








## Critical Review

# Neuropsychological application of the International Test Commission Guidelines for Translation and Adapting of Tests

Christopher Minh Nguyen<sup>1</sup> , Shathani Rampa<sup>2</sup>, Mathew Staios<sup>3</sup> , T. Rune Nielsen<sup>4,5</sup>, Busisiwe Zapparoli<sup>6</sup>, Xinyi Emily Zhou<sup>7</sup>, Lingani Mbakile-Mahlanza<sup>8</sup>, Juliet Colon<sup>9</sup>, Alexandra Hammond<sup>10</sup>, Marc Hendriks<sup>11,12</sup>, Tumelo Kgolo<sup>8</sup>, Yesenia Serrano<sup>13</sup>, María J. Marquine<sup>14</sup> , Aparna Dutt<sup>5,15</sup>, Jonathan Evans<sup>16</sup>  and Tedd Judd<sup>17</sup> 

<sup>1</sup>Department of Psychiatry and Behavioral Health, The Ohio State University College of Medicine, Columbus, OH, USA, <sup>2</sup>Queens College and The Graduate Center, CUNY, Queens, NY, USA, <sup>3</sup>Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Melbourne, Australia, <sup>4</sup>Danish Dementia Research Centre, Copenhagen University Hospital – Rigshospitalet, Copenhagen, Denmark, <sup>5</sup>Neuropsychology & Clinical Psychology Unit, Duttanagar Mental Health Centre, Kolkata, WB, India, <sup>6</sup>The Hospital for Sick Children, Toronto, ON, Canada, <sup>7</sup>University of Missouri-Columbia, Columbia, MO, USA, <sup>8</sup>University of Botswana, Gaborone, Botswana, <sup>9</sup>Columbia University, New York, NY, USA, <sup>10</sup>Utah State University, Logan, UT, USA, <sup>11</sup>Neuropsychology and Rehabilitation Psychology, Donders Institute for Brain Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, <sup>12</sup>Academic Centre of Epileptology, Kempenhaeghe, Heeze, The Netherlands, <sup>13</sup>Department of Veterans Affairs, VISN04 Clinical Resource Hub, Pittsburgh, PA, USA, <sup>14</sup>Department of Psychiatry and Behavioral Sciences and the Duke Center for the Study of Aging and Human Development, Duke University School of Medicine, Durham, NC, USA, <sup>15</sup>School of Psychological Sciences, University of Bristol, Bristol, UK, <sup>16</sup>School of Health & Wellbeing, University of Glasgow, Glasgow, UK and <sup>17</sup>Universidad del Valle de Guatemala, Guatemala

## Abstract

**Objective:** The number of test translations and adaptations has risen exponentially over the last two decades, and these processes are now becoming a common practice. The International Test Commission (ITC) Guidelines for Translating and Adapting Tests (Second Edition, 2017) offer principles and practices to ensure the quality of translated and adapted tests. However, they are not specific to the cognitive processes examined with clinical neuropsychological measures. The aim of this publication is to provide a specialized set of recommendations for guiding neuropsychological test translation and adaptation procedures. **Methods:** The International Neuropsychological Society's Cultural Neuropsychology Special Interest Group established a working group tasked with extending the ITC guidelines to offer specialized recommendations for translating/adapting neuropsychological tests. The neuropsychological application of the ITC guidelines was formulated by authors representing over ten nations, drawing upon literature concerning neuropsychological test translation, adaptation, and development, as well as their own expertise and consulting colleagues experienced in this field. **Results:** A summary of neuropsychological-specific commentary regarding the ITC test translation and adaptation guidelines is presented. Additionally, examples of applying these recommendations across a broad range of criteria are provided to aid test developers in attaining valid and reliable outcomes. **Conclusions:** Establishing specific neuropsychological test translation and adaptation guidelines is critical to ensure that such processes produce reliable and valid psychometric measures. Given the rapid global growth experienced in neuropsychology over the last two decades, the recommendations may assist researchers and practitioners in carrying out such endeavors.

