



HAL
open science

Guidelines and recommendations for cross-linguistic aphasia assessment: a review of 10 years of comprehensive aphasia test adaptations

Silvia Martínez-Ferreiro, Seçkin Arslan, Valantis Fyndanis, David Howard, Jelena Kuvač Kraljević, Ana Matić Škorić, Amaia Munarriz-Ibarrola, Monica Norvik, Claudia Peñaloza, Marie Pourquié, et al.

► **To cite this version:**

Silvia Martínez-Ferreiro, Seçkin Arslan, Valantis Fyndanis, David Howard, Jelena Kuvač Kraljević, et al.. Guidelines and recommendations for cross-linguistic aphasia assessment: a review of 10 years of comprehensive aphasia test adaptations. *Aphasiology*, 2024, pp.1-25. 10.1080/02687038.2024.2343456 . hal-04588233

HAL Id: hal-04588233

<https://hal.science/hal-04588233v1>

Submitted on 17 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Guidelines and recommendations for cross-linguistic aphasia assessment: A review of 10 years of Comprehensive Aphasia Test adaptations

Silvia Martínez-Ferreiro¹, Seçkin Arslan^{2†}, Valantis Fyndanis^{3,4}, David Howard⁵, Jelena Kuvač Kraljević⁶, Ana Matić Škorić⁶, Amaia Munarriz-Ibarrola⁷, Monica Norvik^{4,8}, Claudia Peñaloza⁹, Marie Pourquié¹⁰, Hanne Gram Simonsen⁴, Kate Swinburn¹¹, Spyridoula Varlokosta¹², Eva Soroli¹³
§ On behalf of Working Group 2: “Aphasia Assessment and Outcomes” The Collaboration of Aphasia Trialists

¹ University of A Coruña, Spain

² CNRS & Université Côte d’Azur, Nice, France

³ Cyprus University of Technology, Cyprus

⁴ University of Oslo, Norway

⁵ Newcastle University, UK

⁶ University of Zagreb, Croatia

⁷ University of the Basque Country UPV/EHU, Spain

⁸ Statped, Norway

⁹ University of Barcelona, Spain

¹⁰ IKER-CNRS (UMR 5478), France

¹¹ University College London, UK

¹² National and Kapodistrian University of Athens, Greece

¹³ University of Lille & STL-CNRS (UMR 8163), France

Correspondence

†Dr. Seçkin Arslan, CNRS - Université Côte d’Azur Campus Saint Jean d’Angely - SJA3 / MSHS-SE 25 avenue François Mitterrand 06300 Nice CEDEX Tel.: + 33 04 89 15 22 71

Email: Seckin.ARSLAN@univ-cotedazur.fr

Disclosure statement:

The authors report there are no competing interests to declare.

Acknowledgements

We are grateful to the members of Working Group 2 *Aphasia Assessment and Outcomes* of the Collaboration of Aphasia Trialists (CATs) for their support and contributions towards aphasia test adaptations to several languages in the last 10 years. We also thank the Croatian Science Foundation for allowing the members of Working Group 2: *Aphasia Assessment and Outcomes* to use some of the pictures from the Croatian database created by the professional illustrator Marko Belić (grant no. HRZZ-2421-UIP-11-2013), as well as Eva Soroli who also created drawings for different language versions of the CAT.

Funding

This work has been supported by the Collaboration of Aphasia Trialists: COST IS 1208, Tavistock Trust for Aphasia (TTA, 2017, 2023). We also acknowledge the following funding support: Silvia Martínez-Ferreiro is supported by the Ramón y Cajal fellowship, Ministerio de Ciencia e Innovación, Spain (grant no. RYC2020-028927-1). Seçkin Arslan is supported by the European Union's Horizon 2020 research and innovation programme, Marie Skłodowska-Curie grant (agreement no. 838602). Valantis Fyndanis, Monica Norvik, and Hanne Gram Simonsen received support from the Research Council of Norway through its Centres of Excellence funding scheme (project number 223265). Jelena Kuvač Kraljević and Ana Matić Škorić received support from the Adult Language Processing project (HRZZ-2421-UIP-11-2013) of the Croatian Science Foundation (HRZZ). Amaia Munarriz-Ibarrola is supported by the Basque Government (IT627-22). Claudia Peñaloza is supported by grant RYC2021-034561-I funded by MCIN/AEI/10.13039/501100011033 and the European Union Next GenerationEU/PRTR. Marie Pourquoié receives support from the Region Nouvelle Aquitaine (ESR2019) and GEROA foundation. Eva Soroli is supported by the academic council of the University of Lille, the National Council of Universities, and the French Ministry of Higher Education, Research & Innovation CRCT and PEDR grants.

Abstract

Background. Standardised aphasia assessment tools may not always be available in a variety of languages, posing challenges for speech and language therapists to adequately assess and diagnose aphasia in speakers of those languages. In 2013, *Working Group 2 (WG2) Aphasia Assessment & Outcomes*, part of the Collaboration of Aphasia Trialists network, was formed with the purpose of developing reliable and valid aphasia assessment tools and their cross-linguistic adaptations. Over the past decade, WG2 has undertaken important adaptation projects, including the cross-linguistic adaptation of the *Comprehensive Aphasia Test (CAT)*; Swinburn et al., 2004).

Aims. This review aims to achieve three objectives: (a) describe the adaptation procedure of the CAT within WG2, (b) summarise common guidelines and recommendations for future adaptations, and (c) provide concrete solutions for specific cross-linguistic and cross-cultural challenges encountered during the adaptation and validation procedures of the CAT.

Methods. Between 2013 and 2023, WG2 employed a committee approach and fully adapted the CAT into Catalan, Croatian, Dutch, French, Hungarian, Norwegian, Spanish, and Turkish. Further adaptations are in progress for Arabic (Moroccan), Basque, Cantonese Chinese, German, Greek, Icelandic, Lithuanian, Serbian, Slovenian and Swedish. The review comprehensively addresses the linguistic/cultural adaptation and validation procedure for the three components of the battery: the Cognitive screening, the Language battery and the Aphasia Impact Questionnaire. Critical outcomes and some best practice recommendations from psychometric norming and piloting are also discussed.

Outcomes and results. This review builds upon prior work (Fyndanis et al., 2017) and serves as a practical guide for researchers and clinicians undertaking cross-linguistic adaptations of the CAT, with specific conclusions and recommendations drawn from WG2's adaptations in 19 languages with diverse typological properties. Building on the work exemplified in this paper, future initiatives can direct their efforts towards adapting the CAT for PWA from different linguistic backgrounds for whom validated assessment instruments may be unavailable. This can be achieved through rigorous systematic adaptation procedures for the establishment of comparable language versions of this tool, valuable for various clinical applications. Such endeavours have the potential to provide access to valuable shared datasets

for their use across international aphasia trials, and for comparable clinical work within the aphasiology community.

WORDCOUNT: 355/400

Keywords. Aphasia assessment, Comprehensive Aphasia Test, Psycholinguistic variables, Linguistic and Cultural adaptations, Validation challenges, Best practice recommendations

1. Introduction

Aphasia assessment is essential to determine the extent to which different language abilities are negatively impacted in people with aphasia (PWA). Although the tendency is progressively changing, there is a bias in the availability of standardised aphasia assessment tools towards English, while assessment tools in other languages are largely unavailable or have not undergone a thorough adaptation process (see also Ivanova & Hallowell, 2013; Spreen & Risser, 2003 for overviews). Ivanova and Hallowell (2013) highlight the lack of standardised and normed aphasia assessment tools in multiple languages independently of the number of speakers. The authors also show that many of the existing tests in languages other than English are direct translations from well-established English-language tests, such as the *Boston Diagnostic Aphasia Examination* (BDAE; Goodglass & Kaplan, 1972 and subsequent editions) or the *Western Aphasia Battery* (WAB and WAB-R; Kertesz, 1982, 2006), often without critically controlling for linguistic/typological, cultural and standard psychometric properties. Further, comparable aphasia assessment tools across languages are lacking, as the existing batteries and their operating procedures largely vary across countries (Brady et al., 2014; Fyndanis et al., 2017).

During the last decade, Working Group 2 (WG2) *Aphasia Assessment and Outcomes* has operated under the Collaboration of Aphasia Trialists, an international network of multidisciplinary aphasia researchers, to generate psychometrically accurate, reliable, validated, and cross-linguistically/cross-culturally comparable aphasia assessment tools. Such tools are important for research, to establish larger aphasia datasets comparable across languages, as well as for clinical purposes in languages without existing valid assessment instruments. One of these initiatives is the cross-linguistic adaptation of the *Comprehensive Aphasia Test* (CAT), which was originally developed for English by Swinburn, Porter, and Howard (2005). Between 2013 and 2023, WG2 adapted the CAT into Catalan, Croatian, Dutch,

French, Hungarian, Norwegian, Spanish and Turkish. Other adaptations are in progress, including Arabic (Moroccan), Basque, Cantonese Chinese, German, Greek, Icelandic, Lithuanian, Serbian, Slovenian, and Swedish¹. The cross-linguistic versions of the CAT described in this paper are subject to an exhaustive, committee-based adaptation and validation process. Psycholinguistic variables such as word frequency, imageability (Rofes et al., 2018), and to a lesser extent animacy, are controlled for alongside linguistic variables such as word length, orthographic regularity, sentence length and complexity, and transcultural factors (Fyndanis et al., 2017; Martínez-Ferreiro et al., 2024).

