 (0.6)	3.6 (0.5)	7.332
Median [Min, Max]	4.0 (2.0, 4.0)	4.0 (2.0, 4.0)	4.0 (2.0, 4.0)	4.0 (3.0, 4.0)	
Comprehension written words					
Mean (SD)	29.2 (1.3)	29.0 (1.4)	29.3 (1.1)	29.4 (1.3)	5.229
Median [Min, Max]	30.0 (25.0, 30.0)	30.0 (25.0, 30.0)	30.0 (26.0, 30.0)	30.0 (26.0, 30.0)	
Comprehension written sentences					
Mean (SD)	28.8 (3.0)	27.9 (3.2)	29.8 (2.4)	30.0 (2.4)	0.172
Median [Min, Max]	30.0 (20.0, 32.0)	28.5 (20.0, 32.0)	30.0 (25.0, 32.0)	31.0 (24.0, 32.0)	
Repetition of words					
Mean (SD)	31.4 (1.3)	30.8 (1.7)	32.0 (0.2)	31.9 (0.3)	0.015
Median [Min, Max]	32.0 (26.0, 32.0)	32.0 (26.0, 32.0)	32.0 (31.0, 32.0)	32.0 (31.0, 32.0)	
Repetition of complex words					
Mean (SD)	5.9 (0.4)	5.9 (0.5)	6.0 (0.0)	6.0 (0.0)	0.371
Median [Min, Max]	6.0 (4.0, 6.0)	6.0 (4.0, 6.0)	6.0 (6.0, 6.0)	6.0 (6.0, 6.0)	
Repetition of nonwords					
Mean (SD)	8.9 (1.5)	8.8 (1.7)	9.1 (1.0)	9.1 (1.7)	4.971
Median [Min, Max]	10.0 (4.0, 10.0)	10.0 (4.0, 10.0)	10.0 (8.0, 10.0)	10.0 (6.0, 10.0)	

(*Continued*)

(Continued).

Task	Total (N=82)	Chile (N=44)	Colombia (N=28)	Spain (N=10)	BF
Repetition of digit strings					
Mean (SD)	12.1 (1.9)	11.7 (2.0)	12.9 (1.8)	11.6 (1.6)	0.523
Median [Min, Max]	12.0 (8.0, 14.0)	12.0 (8.0, 14.0)	14.0 (8.0, 14.0)	12.0 (10.0, 14.0)	
Repetition of sentences					
Mean (SD)	11.9 (0.7)	11.8 (0.9)	12.0 (0.0)	12.0 (0.0)	2.935
Median [Min, Max]	12.0 (8.0, 12.0)	12.0 (8.0, 12.0)	12.0 (12.0, 12.0)	12.0 (12.0, 12.0)	
Naming objects					
Mean (SD)	41.6 (3.4)	41.3 (3.5)	41.7 (3.3)	42.2 (2.8)	6.173
Median [Min, Max]	42.0 (32.0, 46.0)	42.0 (32.0, 46.0)	42.0 (34.0, 46.0)	42.0 (38.0, 46.0)	
Naming actions					
Mean (SD)	9.8 (0.8)	9.7 (1.0)	9.9 (0.4)	10.0 (0.0)	4.005
Median [Min, Max]	10.0 (4.0, 10.0)	10.0 (4.0, 10.0)	10.0 (8.0, 10.0)	10.0 (10.0, 10.0)	
Spoken picture description					
Mean (SD)	15.0 (1.7)	15.4 (1.2)	13.9 (2.1)	16.4 (1.1)	0.001
Median [Min, Max]	15.0 (11.0, 17.0)	15.0 (13.0, 17.0)	13.5 (11.0, 17.0)	17.0 (14.0, 17.0)	
Reading aloud words					
Mean (SD)	47.6 (0.9)	47.3 (1.1)	48.0 (0.0)	48.0 (0.0)	0.041
Median [Min, Max]	48.0 (44.0, 48.0)	48.0 (44.0, 48.0)	48.0 (48.0, 48.0)	48.0 (48.0, 48.0)	
Reading aloud complex words					
Mean (SD)	5.8 (0.7)	5.6 (0.9)	6.0 (0.0)	5.8 (0.6)	0.812
Median [Min, Max]	6.0 (2.0, 6.0)	6.0 (2.0, 6.0)	6.0 (6.0, 6.0)	6.0 (4.0, 6.0)	
Reading function words					
Mean (SD)	6.0 (0.3)	5.9 (0.4)	6.0 (0.0)	6.0 (0.0)	3.922
Median [Min, Max]	6.0 (4.0, 6.0)	6.0 (4.0, 6.0)	6.0 (6.0, 6.0)	6.0 (6.0, 6.0)	
Reading nonwords					
Mean (SD)	9.7 (0.9)	9.4 (1.1)	10.0 (0.0)	10.0 (0.0)	0.071
Median [Min, Max]	10.0 (6.0, 10.0)	10.0 (6.0, 10.0)	10.0 (10.0, 10.0)	10.0 (10.0, 10.0)	
Writing: Copying					
Mean (SD)	26.1 (1.9)	26.0 (2.0)	26.8 (1.1)	25.2 (2.8)	0.827
Median [Min, Max]	27.0 (16.0, 28.0)	27.0 (16.0, 27.0)	27.0 (21.0, 27.0)	27.0 (22.0, 28.0)	
Writing picture names					
Mean (SD)	20.3 (1.2)	20.3 (1.2)	20.3 (1.4)	20.7 (0.5)	5.544
Median [Min, Max]	21.0 (17.0, 21.0)	21.0 (17.0, 21.0)	21.0 (17.0, 21.0)	21.0 (20.0, 21.0)	
Writing to dictation					
Mean (SD)	27.0 (1.1)	26.8 (1.3)	27.3 (0.8)	27.1 (0.9)	2.756
Median [Min, Max]	27.0 (22.0, 28.0)	27.0 (22.0, 28.0)	27.0 (26.0, 28.0)	27.0 (26.0, 28.0)	
Written picture description					
Mean (SD)	12.0 (1.9)	11.6 (2.1)	12.2 (1.7)	13.0 (1.4)	0.76
Median [Min, Max]	13.0 (7.0, 14.0)	12.5 (7.0, 14.0)	12.0 (8.0, 14.0)	14.0 (11.0, 14.0)	

Note. BF represents the Bayes factor supporting no differences between countries (null model). Therefore, small values support the hypothesis of differences between countries.