**Keywords:** Cross-cultural neuropsychology; test development; test translation; test adaptation; assessment; cultural diversity

(Received 1 June 2023; final revision 22 May 2024; accepted 24 May 2024)

The rapid expansion of global neuropsychology has led to an increasing need for translating, adapting, and norming tests across several languages and cultures (Kosmidis et al., 2012; Messinis et al., 2011). Research consistently demonstrates that tests and normative data originally designed for middle-class Western populations can lead to misclassifications when used with ethnic minorities or non-Western populations (Daugherty et al., 2017; Heaton et al., 2003). Misclassifications often arise from cultural and linguistic differences, varying levels and quality of education,

the use of non-representative data, culturally biased test content, and differences in test-taking attitudes (Rivera Mindt et al., 2020; Shuttleworth-Edwards, 2016). These issues affect a range of neuropsychological tests, including those for general intelligence, verbal and visual memory (Walker et al., 2009), language assessments (Patricacou et al., 2007), non-verbal visuo-construction tests (Nielsen & Jørgensen, 2013; Rosselli & Ardila, 2003), and executive functioning measures (Agranovich et al., 2011; Messinis et al., 2011).

**Corresponding author:** Christopher Nguyen; Email: [christopher.nguyen@osumc.edu](mailto:christopher.nguyen@osumc.edu)

**Cite this article:** Nguyen C.M., Rampa S., Staios M., Nielsen T.R., Zapparoli B., Zhou X.E., Mbakile-Mahlanza L., Colon J., Hammond A., Hendriks M., Kgolo T., Serrano Y., Marquine M.J., Dutt A., Evans J., & Judd T. (2024) Neuropsychological application of the International Test Commission Guidelines for Translation and Adapting of Tests. *Journal of the International Neuropsychological Society*, 30, 621–634, <https://doi.org/10.1017/S1355617724000286>

© The Author(s), 2024. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Efforts to tackle these challenges have involved creating culturally relevant test content and normative data for various cultural and ethnic groups globally, including minority groups in the United States (Norman *et al.*, 2011; Rivera Mindt *et al.*, 2020) and both majority and minority groups in Europe (33; Nielsen *et al.*, 2019). These initiatives have generally enhanced diagnostic accuracy. Nevertheless, the development of culturally appropriate neuropsychological tests and normative data remains an ongoing and iterative process, particularly when assessing ethnic minority and immigrant populations with limited education. Surprisingly, explicit guidelines for translating and adapting neuropsychological tests for cross-cultural use are currently lacking.

The field of test translations and adaptations has grown significantly, and there is a noticeable increase in the development of guidelines to improve these processes (Hernández *et al.*, 2020). Professional organizations such as the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) have joint efforts to sponsor the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) and issued guidance and standards related to testing, with a focus on fairness. This guideline approaches fairness in testing from different perspectives, including equitable treatment during testing, the absence of measurement bias, accessibility to the concepts being measured, and the validity of test scores for their intended purposes. While the specifics of their guidelines and recommendations may differ, their overarching goal is to promote fair and valid educational and psychological assessment practices.

One standard currently employed for such processes is the Guidelines for Translating and Adapting Tests developed by the International Test Commission (ITC). The ITC comprises 20 members representing national psychological associations, 65 affiliates, and over 700 individual members from 63 countries globally. The ITC's primary goal is to promote collaboration and information exchange among its members, dealing with psychological test development, distribution, and utilization issues. In 2005, the ITC introduced the Guidelines for Translating and Adapting Tests, which underwent a Second Edition revision in 2017. These guidelines offer a comprehensive framework to ensure the quality and validity of psychological tests when translated and adapted for use in diverse linguistic and cultural contexts. While the second edition of the ITC guidelines (International Test Commission, 2017) has made significant strides by offering practical suggestions, these recommendations are not specifically tailored to the theoretical underpinnings of cognitive processes related to neuropsychological assessment. Consequently, their applicability in this context is limited.

In response to these limitations, the Cultural Neuropsychology Special Interest Group of the International Neuropsychological Society formed a working group aimed at furthering the application of the ITC Guidelines for neuropsychology. This group has crafted a set of recommendations specifically designed to steer the translation and adaptation of neuropsychological tests. Authors from over ten different nations worldwide contributed to these guidelines. The process involved conducting a comprehensive review of literature and research experiences pertaining to the translation, adaptation, and development of neuropsychological tests.