This review aims (i) to provide a detailed overview of the CAT adaptation process conducted under WG2, (ii) to distil and expound upon prevalent cross-linguistic standards in aphasia assessment tool adaptation and standardisation, and (iii) to provide concrete solutions to linguistic, cultural and psychometric challenges encountered in the process of adapting the CAT into different languages. Embracing a holistic approach, this review offers valuable insights, a blueprint for similar initiatives, and positions itself as a resourceful guide for researchers and clinicians engaging in CAT adaptations.

2. The Comprehensive Aphasia Test

No single assessment battery can cover all dimensions of aphasia, which makes combining assessments and protocols a common practice (WHO, 2001). As most operational definitions of aphasia revolve around the formula “acquired language disorder”, most assessment tools include more or less exhaustive language production and comprehension subtests. Other key

¹ The Danish adaptation of the CAT (Swinburn et al., 2014) has been independently completed by two SLTs and is not necessarily comparable to the adaptations reported here. This paper version (with manual) includes new illustrations for all items, however, the imageability data used were taken from English and the test was normed only with 20 healthy adults. There is also a version in Japanese by Watamori and colleagues close to completion. As in the case of the Danish CAT, this adaptation does not adhere **strictly** to the methodology described here.

components such as the effect of brain injury on cognitive processes, and its impact on daily life activities are more frequently neglected (Castro et al., 2023).

From its first edition (Swinburn et al., 2005), the CAT has distinguished itself by being a comprehensive assessment battery envisioned to characterise the performance of PWA across all language modalities, and identify associated cognitive deficits, and the psychological and social impact of aphasia in their everyday life. Hence, in line with the International Classification of Functioning, Disability, and Health (WHO, 2001), it adopts the broader view of aphasia in terms of “body function and structure, performance of activities, participation in relevant life situations, and the influence on functioning of environmental and personal factors” (Simmons-Mackie et al., 2005: 12). Moreover, the CAT can be used to evaluate changes in severity over time, and to plan interventions, avoiding syndromic classifications associated with localizationist approaches to aphasia (Howard et al., 2010).

The first edition of the CAT (Swinburn et al., 2005) comprises three parts and, in total, 34 subtests: (I) a cognitive screening with 6 subtests, (II) a language battery with 21 subtests, and (III) a Disability questionnaire with 7 subtests. For an overview, see Table 1.

Table 1. A summary of subtests and tasks in the original CAT

Part I: Cognitive screening	Part II: Language battery					Part III: Disability questionnaire
1 Line bisection	<i>Comprehension</i>	<i>Repetition</i>	<i>Naming</i>	<i>Reading</i>	<i>Writing</i>	28 Talking
2 Semantic memory	7 of spoken words	12 of words	17 objects	20 words	24 copying	29 Understanding
3 Word fluency	8 of written words	13 of complex words	18 actions	21 complex words	25 picture names	30 Reading
4 Recognition memory	9 of spoken sentences	14 of non-words	19 Spoken picture description	22 function words	26 dictation	31 Writing
5 Gesture object use	10 of written sentences	15 of digit strings		23 non-words	27 Written picture description	32 Intrusion
6 Arithmetics	11 of spoken paragraphs	16 of sentences				33 Self-image
						34.Emotional consequences

Part I includes a cognitive screening with six subtests that assess PWA’s cognitive abilities (which may influence their performance in the language battery; Howard et al., 2010),

including verbal memory, word fluency, and visual neglect. Part II contains a detailed language assessment battery consisting of 21 subtests that tap into auditory and written language comprehension (subtests 7 to 11), repetition (subtests 12 to 16), oral production – including naming – (subtests 17 to 19), reading (subtests 20 to 23), and writing (subtests 24 to 27). All subtests were designed to retrieve information about the linguistic behaviour of PWA at the word, sentence, and paragraph levels, controlling for potentially confounding variables such as imageability, frequency, word length, animacy, and syntactic complexity. Part III includes a concise psychosocial impact questionnaire. While the original version of the CAT (Swinburn et al., 2005) measured psychosocial impact via the *Disability Questionnaire* (DQ), its second edition (Swinburn et al., 2023) replaced it by the *Aphasia Impact Questionnaire* (AIQ-21; Swinburn et al., 2018). The inclusion of aphasia-related psychosocial aspects underscores the comprehensiveness of the CAT relative to other prominent aphasia batteries used in clinical and research contexts such as the BDAE (Goodglass & Kaplan, 1972), the WAB (Kertesz, 2006), and the *Porch Index of Communicative Ability* (PICA; Porch, 1967; 1971).

The CAT is standardised and normed. Although it includes a relatively large number of language and cognitive subtests, these contain a manageable number of items to ease its administration and keep time requirements to a minimum (90-120 minutes; Howard et al., 2010). This renders the CAT a valuable resource for speech and language therapists/practitioners and researchers.

3. Test adaptations

The adaptation of a language test for a population with language impairment is a challenging task. Adaptations should be accurate and sensitive to reflect scientific integrity and social equity in standard care. All languages, irrespective of surface similarities, are inherently different. According to Ivanova and Hallowell (2013: 8), “there is not a one-to-one match

between words and syntactic structures across any two languages (even similar languages with common origins, such as Romance languages)”. Biases may also result from the pictorial representations used in language tests. Thus, when adapting existing tools, direct translation alone is inappropriate (Paradis & Libben, 1987).

3.1 Adaptation standards

In recent decades, various professional committees and scientific associations including the International Test Commission (ITC, 2017) and the European Federation of Psychologists’ Associations (EFPA, Evers et al., 2013) have elaborated guidelines for test adaptation, which range from lists of best practices to well-developed evidence-based recommendations. Since the validity of the results collected in the test phase depends on the adaptation process, it is necessary to define the test adaptation procedure as a list of standards (Van de Vijver, 2016).

Before starting to adapt the test content, time requirements, effort, and personnel available should be ensured as suitable resources. All adaptations must then undergo three necessary steps.

(i) *Context*: It is important to address all contexts and factors that will influence test adaptation. The constructs to be adapted should be considered from a linguistic viewpoint, accounting for the linguistic peculiarities of each language, besides other non-linguistic factors (Hambleton, 1996). Considering the context helps to control for any confounding factors related to the background variables.

(ii) *Test development and adaptation*: The test must reflect the linguistic and cultural characteristics of its target population, whilst ensuring that the test items used in the original and the newly adapted test are comparable in difficulty, readability, grammar usage, writing style, and punctuation (Hambleton, 1996).

(iii) *Administration*: The choice of tasks, item formats, test instructions, and test administration procedures must be accessible to the target population (Van de Vijver & Hambleton, 2014).

If these steps are not seriously considered, the newly adapted test may differ from the original in difficulty and reliability (Mokkink et al., 2010). The entire adaptation process, including the methods used to assess equivalence and the rationale for changing items between the original and the adapted versions, should be fully justified and documented. An example of such documentation for the adaptation of the CAT into different languages - the focus of the present paper - is also provided by Fyndanis et al. (2017).

3.2 Standardisation principles

Aphasia assessment tools must be standardised and normed to return reliable scores. This requirement reflects two fundamental psychometric properties - validity and reliability - which are pivotal in the standardisation process. Validity is a multifaceted construct that refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests (Van de Vijver, 2016). Reliability means that the results are reproducible. If the test is reliable, changing the examiner or taking repeated measurements of the same participant within the same period of time should only result in small changes in measured scores (see e.g., Özdemir et al., 2022 and; Zakariás & Lukács, 2022; for examples from the Turkish and Hungarian CAT adaptations, respectively).

Psychometric properties including validity and reliability are multi-component constructs. Although it would be desirable that newly adapted assessment tools report on all types of validity and reliability, this is often not feasible. Moreover, only a small number of psychometric appraisal instruments are currently available and they cover a mix of standards

of quality (Rosenkoetter & Tate, 2017) making it difficult to determine which aspects of validity and reliability should be prioritised. The COSMIN Study Design checklist (Mokkink et al., 2010) could be considered as a comprehensive point of reference for the report of test psychometric properties (content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity and responsiveness). However, other tools available for this purpose (Rosenkoetter & Tate, 2017) may also allow for a combination of properties that minimally demonstrate that the newly adapted test is reliable, valid, standardised and free from bias. Construct validity means that a test measures the theoretical construct that it purports to measure. The CAT is based on the language processing model specified in Patterson and Shewell (1987)², a comprehensive model that includes all four language modalities - comprehension, speaking, writing and reading. The aim of the model is to understand the underlying nature of the language impairment and to define the strengths and weaknesses of the language functioning of a PWA in order to enable impairment-based therapy. Therefore, since the aim of the CAT was to describe the performance of PWA in different language domains and modalities and not to classify disorders, its structure reflected Patterson and Shewell's (1987) language processing model. Importantly, whether a construct has the same qualities across cultures (linguistic and cultural appropriateness) should be borne in mind.

Even when the adapted version respects the structure (subtests, items) and rationale of the original test (underlying hypotheses and variables), to keep content/construct validity constant across languages, reliability and validity still need to be measured for each new test. These psychometric features are especially important in clinical settings. When the test is used to assess a neurologically vulnerable population, such as PWA, it is critical that tools can

² Construct validity reflecting lexical processing refers only to the language part of the CAT.

reliably distinguish between individuals with and without language and communication difficulties by achieving both sensitivity (i.e., a test's ability to correctly identify patients with a given disorder) and specificity (i.e., a test's ability to correctly identify people without the disorder). The use of valid and reliable tests allows the clinician to minimise false positives (i.e., cases misdiagnosed as having a disorder) and false negatives (i.e., cases with a disorder not detected). Tests with good psychometric properties, a clearly defined phenomenology of the disorder, and clear administration procedures for examiners help reduce diagnostic errors.

In the validation of aphasia assessment tools, the inclusion of neurotypical individuals as a control group, is crucial to establish a reference for within-normal-limit performance for comparative analysis. In other words, neurotypical individuals serve as a reference population, allowing researchers to differentiate between language impairments associated with aphasia and variations that may arise naturally within the broader population. The CAT adaptations reported here systematically included neurotypical individuals in the whole adaptation procedure: first in order to pre-test the validity of the selected items for the adapted versions (e.g., within a name agreement task), and then to pilot the adapted versions of the test and further identify the boundaries between normal language function and deviations indicative of aphasia. To enhance the generalizability of findings, inclusion criteria required native participants of the target languages, with ages between 50 and 80 - mirroring the age range commonly affected by aphasia, with varied educational levels, and without any attested psychiatric/neurological disorders or visual/hearing impairments (other than corrected-to-normal vision/audition).

Statistical analyses are necessary to compute basic test measurements and to compare performance across languages (Matić Škorić et al., 2023; this issue). Studies aimed at examining cross-cultural validity, i.e. comparing original and adapted tests or adapted tests

between them, often show differences in participants' language behaviour. Although one of the most important requirements for testing cross-cultural validity is the uniformity of the sample in two languages (e.g., the level of education or the number of languages participants use in everyday life), this is sometimes difficult to achieve as such factors vary across cultures. For example, the Croatian (CAT-HR) and Norwegian (CAT-N) adaptations of the CAT have been recently compared (Matić Škorić et al., 2023; this issue) and it was shown that differences in performance of PWA on these adapted tests were caused by between-group sociodemographic differences and differences in item difficulty, although the same test development instructions were used for both versions. Furthermore, the French version of the CAT (Jacquemot et al., in preparation) was used to assess the performance of both first and second language French-speaking PWA and neurotypical participants (Python et al., 2023). Python et al.'s (2023) study reported that, in addition to the lower scores achieved by PWA in comparison to neurotypical individuals, near-native (but second language) neurotypical French speakers also demonstrated reduced scores compared to native French speakers in specific subtests (including sentence repetition, oral and written naming, nonword reading and word dictation subtests). This performance discrepancy is attributed to second language proficiency and everyday use and is important to take into account during the adaptation and validation procedure of a test. For example, one needs to be able to differentiate between language impairments associated with aphasia and variations that may arise naturally for other reasons (e.g., within the broader ageing population or within not-fully-native speakers of a language). Such findings suggest that linguistic and sociodemographic characteristics of the populations tested should be taken systematically into account in language assessment.

4. Cross-linguistic adaptations of the CAT: cultural/linguistic challenges and solutions

A consensus meeting in 2015 involving all WG2 teams working on the adaptation of the CAT in languages other than English (hereafter language teams) set the basic ground rules for the CAT's adaptation process. All language teams pledged to follow the same set of principles underlying the design of the original English version of the CAT (Swinburn et al., 2005). Changes tailored to be in tune with cultural and linguistic differences needed to adhere to a set of guidelines created by the 2015 consensus group.

All CAT adaptations must fully meet eight general requisites (Fyndanis et al., 2017) for the final outcome to be considered “an adaptation of the CAT”:

- (i) The adapted test maintains the structure of the original English CAT and includes the same number of items per subtest.
- (ii) Items and visual materials can be altered or replaced with culturally appropriate ones following strict adaptation guidelines.
- (iii) All relevant items, replaced or not, in a given language version must first be normed based on a ‘name agreement task’ to ensure that each entry elicits a consistent response (e.g., for the object and action naming subtests). This is especially important for target items across subtests. The cut-off is set at 90% (although it can be lowered to 85% in exceptional cases, e.g. in contexts of language contact in which the use of loanwords is habitual, and in languages with high dialectal variation). By consensus, a minimum of 20 neurotypical individuals (including a balanced number of men and women, in the 30 to 60 age range, native speakers of the target language, and with varied educational levels) must be tested with the name agreement task in order to select the most appropriate items to be included in the adapted version of the CAT.
- (iv) To ensure that items are equivalent to the original, an imageability study must be conducted in the absence of an available database in the language of the adaptation. A 7-point Likert scale is recommended to rate items in a minimum population of 20

neurotypical individuals, although certain language teams (e.g., Croatian and Spanish) have opted for a 5-point scale (Rofes et al., 2018). Each participant can rate a maximum of 100 items³.

- (v) Frequency data, preferably based on oral language production, should be provided. If no spoken corpora are available in a given language, film subtitle data or large written corpora could be used (see Coltheart, 1981, on how to combine different datasets). In the absence of those, a frequency or familiarity study must be completed. Items should be rated using a 7-point Likert scale. Again, at least 20 participants should be tested for each word, with no more than 100 words per participant⁴.
- (vi) When the first version of an adapted test is developed, it needs to be piloted on a small sample of healthy speakers (at least 10-20 individuals aged 30-60). For languages or dialects spoken in different geographical areas (e.g., French or Spanish), the pilot study should include representatives from more than one area to rule out potential dialectal and cultural differences.
- (vii) Following the piloting phase, each adapted version needs to undergo norming and psychometric validation in accordance with standard procedures.
- (viii) A focus group of people with chronic aphasia (minimum n=3) should be engaged to advise on the adaptation of the AIQ-21 (Swinburn et al., 2018), focusing on the cultural appropriacy of the items and images.

³ A gap between the high and low imageability ratings is necessary. For instructions on how to conduct an imageability study, see Paivio et al. (1968) and Toggia & Battig (1978). See Rofes et al. (2018) for further details and data on imageability ratings in published and ongoing CAT adaptations.

⁴ A gap between low and high frequency ratings is necessary. For reference, the range in the original English version was set at 2-15 occurrences per million for low-frequency items and 50-150 per million for high-frequency items in naming subtests, and above 100 per million for high frequency items in repetition subtests (Swinburn et al., 2005).

Once the validation process is completed, the adapted resource is ready for dissemination. For all adaptations, the names of the original authors come first, followed by the authors of the respective adaptations. Authors must determine whether there is need for a printed, an electronic version (e-CAT) or both. Across the 2019 and 2022 WG2 consensus meetings, details of the publication process of both printed and electronic versions were discussed with the owner of the CAT copyright, Psychology Press (a part of the Taylor and Francis Group) and the developers of the electronic version (PLORAS⁵), based at University College London. With the support of the Taylor and Francis Group, the decision was made that, for printed versions, interested teams should contact a local publisher. Local e-CATs, however, are developed with support from both PLORAS and the Taylor and Francis Group. A common goal across WG2 adaptation teams is to have an e-version of the CAT in all languages in the future.

The WG2 consensus meetings held in 2019 and 2022 brought about an important breakthrough. As pointed out by Peña-Casanova et al. (2019), tests are a reflection of the environment in which they are created. As such, the most widely used assessment tools show a bias towards western urban middle-class elements produced in the standard dominant register. The use of a formal register, which tends to overlap with the concept of official language, enters in conflict with the individual realities of PWA in many countries, especially in the case of minority and non-normalised languages (e.g., Basque). This issue tends to be overlooked in neuropsychological tests. WG2 aims at promoting a richer scenario favouring the inclusion of varieties/dialects with or without official recognition and varieties restricted to oral registers. Different scenarios emerge with the introduction of these new varieties. In what follows, we introduce some recommendations based on real examples.

⁵ Predicting Language Outcome and Recovery After Stroke. See <https://www.ucl.ac.uk/ploras/project-information-versions/project-info-standard>

The controversy between language and dialects is illustrated through the case of Basque, where Euskara Batua (the standardised variety developed in 1968) coexists with five dialects (Western, Central, and Navarrese in Spain, and Navarrese–Lapuradian and Zuberoan in France; Zuazo, 2019). The coexistence of an agreed-upon standard variety with different dialects may raise issues during the name agreement phase, similar to those observed in languages with wide geographical distributions (see Martínez-Ferreiro et al., 2024, for Spanish; Pourquoié & Munarriz-Ibarrola, this issue, for Basque). Whenever possible during the adaptation process, unequivocal items across dialects should be included (e.g., avoiding terms such as *aubergine* (UK)/*eggplant* (US); *serviette* (UK)/*napkin* (US)). When this is not possible, item selection should be based on inclusiveness/representativeness. In these cases, correct responses in both varieties should be credited. In the case of Basque, name agreement was a great challenge considering the heterogeneous knowledge and use of the standard variety (mostly in the Northern Basque Country and in speakers older than 50 who did not attend Basque schools). Considering the major difficulties of reaching 90% name agreement in subtests with pictures, different answers were considered as correct if they met the required psycholinguistic and linguistic properties in the subtest irrespective of being part of the standard dictionary. For example, for the word *intestine*, two phonological variants were accepted: the standard *heste* (NA 51,5%) and the non-standard form *hertze* (NA 27.3%), as both forms are inanimate, low frequency, short words.