The neuropsychological applications of the ITC Guidelines for Test Translation and Adaptation are presented by Judd *et al.* (2024) and summarized in the subsequent sections of this publication. For clarity, test translation is a simplified language

transfer that maintains accuracy, serving as one component of the broader adaptation process. In contrast, test adaptation encompasses evaluating a test's ability to measure the same concept in a different context effectively. This process involves selecting skilled translators, establishing translation criteria, making necessary accommodations, modifying the test format, ensuring equivalence, and conducting essential validity studies for a comprehensive assessment. The 18 ITC Guidelines are categorized into six sections: Pre-condition, Test Development, Confirmation, Administration, Scoring and Interpretation, and Documentation. The first three guidelines in the Pre-condition section emphasize the importance of making well-informed decisions before embarking on the translation or adaptation process. The Test Development section comprises five guidelines that discuss specific test adaptation procedures. Four guidelines in the Confirmation section address the systematic gathering of empirical evidence for assessing a test's equivalence, reliability, and validity across various linguistic and cultural contexts. The last three sections each contain two guidelines, encompassing the categories of Administration, Score Scales and Interpretation, and Documentation. A summary of each guideline is presented in Table 1.

### Pre-condition (PC) guidelines

#### ***PC-1 (guideline 1): obtain permission from the intellectual property rights owner***

When choosing tests to adapt that were originally created by an individual or corporation, it is important to take intellectual property and copyright laws into account. Intellectual property comprises two main subdivisions: (i) industrial property, which encompasses patents safeguarding inventions, industrial designs, trademarks, and commercial names, and (ii) copyright, which pertains to artistic and technology-based creations. Test adapters should recognize copyright law and agreements for the original test. Before starting a test adaptation, one should have a signed agreement from the intellectual property owner (i.e., the author or the publisher). The agreement should specify the modifications in the adapted test that will be acceptable regarding the original test's characteristics and should clarify who would own the intellectual property rights in the adapted version. Seeking permission to undertake test adaptation is common practice for several measures under copyright from major publishers (e.g., the Wechsler scales). Contacting the original test developers is advised to avoid intellectual property and copyright breaches and to gain clarity around these procedures.

#### ***PC-2 (guideline 2): evaluate the overlap between the test construct and item content***

Adapting and developing neuropsychological tests demand meticulous attention to cultural factors and the concept's equivalence when modifying tests for a specific population. For instance, one must consider how intelligence is conceptualized within the target population. In neuropsychology, intelligence is a frequently measured trait, yet a consensus on its definition remains elusive. An illustrative example comes from Kenya, where intelligence is delineated in the DhoLuo language using four distinct terms: *rieko* (knowledge and skills), *luoro* (respect), *winjo* (comprehension of real-life problem-solving), and *paro* (initiative), which deviate from the Western understanding (Grigorenko *et al.*, 2001).