The case of Cypriot Greek exemplifies dialects used mostly in oral registers. While Greek Cypriots predominantly use this oral variety of Greek in informal (and private) interactions, they commonly resort to Standard Modern Greek in formal settings, including written subtests (e.g., Fotiou & Grohmann, 2022; Tsiplakou, 2014). In this and other scenarios, and to preserve the integrity of the test, oral and auditory subtests can be adapted to the dominant spoken dialect (i.e., Cypriot Greek), whereas reading and writing subtests should be

adapted to the language variety commonly used in the respective language modalities (i.e., Standard Modern Greek). Moreover, since Cypriot Greek speakers often switch to Standard Greek in formal settings, correct responses in Standard Greek should be credited both in reading and writing subtests (which are in Standard Greek), and in oral and auditory subtests (which are in Cypriot Greek).

5. The Cognitive screening, the Language battery, and the AIQ-21

Unlike many assessment tools, the linguistic and psycholinguistic properties underlying the choice of subtests and items in the CAT are stated explicitly in the original manual, which greatly facilitates the adaptation procedure. The linguistic variables included in the CAT design are word length, phonological and semantic relatedness, orthographic regularity, morphological complexity, sentence length, and syntactic complexity. The controlled psycholinguistic variables in the CAT are frequency, imageability, and, to a lesser extent, animacy. These variables can impact performance in PWA on a variety of tasks (see e.g., Bastiaanse et al., 2009, 2016; Whitworth et al., 2013). Below we provide a step-by-step guide of the adaptation procedure with respect to these variables, the challenges experienced during the adaptation to particular languages, and the solutions proposed.

5.1. Part I – Cognitive screening

Given the nature of Part I, the cognitive screening does not entail great cross-linguistic adaptation difficulties. Subtest 1 (line bisection), subtest 3 (word fluency), subtest 5 (gesture object use) and subtest 6 (arithmetic) can generally be reproduced from the original English version. However, even in these subtests, if problems in any of the languages arise during the piloting phase, changes can be introduced. For instance, in the case of subtest 3, *Word fluency*, which aims at eliciting words starting with a given letter (letter “S”) and words belonging to a

specific category (*animals*), phonological variations such as Spanish *seseo* accent, which causes the assimilation of the phonemes /s/ and /θ/ (represented by the spellings ‘c’ before ‘e’ or ‘i’, ‘z’ and ‘s’) into /s/, may interfere with the completion of the subtest. Given the transparency of the Spanish orthographic system, this can be a potential confounding factor for people with low formal education⁶. If necessary, the letter can be substituted for a more appropriate, but high frequent one. In subtest 5, *Gesture object use*, although common objects are depicted, these can be substituted if they are not culturally relevant.

Contrary to previous subtests, subtest 2, *Semantic memory*, and subtest 4, *Recognition memory*, require adjustments to the different linguistic and cultural realities of most countries. In subtest 2, certain stimuli can raise identification problems (e.g., *nun* – a culturally marked term, or *eskimo* – whose frequency of use varies significantly across countries). Potentially problematic items should be replaced.

A final aspect to consider both within the cognitive and the language sections has to do with cascade effects, as certain substitutions may trigger changes in subsequent subtests and decisions must be maintained throughout the test adaptation process. This is for instance the case of subtest 4, *Recognition memory*, which must be adjusted to changes in subtest 2, *Semantic memory*, as all target items in the former must necessarily be present in the latter.

5.2. Part II – Language battery

A. Comprehension subtests

Receptive language in the CAT can be measured across five subtests of different complexity levels. Subtests 7 and 8, *Comprehension of spoken and written words*, have been reported by the original authors of the CAT as too easy, thus reducing their degree of informativeness.

⁶ For the solution implemented in Spanish, see Martínez-Ferreiro et al. (2024).

Variation in the number of distinctive features between the target word and the phonological distractor plays a strong role in item selection. It is recommended that complexity is increased by adding more items varying in only one distinctive feature, and by using less frequent words, while avoiding phonemic similarity in consecutive items and the use of compound words. The number and position of the distinctive features between the target and the phonological distractor (one, two or three in the original English CAT) do not lead to differences in the final score, that is, all items receive a maximum of two points independently of these factors. Of note, complexity across subtests 7 and 8 must remain balanced. Items must be equally demanding in both subtests to ensure comparability between auditory and written comprehension.

As an example, in the Norwegian and the Spanish adaptations (Jensen et al., 2024; Martínez-Ferreiro et al., 2024), words were first selected based on the availability of a phonological distractor with a distance from the target of one or two distinctive features, with changes restricted to consonants only. Adaptations such as the Slovenian version (Kocjančič Antolik & Širca Ule, 2017) have balanced the frequency of the phonological distractors with that of the target words. Overall, new words were also less frequent than their English counterparts. Unrelated distractors were chosen within the semantic family of phonological distractors mimicking the relationship between the target and the semantic distractor (e.g., for Norwegian: Target: *mur* ‘wall’; Phonological distractor: *bur* ‘cage’; Semantic distractor: *gjerde* ‘fence’; Unrelated distractor: *fengsel* ‘jail’). Although in the original test changes appear in initial, middle and final position, this is not always possible across languages. In Spanish, singular nouns ending in consonants are scarce; therefore, changes were restricted to initial and middle positions. This was also the case in Basque, as the citation form includes the determiner (e.g., *oihan-a* ‘forest-the’).

In subtests 9 and 10, *Comprehension of spoken and written sentences*, certain language-specific adjustments were required. In the original version, the subtests include simple sentences with one or two arguments (incorporating unergative and transitive verbs such as *walk* or *paint*), locatives (e.g., *The pen is under the paper*), semantically reversible sentences (actives and passives; e.g., *The policeman paints the dancer* vs. *The dancer is painted by the policeman*), and embedded clauses (e.g., *The carpet the cat is on is green*). To keep lexical, semantic, and syntactic complexity consistent across adaptations, the following measures were implemented. In cases of metonymy observed during the pilot phase, the problematic items were replaced. For example, in French, the ‘pen’ in the target sentence ‘*The pen is under the paper*’ was sometimes referred to as the ‘bic’. To avoid any semantic ambiguity, the item was replaced by a variant that involved a pencil instead. In languages that have syncretism between masculine and feminine pronouns, plural pronouns were used (e.g., in Basque, *Hark barre egiten du* '(S)he is laughing' vs. *Haiek barre egiten dute* 'They are laughing'). Similarly, in languages that may not have productive passive constructions, these were replaced by another non-canonical structure, e.g., object-extracted cleft sentences, pseudoclefts, and sentences with object-extracted relative clauses. For example, the original English sentence stimuli ‘*The policeman is painted by the dancer*’ and ‘*The butcher is chased by the nurse*’ have been adapted with modifications. In the Croatian version, passive constructions were replaced by active sentences with non-canonical word order (OVS; e.g. *Policaјca crta plesačica*. ‘The policeman-3sg-OBJ-ACC paints the dancer-3sg-SUB-NOM’; Swinburn et al., 2020). In Slovenian, a relative pronoun had to be added (e.g. *Preproga, na kateri je mačka, je zelena*. lit. ‘Carpet, on which is cat, is green’ ‘The carpet the cat is on is green’; Kocjančič Antolik & Širca Ule, 2017). Given the lack of passive structure in Basque, a similar structure was targeted in this version, active non-canonical OSV structures (*Sukaldaria sendagileak jarraitzen du* 'cook-D-ABS doctor-D-ERG follows AUX' 'The cook is chased by the doctor'; Pourquoié & Munarriz-

Ibarrola, this volume). In Turkish, structures with clefts were used to test complexity (e.g., *Aşçıyı kovalayan adam bir doktor* lit. ‘Cook-ACC chase-PART man-NOM one-INDEF doctor-BE’ ‘It is a doctor who chases the cook’) (Özdemir et al., 2022). In Catalan, in addition to passive constructions, clitic left dislocations were used (e.g., *El bomber és fotografiat pel pallasso* ‘The firefighter is being photographed by the clown’, and *Al bomber, el fotografia el pallasso* ‘To the firefighter, the clown is taking a picture of ’; Salmons et al., 2021). In the forthcoming Modern Greek version, passives were replaced by sentences with object-extracted relative clauses (e.g., *Dikse mu ti nosokoma pu sproxni i ciria*. ‘Show me the nurse that the lady is pushing’).

These modifications are based both on typological differences and on research evidence. For Catalan, impaired performance on clitic left dislocation, a frequent construction in this language, has been reported in PWA (Salmons & Gavarró, 2022). Structures with derived orders are difficult for PWA speaking Basque (OSV/VOS; Arantzeta et al., 2017) and Turkish (OVS/OSV; Mavis et al., 2020). Moreover, although comprehension of both reversible passive sentences and reversible object-extracted relative clauses can be impaired in Greek-speaking PWA (e.g., Fyndanis et al., 2013; Nerantzini et al., 2014; Varlokosta et al., 2014; but see Terzi & Nanousi, 2018), passives with the agent explicitly stated are infrequent in Greek (see Warburton, 1975, and references therein). This is also the case for Croatian (Ham, 1989) and Slovenian (Agović, 2019). In the latter, the passive voice is built either with the participle ending in *-n/-t* or with *se*. As the participle can express both passive voice and state, the statement can be ambiguous.

To promote gender balance and avoid cues based on the gender of subject and object noun phrases, in the CAT adaptations both participants had the same gender. Consequently, *nurse* was systematically replaced by *doctor*, *policeman* by *witch*, and *the dancer* was female.

To control for emotional valence of the sentence stimulus, phrases such as *butcher, kill, shoot,* and *hit* were replaced by more neutral phrases including *cook, kiss, carry,* and *push*.

In subtest 11, *Comprehension of spoken paragraphs*, reported by the original authors as an easy subtest, speech tempo becomes an important factor to evaluate performance. In the electronic versions, this is controlled for by default. Stimuli in this subtest were adapted as close as possible to the original content, whereas the names of the characters, contextual details and measurement units were replaced by culturally suitable ones. In Turkish, *Sally and Richard* were replaced by *Selma and Remzi*. *London, miles,* and *pounds* were substituted by *Ankara, kilometres,* and *lira* (Maviş et al., 2022). Currency may be an issue for languages spoken in different countries (e.g., Spanish speakers may use euros, pesos, or even dollars depending on region of residence). For the sake of consistency, in these cases, it is recommended to replace specific quantities with expressions like “extensive material damage” (Martínez-Ferreiro et al., 2024).

B. *Production subtests*

The expressive language section is the longest of the battery, with 16 subtests including repetition, naming, and sentence/discourse production subtests, as well as reading, and writing subtests. Except for subtest 15, *Repetition of digit strings*, all subtests in this section require a thorough adaptation process.

Repetition subtests are defined by the original authors as highly informative in clinical terms. In subtest 12, *Repetition of words*, the 16 items included are balanced for frequency and imageability as they are equally divided into high- vs. low-frequency words, and high- vs. low-imageability words. This balance must be kept in all adaptations. However, since all items need to be depicted and aim at a name agreement score of 90% in order to be included, low imageability should be interpreted as ‘lower’ imageability (e.g., the imageability of *weasel* is

lower than that of *dog*, although still significantly higher relative to terms such as *consciousness*). Additionally, although phoneme frequency is immaterial in this subtest, the number of syllables needs to be controlled for. The original version includes half monosyllabic and half trisyllabic words. However, in certain languages such as Greek, Croatian, and Slovenian, monosyllabic words are scarce (e.g., Kuvač Kraljević et al., 2020). Disyllabic and tetrasyllabic words can be used instead, since the key property of the subtest to be kept is variability in the number of syllables (e.g., *plant* in Croatian is a highly imageable, high-frequency word as in English, but disyllabic ‘*biljka*’). Compound words and homophones should be avoided throughout the subtest.

Subtest 13, *Repetition of complex words*, includes words with prefixes and suffixes, making this subtest highly susceptible to variation across languages. Although the number and position of morphemes need to be considered in the adaptation, the complex words of this subtest are only evaluated for correctness (not for frequency or morphological complexity). In languages with rich inflectional morphology and with most words being inflected (e.g., Basque, Greek), both inflectional and derivational morphemes should be present, as only derived words are generally considered to be morphologically complex. Subtest 21, *Reading complex words*, follows the same rationale, and should be equivalent to subtest 13 in terms of balance between derivational and inflectional morphemes.

The original subtest 14, *Repetition of non-words*, includes monosyllabic and disyllabic non-words with 3 to 7 phonemes. Adaptations such as the Dutch (Visch-Brink et al., 2014) and the Norwegian (Swinburn et al., 2021) versions include items highly similar to the English non-words (e.g., the English non-words *gart*, *clup*, *spenk*, *trimpy* and *prastode* became *gamt*, *klupp*, *spenk*, *brimti* and *praston*, respectively, in the Norwegian version). In cases where keeping non-words close to the original is not possible, the number of phonemes (3-7) needs to be controlled for. Non-words should not resemble existing words. Subtest 23, *Reading non-words*,

should follow the same pattern as subtest 14, *Repetition of non-words*, clearly differentiating non-words from real words.

In subtest 16, *Repetition of sentences*, the critical underlying principle is sentence length. Ideally, the number of content words per item, and, if feasible, also that of function words, should be maintained identical to the original English version. For instance, in Croatian (and Slovenian), simple past tense requires the periphrastic *be* + verb form (e.g. *Čovjek je otišao* ‘(The) Man left’, Kuvač Kraljević et al., 2020), adding an extra function word to the stimulus material that needs to be compensated for. In this case, this occurs naturally due to the absence of an obligatory determiner. Notwithstanding potential unavoidable alterations in the number of function words, the complexity of the syntactic structure, originally constructed with one coordinated and one complex sentence per block, needs to be equivalent across adaptations.

Acknowledging the importance of adding verbs to elicitation tasks (Mätzig et al., 2009; Rofes & Mahon, 2021), the CAT includes two naming subtests, one for objects, subtest 17, and one for verbs, subtest 18. In subtest 17, *Naming objects*, phonological complexity is not controlled for; however, number of syllables and frequency are important factors in the selection of the 24 items. For low-frequency nouns, animacy is also taken into consideration. Half of the low-frequency items should represent animate objects and the other half inanimate objects. Regarding the number of syllables, the decision taken by each language team should be the same as in subtest 12. Simple words that can be shortened during production (e.g. *bicycle* – *bike*; *television* - *tv*) and compounds should be avoided, as these may alter the difference between long and short words in this subtest.

Subtest 18, *Naming actions*, includes five unambiguous verbs. The elicited verb form in different languages needs consideration. While in languages like English the gerund is used in response to “What is the person doing?” (e.g. *eating*), in other languages such as Catalan or Greek, verbs are generally cited in infinitive (e.g. Catalan: *menjar* ‘to eat’), or in 3rd person

singular present tense (e.g. Catalan: *menja*; Greek: *troi* ‘S/he eats/is eating’). The selection of verbs can vary across languages, although forms should preferably be disyllabic. As an example, the Norwegian adaptation maintained the verbs ‘eating’ (*spiser*) and ‘sawing’ (*sager*) and replaced the rest with ‘throwing’ (*kaster*), ‘bathing’ (*bader*), ‘painting’ (*maler*) and ‘diving’ (*dykker*), all in the present tense (Swinburn et al., 2021).

Subtest 19, *Spoken picture description*, and subtest 27, *Written picture description*, have experienced major changes with respect to the original English version. Many adaptations have adopted the principles of the Dutch scoring system with slight variations (Visch-Brink et al., 2014). To simplify the scoring and improve its reliability, the new system includes a short form with ratings on content, form, and (in some cases) a general evaluation over a 3 to 4-point Likert scale.

In the Norwegian, French and Spanish versions, the scoring system builds on the Dutch one, but is developed further. It contains three parameters for form: tempo/fluency (only relevant for the spoken description), grammatical complexity, and grammatical correctness, scored on a scale from 3 (good) via 2 (medium) and 1 (weak) to 0 (missing). Based on discussions with the Dutch language team, four main concepts⁷ were identified: man sleeps; girl points/awakens/alerts; cat chases fish; books fall (E. Visch-Brink, personal communication, 22.10.2019). These are scored on a scale from 2 (complete and precise) via 1 (present, but not complete and/or precise) to 0 (missing). An example of the Norwegian scoring sheet for the spoken description is provided in Figure 1. By necessity, the scoring of picture descriptions is more subjective than the scoring of other tests in the Language battery. However, in the Norwegian norming study, inter-rater and intra-rater reliability of the scoring for these subtests were acceptable, although lower for spoken than for written description

⁷ More about the main concepts produced by PWA and HS from the Spoken picture description can be found in Kuvač Kraljević et al. (2023).

(Jensen et al., 2024). Potential differences in the scoring system between language versions need to be taken into account when comparing these two subtests. Even if some of the early adaptations still include the original scoring system (e.g., the Croatian adaptation; Swinburn et al., 2020), all new test adaptations should incorporate the simplified scoring system. To increase the reliability even more, it is recommended to include examples of the different levels of grammatical correctness and complexity (from 3 (good) via 2 (medium) and 1 (weak)) in the manual of each language version.