**Table 1.** Summary of the International Test Commission Guidelines and its applications for neuropsychology

Guideline	Description	Neuropsychology application
<i>Pre-condition</i>		
PC-1 (1)	Obtain the necessary permission from the holder of the intellectual property rights relating to the test before any adaptation.	Iterative collaboration with test developers and intellectual property rights owners to maintain construct validity and test integrity.
PC-2 (2)	Evaluate that the amount of overlap in the definition and content of the construct measured by the test and the item content in the populations of interest is sufficient for the intended use (or uses) of the scores.	Establish construct equivalence through consultations with language and cultural experts in the population of interest.
PC-3 (3)	Minimize the influence of cultural and linguistic differences irrelevant to the test's intended uses in the populations of interest.	Test user consultation and piloting are important to identify potential test artifacts and examine the feasibility of the test for the target population.
<i>Test development</i>		
TD-1 (4)	Ensure that the translation and adaptation processes consider linguistic, psychological, and cultural differences in the intended populations by choosing experts with relevant expertise.	Consult with a multidisciplinary team with culture, language, and neuropsychological testing expertise during the translation and adaptation process.
TD-2 (5)	Use appropriate translation designs and procedures to maximize the suitability of the test adaptation in the intended populations.	Consider test-specific factors that affect test performance when evaluating the validity of test translations.
TD-3 (6)	Provide evidence that the test instructions and item content have similar meanings for all intended populations.	Translated/adapted instructions for test-takers and administrators should have the same understanding and purpose as for source language users.
TD-4 (7)	Provide evidence that the item formats, rating scales, scoring categories, test conventions, modes of administration, and other procedures are suitable for all intended populations.	Be mindful of cultural, linguistic, and educational diversity and include allowances for variations in test-taking approaches.
TD-5 (8)	Collect pilot data on the adapted test to enable item analysis, reliability assessment, and small-scale validity studies so that any necessary revisions to the adapted test can be made.	Qualitative and quantitative piloting is required to ensure method and item suitability and validity and reliability of the adapted test, respectively.
<i>Confirmation</i>		
C-1 (9)	Select a sample with characteristics relevant to the test's intended use and sufficient size and relevance for the empirical analyses.	Consider key elements that define the intended population and might impact test performance.
C-2 (10)	Provide relevant statistical evidence about the construct, method, and item equivalence for all intended populations.	Examine the test's validity and item equivalence to ensure that it and its items are useful in the new language and/or cultural group.
C-3 (11)	Provide evidence supporting the adapted test version's norms, reliability, and validity in the intended populations.	Validity should be demonstrated for each distinctive test with new normative data collected in different contexts.
C-4 (12)	Use an appropriate equating design and data analysis procedures when linking score scales from different language versions of a test.	Consider whether the test measures across a similar construct and if the test scores have strong psychometric properties.
<i>Administration</i>		
A-1 (13)	Prepare administration materials and instructions to minimize any culture- and language-related problems caused by administration procedures and response modes that can affect the validity of the inferences drawn from the scores.	Self-evaluate biases and preconceived notions about the population and consider how these may affect the approach to adaptation and validation procedures.
A-2 (14)	Specify testing conditions that should be followed closely in all populations of interest.	Consider test-specific factors and environment-specific elements that may affect the results or measured construct.
<i>Score scales and interpretation</i>		
SSI-1 (15)	Interpret any group score differences with reference to all relevant available information.	Consider cultural and contextual effects on test performance for a more accurate and comprehensive interpretation.
SSI-2 (16)	Only compare scores across populations when the level of invariance has been established on the scale on which scores are reported.	Interpret test scores cautiously when applying normative data not fully representative of the respondent's cultural or language background.
<i>Documentation</i>		
Doc-1 (17)	Provide technical documentation of any changes, including an account of the evidence obtained to support equivalence, when a test is adapted for use in another population.	Provide comprehensive details of the adaptation process and evidence of the psychometric properties of the adapted test for its intended uses in the new context.
Doc-2 (18)	Provide documentation for test users that will support good practice in using an adapted test with people in the context of the new population.	Provide a user manual to ensure proper administration, scoring, and interpretation of the test results in the new context.

Test adapters and developers should explore how culture influences other constructs, such as working memory and executive functioning, without assuming universality. Verifying that the construct measured by the test is part of the target population's vocabulary and examining the item-content equivalence of the test items are necessary initial steps. Test adapters and developers should know the target population's culture and intended uses of the test to establish construct equivalence early in

the adaptation process. Establishing construct equivalence requires experts in more than just both languages, and considerations for diagnostic validity and adaptive behavior should also be considered.

To achieve effective translation or adaptation of neuropsychological tests, involving language and cultural experts knowledgeable about the target population, including academics, professionals, and local informants, is recommended. Merely

**Table 2.** Test adaptation typology

Type	Description	Example
<i>Same test</i>	Translate and slightly adapt a test to provide psychometric equivalence in both the source and target languages. This allows for direct, quantitative comparisons across populations with what is considered “the same” test in both content and function.	WAIS Matrix Reasoning subtest which uses all the same items, but the order of item difficulty has been found to differ across languages, so the order of item administration has been adapted
<i>Test adapted version</i>	Translate and adapt a test to such a degree that it can serve similar functions in the target language but without psychometric equivalence or comparability, so it is considered a version of the test.	WAIS Information subtest in different languages, in which a minority of the items refer to historical figures specific to the country of testing.
<i>Test family</i>	Construct a test in the target language that follows the same design and principles or paradigm of the model test but with substantially new and culturally relevant content and procedures. Such a test can serve similar functions in the target language without psychometric equivalence or comparability.	Hong Kong Verbal Learning Test is part of the Word List Learning “family” of tests (Chan <i>et al.</i> , 2003; Chan, 2006)
<i>Indigenous test version</i>	Construct a test in the target language with a substantially new design and principles and culturally relevant content and procedures, but to measure the same construct(s) as in the original language.	“Which Car Test” for northern Australian Aborigines to assess executive functions and cognitive flexibility by using a social judgment problem specific to their cultural norms (Rock & Price, 2019)
<i>Indigenous measure</i>	Construct a test in the target language with a substantially new design, principles, culturally relevant content, and procedures to measure indigenous construct(s).	Measures of traditional beliefs, especially regarding health conditions and treatments, may also apply to measures of adaptive functions that apply only to specific cultural contexts; for example, measures of competence to stand trial only apply within a specific criminal justice system.