Figure 1: *Scoring sheet for spoken picture description in several adapted versions*

Content parameters	Score	Form parameters	Score
<i>Man sleeping</i>	0 1 2	Tempo/fluency	0 1 2 3
<i>Girl pointing/alerting/waking</i>	0 1 2	Grammatical correctness	0 1 2 3
<i>Cat trying to catch fish</i>	0 1 2	Grammatical complexity	0 1 2 3
<i>Books falling</i>	0 1 2		
Sum	/8	Sum	/9
Sum (Content parameters + form parameters)			/17

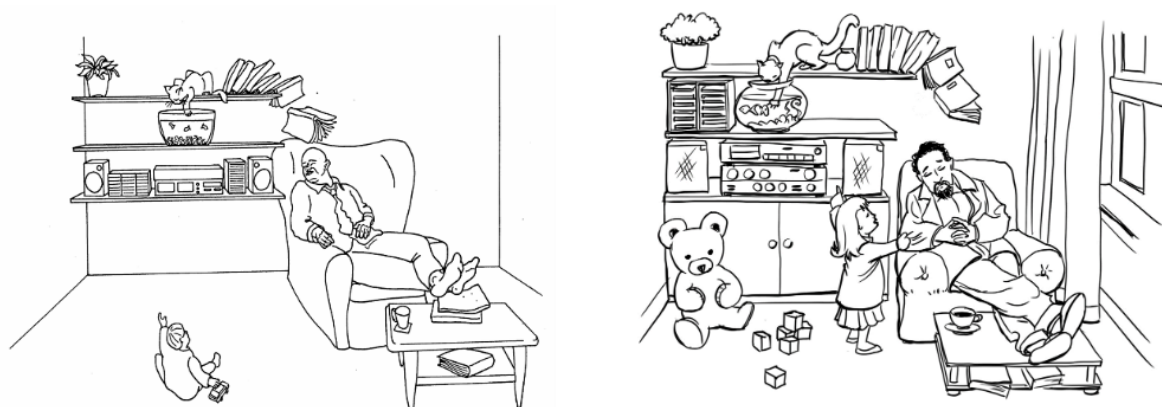


Figure 2. Picture description examples. Visual material used in the oral and written picture description subtests (19 and 27, respectively) of the original English version to the left

(Swinburn et al., 2005), and of several of the adapted versions to the right (e.g., Swinburn et al., 2020; 2021). The first image was reproduced with the permission of INFORMA UK LTD and the second with the permission of Naklada Slap.

It is worth noting that, although respecting the main elements, in 2015 the decision was made to update and, in some cases, culturally adapt the existing picture to adjust to the reality of the countries where the adaptations will be applied (e.g., the physical appearance of the man varies across adaptations; Figure 2). The same changes should be maintained in subtest 27, *Written picture description*, which makes use of the same picture, thus providing the perfect scenario to contrast oral vs. written skills.

Finally, the Language battery of the CAT further includes a set of subtests envisioned to assess reading and writing skills and account for deficits in the respective language modalities. Reading subtests include regular and irregular, short, long, and complex, content and function words, as well as non-words, whereas written subtests include letters, words and a short paragraph.

Subtest 20, *Reading words*, takes into account word frequency, imageability and number of syllables, which should be respected in all adaptations. The procedure to adapt this subtest is the same as in subtests 12, *Repetition of words*, and 17, *Naming objects*. Spelling-to-sound and sound-to-spelling regularity is another important factor here. Spelling-to-sound regularity (a.k.a. orthographic regularity) relates to the predictability of the pronunciation of the word on the basis of the written form, whereas sound-to-spelling regularity refers “to the relative ease with which the orthography of a word can be predicted on the basis of its phonetic realisation” (Fyndanis et al., 2017: 703). Both parameters vary across languages. In languages with opaque orthography, such as English and French, regularity can be easily manipulated, whereas in languages with transparent orthography, such as Basque, Spanish and Turkish, it can become

harder. Some language teams (e.g., the Basque team) have decided to drop regularity as a factor. It is important to note that some languages (e.g., Greek) are spelling-to-sound regular, but sound-to-spelling irregular, as they include sound units that have more than one possible orthographic realisation. In Greek, some sound units may correspond to one grapheme or to the combination of two graphemes. For example, in the Greek noun *ήχοι* [íçi] ‘sounds’, the vowel [i] occurs twice. In its first occurrence, it corresponds to the grapheme <η>, whereas in its second occurrence it corresponds to the combination of the graphemes <ο> and <ι> (<οι>). The spelling of words including such sound units cannot be predicted on the basis of their phonetic realisation. For languages like Greek, therefore, only the sound-to-spelling regularity variable could be adequately manipulated by including both sound-to-spelling regular words and sound-to-spelling irregular words.

In highly regular languages, transparency, that is, the systematicity with which a given letter maps onto its corresponding phoneme, can be manipulated. This was the solution adopted by the Hungarian language team (e.g., in Hungarian, transparent words such as *mód* /mɔ:d/ ‘mode’ were contrasted to words such as *jacht* /jaht/ ‘yacht’) (Zakariás & Lukács, 2022). Other teams opted for manipulating syllabic structure, using less frequent syllable patterns (e.g. complex consonant clusters in Spanish).

Finally, in subtest 22, *Reading function words*, subtest 24, *Writing: copying*, subtest 25, *Writing picture names*, and subtest 26, *Writing to dictation*, item selection should resemble the original English words as much as possible. In the original version, subtest 24 includes an apparently random selection of frequent letters and words, with one of the words being longer than the rest of them. Thus, the letters and words chosen for the adapted versions of CAT must be frequent in the different languages the test is adapted to, and one of the words should be longer than the others. When possible, compound words should be avoided. In subtest 25, the number of letters per word has to be maintained. However, in some languages it is not possible

to find content words that adjust to this constraint while adhering to the frequency and imageability requirements. In these cases, the number of letters in each word can be altered, but the total number of letters in the subtest should remain 21. Also, when feasible, one of the five items should be a homophone (e.g. *pear* in the original English version). The same holds for subtest 26. In this subtest, the fourth word should maintain its morphological complexity. For the non-word, the English CAT includes an unequivocally spelled item. To create the non-word in other languages, the preferred option is to opt for a vowel change in an otherwise existing word, thus respecting the phonotactic constraints of the target language.

During the adaptation phase, the underlying principles of the CAT should guide item selection (words/sentences, pictures) in all cases, and picture choices should be the last step. Other measures that apply across subtests are the following:

- (i) In the preliminary stages, additional items should be added to all subtests to compensate for item loss during the piloting phase. In subtests such as 12 and 17, this may imply doubling the number of proposed items to be able to fulfil the aimed distribution of items across imageability, frequency, and length parameters after eliminating those that do not reach the 90% name agreement threshold.
- (ii) Items used as distractors in one subtest can be used as targets in another section of the test.
- (iii) In languages with case marking, nouns should appear in the nominative case, assuming this is the prototypical citation form of the language.
- (iv) The development of new items may entail the design of new pictorial materials. The new pictures should remain black and white line drawings as in the original CAT. A database of existing pictures is available upon request.

- (v) The format of the stimulus and scoring books should be as consistent as possible in size and quality within the test and with the original CAT, as inconsistencies might result in non-comparable responses. Exact size of the visual materials is difficult to decide as publishers might use different proportions and may propose specific fonts. Our recommendation is to use a border line around visuals for attentional purposes.

5.3. The Aphasia Impact Questionnaire

There are differences between Disability Questionnaires (DQs) in early adaptations (e.g., the DQ only includes language-related questions in the Dutch adaptation; Visch-Brink et al., 2014). The DQ has since been modified and developed into the Aphasia Impact Questionnaire (AIQ) (Swinburn et al., 2018), extensively used in the UK, thus leading to the abandonment of the DQ.

The AIQ is a patient-reported outcome measure in its own right co-produced with the assistance of PWA. It covers three domains: communication, participation in everyday life activities, and well-being/emotional consequences of living with aphasia. The AIQ aims to balance the information related to cognitive and linguistic impairments, gathered from a professional perspective, with the perspective of the person who lives with the impact of these impairments on their everyday life.

Although some early adaptations still include the DQ (e.g., the Croatian adaptation; Swinburn et al., 2020), all new test adaptations should incorporate this update. For the Norwegian adaptation of the AIQ, a focus group interview of five PWA was carried out. This resulted in the addition of one question about digital communication, raising the final number of questions in the Norwegian version of the AIQ to 22 (as compared to 21 in the original AIQ; Swinburn et al., 2018). Although the AIQ is a part of the CAT, it is not to be seen as a language test, but rather as a structured starting point for PWA to express their subjective experiences of

living with aphasia; thus, comparability of different language versions of the AIQ should not be a linguistic issue, but may yield interesting cultural and/or sociological comparisons of responses and adaptations to language disability between languages.

5.4 Accompanying documents

The accompanying documents of the CAT including the user manual, the rating scale booklet and a scoring booklet also require attention. To be clinically useful, the language teams need to ensure that all these documents (including the administration procedures within the user manual) are clear and culturally appropriate.

6. Norms and validation

Once the materials have been adapted and the pilot phase has been concluded, it is time to evaluate the psychometric properties of each adapted version. There is no overall consensus on the exact number of people with and without acquired neurological damage to be included in the validation phase, although Ivanova and Hallowell (2013) recommend including 100 individuals in normative groups.⁸ The number may depend on the number of language users in the country or the language community. The original CAT (Swinburn et al., 2005) was based on the performance of 27 people without aphasia (aged 25-79), 56 PWA tested 1, 3, 6 to 12 months post onset, and 57 PWA tested on a single occasion, whereas more than 250 PWA and 500 controls participated in the Dutch validation process (Visch-Brink et al., 2014). The Norwegian adaptation was normed on 85 PWA and 84 healthy controls (Jensen et al., 2024), the Turkish version on 90 PWA and 200 controls (Özdemir et al., 2022), the Hungarian version

⁸ Please note that in reliability studies normative sample sizes are often recommended to be larger, such as 400 individuals (Karakaya et al, 2022). Here we follow recommendations for adapting tests; see Ivanova & Hallowell (2013).

on 100 PWA and 34 controls (Zakariás & Lukács, 2022), and the Croatian version on 114 PWA and 123 controls (Kuvač Kraljević et al., 2020).

Other important measures include internal consistency and test-retest reliability. The results of 137 PWA were examined for internal consistency in the original English version (Swinburn et al., 2004). In Croatian (n=114 PWA), the internal consistency of the subtests and modalities ranges from moderate to high and the validity measures on the modality mean can discriminate 85% of PWA from neurologically healthy individuals, i.e. those without aphasia. As expected, the reliability of modality scores is generally higher than those of the separate subtests and reliability is lower for subtests that include a smaller number of items. The Turkish and the Hungarian adaptations yielded similar results (Cronbach alpha coefficients of 0.88-0.89 and 0.7-0.97, respectively; Özdemir et al., 2022; Zakariás & Lukács, 2022). Additionally, the concurrent validity of Croatian CAT was determined using two tests: the *Peabody Picture Vocabulary Test* (PPVT-III-HR; Dunn, Dunn, Kovačević et al., 2010) and the *Test for Reception of Grammar* (TROG-2:HR; Bishop et al., 2014), which showed a moderate correlation between PPVT-III-HR/TROG-2:HR and all subtests from CAT that focus primarily on comprehension and production at the word and sentence level (Swinburn et al., 2020; Kuvač Kraljević et al., 2020). Again, the Hungarian adaptation produced similar outcomes when compared to the Western Aphasia Battery (WAB; Osmánné Sági, 1991) and the Test for the Reception of Grammar (TROG-H; Lukács et al., 2012) (Zakariás & Lukács, 2022).

The test-retest reliability was performed in such a way that the same group of raters performed the evaluation for the second administration of the test to the same group of PWA. In the original English version of the CAT (Swinburn et al., 2005), test-retest reliability was evaluated on a sample of 21 participants with chronic aphasia that undertook parts I and II of the test (Cognitive screening and Language battery) at different times post onset (1, 3, 6, and

12 months). In the Dutch adaptation (Visch-Brink et al., 2014), the performance of 19 chronic participants was evaluated. In 2015, consensus amongst WG2 members was established to target 10 chronic patients (at least 1 year post onset) to assess the test-retest reliability of each adaptation. The test-retest reliability of the Croatian version was examined by retesting 22 PWA six months after the first test was administered (Swinburn et al., 2020). Inter-rater variability was also examined. The original version showed high inter-rater agreement in almost all subtests, with a correlation coefficient around 0.9. This was also the case in the Turkish adaptation (Özdemir et al., 2022).

Regarding the AIQ, 31 PWA were interviewed for the original English version (Swinburn et al., 2018). The Norwegian version was tested on 21 PWA, not to establish norms but to see how the AIQ functions in actual use (Jensen et al., 2024).

7. Conclusions and suggestions for future research

This work describes the adaptation procedures of the CAT (Swinburn et al., 2005) undertaken by several language teams over the last decade and underscores that adapting aphasia assessment tools across languages can be a challenging endeavour. However, it also demonstrates that this goal can be effectively achieved via expert consensus and in consideration of several stimuli-related properties and linguistic and cultural factors that may critically impact the psychometric properties of the original tool and the resulting measurement of language behaviour in PWA. Although not conceived for the study of bi/multilingual speakers with aphasia, the adaptation of the CAT to multiple languages following strict guidelines can serve as a starting point to develop CAT-based versions for bi/multilingual populations in the future. More importantly, it also provides clinicians with a tool that can be used in multilingual contexts in the absence of more specialised tests. This is particularly relevant given the existing disparity in content across aphasia batteries (Castro et al., 2023).

For example, in the Basque Country, where people are either Basque-French bilingual in the North or Basque-Spanish bilingual in the South or may also be Basque-French-Spanish multilingual, the CAT adaptations in Basque, French and Spanish represent a particularly valuable tool to assess and compare aphasia deficits in these languages⁹. Also, sophisticated statistical approaches have been more recently developed to better assess the psychometric properties of newly adapted tests. For instance, structural equation modelling for identifying factorial equivalence across different language versions of an assessment tool (ITC, 2017) could be considered in future research. Important steps towards establishing the equivalence between CAT versions have been recently taken (Matić Škorić et al., 2023). Building on the amount of work that this paper exemplifies, we can look to a brighter future in which initiatives like the one described in this article allow for the combination of comparable cross-linguistic datasets, granting access to shared big data amongst the aphasiology community.

References

- Agović, H. (2019). Razlikovanje med stanjem in trpnikom na primerih iz Korpusa akademske slovenščine. *Jezik in Slovnstvo*, 64(3/4), 117-129. <http://www.dlib.si/details/URN:NBN:SI:doc-YKDJGFU9>
- Arantzeta, M., Bastiaanse, R., Burchert, F., Wieling, M., Martinez-Zabaleta, M., & Laka, I. (2017). Eye-tracking the effect of word order in sentence comprehension in aphasia: evidence from Basque, a free word order ergative language. *Language, Cognition and Neuroscience*, 32(10), 1320-1343. <https://doi.org/10.1080/23273798.2017.1344715>

⁹ See also Python et al. (2023) for a comparison between first versus second language users' performance on the French version of the CAT and the importance to consider second language issues during the standardisation process to better distinguish between stroke-induced language deficits and second language knowledge limitations, and thus to avoid misdiagnosis.

- Bastiaanse, R., Bouma, G., & Post, W. (2009). Linguistic complexity and frequency in agrammatic speech production. *Brain and language*, *109*(1), 18–28.
<https://doi.org/10.1016/j.bandl.2008.12.004>
- Bastiaanse, R., Wieling, M., & Wolthuis, N. (2016). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology*, *30*(11), 1221–1239.
<https://doi.org/10.1080/02687038.2015.1100709>
- Bishop, D.V.M., Kuvač Kraljević, J., Hržica, G., Kovačević, M., & Kologranić Belić, L. (2014). *Test razumijevanja gramatike – TROG-2:HR* [Test for reception of grammar, 2nd edition]. Naklada Slap.
- Brady, M. C., Ali, M., Fyndanis, C., Kambanaros, M., Grohmann, K. K., Laska, A. C., Hernández-Sacristán, C., & Varlokosta, S. (2014). Time for a step change? Improving the efficiency, relevance, reliability, validity and transparency of aphasia rehabilitation research through core outcome measures, a common data set and improved reporting criteria. *Aphasiology*, *28*(11), 1385–1392.
<https://doi.org/10.1080/02687038.2014.930261>
- Castro, N., Hula, W. D., & Ashaie, S. A. (2023). Defining aphasia: Content analysis of six aphasia diagnostic batteries. *Cortex*, *166*, 19–32.
<https://doi.org/10.1016/j.cortex.2023.05.005>
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505.
<https://doi.org/10.1080/14640748108400805>
- Dunn, L. M., Dunn, L.M., Kovačević, M., Padovan, N., Hržica, G., Kuvač Kraljević, J., Mustapić, M., Dobravac, G., & Palmović, M. (2009). *Peabody slikovni test rječnika, PPVT-III-HR* [Peabody Picture Vocabulary Test, 3rd edition]. Naklada Slap.

- Evers, A., Muñiz, J., Hagemester, C., Høtmælingen, A., Lindley, P., Sjöberg, A., Bartram, D. (2013). Assessing the quality of tests: revision of the EFPA review model. *Psicothema*, 25(3), 283-91. <https://doi.org/10.7334/psicothema2013.97>
- Fotiou, C., & Grohmann, K. K. (2022). A small island with big differences? Folk perceptions in the context of dialect levelling and koineization. *Frontiers in Communication*, 6, 770088. <https://www.frontiersin.org/articles/10.3389/fcomm.2021.770088/full>
- Fyndanis, V., Varlokosta, S., & Tsapkini, K. (2013). (Morpho)syntactic comprehension in agrammatic aphasia: Evidence from Greek. *Aphasiology*, 27(4), 398–419. <https://doi.org/10.1080/02687038.2013.770817>
- Fyndanis, V., Lind, M., Varlokosta, S., Kambanaros, M., Soroli, E., Ceder, K., Grohmann, K. K., Rofes, A., Simonsen, H. G., Bjekić, J., Gavarró, A., Grohmann, K., Kuvac, J., Martinez Ferreiro, S., Munarriz, A., Pourquié, M., 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>
- Goodglass, H., & Kaplan, E. (1972). *Boston Diagnostic Aphasia Examination (BDAE)*. Lea & Febiger.
- Ham, S. (1989). Pasiv i norma [Passiv and norm]. *Jezik*, 37 (3), 65-76. <https://hrcak.srce.hr/152496>
- Hambleton, R. K. (1996). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229-244.
- 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>

- International Test Commission (2017). *The ITC Guidelines for Translating and Adapting Tests* (Second edition). Retrieved from www.InTestCom.org
- 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. <https://doi.org/10.1080/02687038.2013.805728>
- Jacquemot, Ch., Python, G. & Soroli, E. (in preparation). The Comprehensive Aphasia Test for French speakers (CAT-FR).
- Jensen, B. U., Norvik, M. I., & Simonsen, H. G. (2024). Statistics and psychometrics for the CAT-N: Documenting the Comprehensive Aphasia Test for Norwegian. *Aphasiology*, 38. <https://doi.org/10.1080/02687038.2023.2200132>
- Karakaya, S. P. Y., & Alparslan, Z. N. (2022). Sample Size in Reliability Studies: A Practical Guide Based on Cronbach's Alpha. *Psychiatry and Behavioral Sciences*, 12(3), 150. <https://doi.org/10.5455/PBS.20220127074618>
- Kertesz, A. (1982). *The Western aphasia battery*. Grune & Stratton.
- Kertesz, A. (2006). *Western aphasia battery-revised (WAB-R)*. Pro-Ed, 10.
- Kocjančič Antolík, T., & Širca Ule, P. (2017). Slovenian translation and adaptation of Comprehensive Aphasia Test. *Abstract Book*, 20. http://www.aphasiatrials.org/images/Abstract_book.pdf
- Kuvač Kraljević, J., Matić, A., & Lice, K. (2020). Putting the CAT-HR out: key properties and specificities. *Aphasiology*, 34(7), 820-839. <https://doi.org/10.1080/02687038.2019.1650160>
- Kuvač Kraljević, J., Matić, Škorić, A. Lice, K. (2023). Main Concepts in the Spoken Discourse of Persons with Aphasia: Analysis on a Propositional and Linguistic Level. *Languages* 8(120). <https://doi.org/10.3390/languages8020120>

- Long, A., Hesketh, A., Paszek, G., Booth, M., & Bowen, A. (2008). Development of a reliable self-report outcome measure for pragmatic trials of communication therapy following stroke: the Communication Outcome after Stroke (COAST) scale. *Clinical rehabilitation*, 22(12), 1083–1094. <https://doi.org/10.1177/0269215508090091>
- Lukács, Á., Rózsa, S., & Győri, M. (2012). A TROG pszichometriai jellemzőinek magyar vizsgálata, a normák kialakítása [The psychometric analysis of Hungarian data from the TROG]. In D. V. M. Bishop (Ed.), *TROG—Test for reception of grammar handbook* (pp. 47–86). OS Hungary.
- Martinez-Ferreiro, S., Quique, Y., Rodríguez, V. A., & Méndez Orellana, C. (2024). Spanish adaptation of the CAT (SP-CAT): linguistic and cultural properties. *Aphasiology*. <https://doi.org/10.1080/02687038.2024.2319362>
- Matić Škorić, A., Norvik, M. I., Kuvač Kraljević, J., Røste, I., Simonsen, H. G. (2023). Comprehensive Aphasia Test (CAT): Comparability of the Croatian and Norwegian versions. *Aphasiology*. <https://doi.org/10.1080/02687038.2023.2250516>
- Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: past studies and new evidence. *Cortex*, 45(6), 738–758. <https://doi.org/10.1016/j.cortex.2008.10.003>
- Maviş, İ., Arslan, S., & Aydin, Ö. (2020). Comprehension of word order in Turkish aphasia. *Aphasiology*, 34(8), 999-1015. <https://doi.org/10.1080/02687038.2019.1622646>
- 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. <https://doi.org/10.1080/02687038.2019.1622646>

- Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC medical research methodology*, *10*, 22. <https://doi.org/10.1186/1471-2288-10-22>
- Nerantzini, M., Varlokosta, S., Papadopoulou, D. & Bastiaanse, R. (2014). *Wh*-questions and relative clauses in Greek aphasia: Evidence from comprehension and production. *Aphasiology*, *28*(4), 490-514. <https://doi.org/10.1080/02687038.2013.870966>
- Osmánné Sági, J. (1991). Az afázia klasszifikációja és diagnosztikája. *Ideggyógyászati Szemle*, *44*(8), 339–362.
- Özdemir, Ş., Maviş, İ., & Tunçer, A. M. (2022). The Validity and Reliability of the Language Battery in Comprehensive Aphasia Test-Turkish (CAT-TR). *Journal of psycholinguistic research*, *51*(4), 789–802. <https://doi.org/10.1007/s10936-022-09850-2>
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, *76*(1), 1–25. <https://doi.org/10.1037/h0025327>
- Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Lawrence Erlbaum Associates, Inc.
- Peña-Casanova, J., Vinaixa, L., Diéguez-Vide, F., Gramunt-Fombuena, N., & Soler-Campillo, A. (2019). Assessment of aphasia: dialectical and cultural considerations in neurology. *Neurologia*, *37*(7), 596-603. <https://doi.org/10.1016/j.nrleng.2019.07.007>
- Porch, B. E. (1967). *Porch Index of Communicative Ability (Vol. 1). Theory and Development*. Consulting Psychologists Press.

- Porch, B. E. (1971). *Porch Index of Communicative Ability (Vol. 2): Administration, scoring and interpretation*. Consulting Psychologists Press.
- Pourquié, M., & Munarriz-Ibarrola, A. (this issue). Overcoming challenges in developing a standardised aphasia assessment tool in Basque, a language in the process of normalisation.
- Python, G., Danna, M., Pfyffer, E., Jacquemot, Ch. & Soroli, E. (2023). Standardization and validation of the French version of the Comprehensive Aphasia Test in L1 and L2 speakers in Switzerland. *Science of Aphasia Conference*, Sept 20, Nice, France.
- Rofes, A., Zakariás, L., Ceder, K., Lind, M., Blom Johansson, M., De Aguiar, V., Bjekić, J., Fyndanis, V., Gavarró, A., Simonsen, H. G., 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., Munarriz-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>
- Rofes, A., & Mahon, B. Z. (2021). Naming: Nouns and Verbs. In E. Mandonnet & G. Herbet (eds). *Intraoperative Mapping of Cognitive Networks*. Springer, Cham.
- Rosenkoetter, U., & Tate, R. L. (2018). Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools. *Brain Impairment*, 19(1), 103–118. <https://10.1017/BrImp.2017.29>
- Salmons, I., Rofes, A., & Gavarró, A. (2021). *Prova integral d'afàsia. Llibre d' items*. Servei de Publicacions de la UAB.
- Salmons, I., & Gavarró, A. (2022). Intervention effects in Catalan agrammatism. *Glossa: a journal of general linguistics*, 7(1), 1-32. <https://doi.org/10.16995/glossa.6348>

- Simmons-Mackie, N., Threats, T. T., & Kagan, A. (2005). Outcome assessment in aphasia: a survey. *Journal of communication disorders*, 38(1), 1–27. <https://doi.org/10.1016/j.jcomdis.2004.03.007>
- Spreen, O., & Risser, A. H. (2003). *Assessment of aphasia*. Oxford University Press.
- Swinburn, K., Porter, G., & Howard, D. (2005). *The Comprehensive Aphasia Test*. Psychology Press.
- Swinburn, K., Porter, G., & Howard, D. (2014). *Comprehensive Aphasia Test* (L. Haaber Hansen & M. Kaae Frederiksen, adaptation). Dansk Psykologisk Forlag.
- Swinburn, K., Best, W., Beeke, S., Cruice, M., Smith, L., Pearce Willis, E., Ledingham, K., Sweeney, J., & McVicker, S. J. (2018). 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., Kuvač Kraljević, J., Lice, K., & Matić, A. (2020). *Sveobuhvatni test za procjenu afazije CAT-HR*. [Comprehensive Aphasia Test - Croatian version]. Naklada Slap.
- 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. (2023). *The Comprehensive Aphasia Test* (2nd edition). Routledge.
- Terzi, A. & Nanousi, V. (2018). Intervention effects in the relative clauses of agrammatics: The role of gender and case. *Glossa*, 3(1), 17. 1–23. <https://doi.org/10.5334/GJGL.274>
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Lawrence Erlbaum.
- Tsiplakou, S. (2014). How mixed is a ‘mixed’ system? The case of the Cypriot Greek *koiné*. *Linguistic Variation*, 14, 1. <https://doi.org/10.1075/lv.14.1.07tsi>

- Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1(2), 89-99. <https://doi.org/10.1027/1016-9040.1.2.89>
- Van de Vijver, F. J. R. (2016). The use of standards in test development. In: Karl Schweizer and Christine DiStefano (Eds.), *Principles and Methods of Test Construction* (pp. 7-25). Hogrefe.
- Varlokosta, S., Nerantzini, M., Papadopoulou, D., Bastiaanse, R. & Beretta, A. (2014). Minimality effects in agrammatic comprehension: The role of lexical restriction and feature impoverishment. *Lingua*, 148, 80-94. <https://doi.org/10.1016/j.lingua.2014.05.013>
- Visch-Brink, E., Vandenborre, D., de Smet, H. J., & Mariën, P. (2014). *Comprehensive Aphasia Test - Nederlandse bewerking - Handleiding*. Pearson.
- Warburton, I. (1975). The passive in English and Greek. *Foundations of Language*, 13(4), 563-578.
- Whitworth, A., Webster, J., & Howard, D. (2013). *A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia: A clinician's guide* (2nd ed.). Psychology Press.
- World Health Organization. (2001). *International Classification of Functioning, Disability and Health (ICF)*. World Health Organization.
- Zakariás, L. & Lukács, Á. (2022). The Comprehensive Aphasia Test–Hungarian: adaptation and psychometric properties. *Aphasiology*, 36(9), 1127-1145. <https://doi.org/10.1080/02687038.2021.1937921>
- Zuazo, K. (2019). *Standard Basque and Its Dialects*. Routledge.