Note. The test adaptation typology is conceptualized as a continuum rather than discrete categories.

having experts present, however, does not guarantee the quality of the final product. Hence, efforts should focus on providing specific recommendations tailored to neuropsychological test translation or adaptation processes. For example, during translation, emphasis should be placed on achieving linguistic equivalence, while adaptation necessitates consideration of cultural equivalence, ultimately aiming for psychometric equivalence as the final result. These concepts are crucial to incorporate into the translation or adaptation process to ensure the validity and reliability of the test across different cultural and linguistic contexts.

### **PC-3 (guideline 3): minimize the influence of cultural and linguistic differences irrelevant to intended test uses in target populations**

To mitigate the impact of cultural and linguistic differences that are irrelevant to the test’s intended purpose on the results in the target populations, a comprehensive approach that involves reviewing, surveying, piloting, and debriefing test-takers, administrators, and focus groups is recommended. Multiple-choice and Likert scale response formats may be unfamiliar or ill-aligned with cultural norms in some cultures. For example, Chinese culture values moderation and may prefer neutral responses on Likert scales, even for positive emotions. Chinese and Japanese individuals tend to choose midpoints when acknowledging positive emotions, differing from European Americans (Wang *et al.*, 2008). Cultural variations also impact the value placed on speed in neuropsychological tests, willingness to guess, acceptance of synonyms and paraphrases, and knowledge of cardinal directions (Ardila, 2005).

There may also be intracultural variability in familiarity with test materials and techniques due to education policies, language policies, and writing systems. Test item content should have similar difficulty and familiarity relative to the culture and across all subcultures of anticipated application. Consulting the full range of intended test users early and addressing these differences through piloting is necessary. At this stage, assessing the feasibility, planning strategies, and establishing project objectives becomes

pivotal. Early in the project, it is essential to determine whether the most feasible strategy is to translate and make slight adaptations to a test, perform a full translation with adaptation, create a test in the target language following the model test’s paradigm, or construct a new test in the target language with a new design, as outlined in the Test Adaptation Typology provided in Table 2. To systematize the adaptation process, it is important to apply adaptation frameworks that are construct-driven, language-driven, culture-driven, theory-driven, and familiarity/recognizability-driven (Malda *et al.*, 2008).

### **Test development (TD) guidelines**

#### **TD-1 (guideline 4): consider linguistic, psychological, and cultural differences by consulting with content experts**

Many widespread languages may have different versions or dialects in different communities and parts of the world, such as English, Spanish, French, Arabic, Chinese, Swahili, and Quechua (Harris, 2022). Test developers should be conscious of whether they are developing localized or language-wide versions of their tests. To ensure that test materials are relevant and appropriate for the target population, the project team should include members with cultural, content, and testing expertise in the target language/population. The definition of an “expert” should include not only knowledge of (1) the languages involved, (2) the cultures, (3) the content of the test, and (4) general principles of testing but also (5) knowledge of the constructs of the test and their measurement. For example, take the item, “How are these two things alike, Orange-Banana?” A translator who is an expert in the content of the test and in Spanish localization would know to translate “banana” as “guineo” for the Caribbean but as “banano” for other parts of Latin America. But if “orange” had been translated” as “anaranjado” (the color orange rather than the fruit orange), this would back-translate fine and could appear correct to a content expert. But a construct expert would recognize that this item misses the intended construct (the similarity of being fruit. It would be like asking, “How are these two things alike? Blue-Banana”).



As demonstrated in the example, relying solely on individuals well-versed in testing fundamentals would fall short. The process should also require a deep understanding of test constructs. Collaboration with professionals possessing expertise in neuropsychological testing and a strong grasp of the language and culture, such as speech-language pathologists, for assessments related to language functions and aphasia tests can be highly beneficial. The translation and localization process should involve a multidisciplinary team with language and neuropsychological testing expertise to ensure the highest quality and validity of the adapted test materials.

***TD-2 (guideline 5): maximize test adaptation suitability for target populations through appropriate translation designs and procedures***

Neuropsychological assessments often rely on language-related processes to gauge various language skills, memory, and executive functions. When translating these assessments, prioritizing maintaining the core concepts rather than merely preserving semantic similarity is recommended for a successful adaptation. In some cases, the semantic content becomes irrelevant to the evaluation, a concept that may be unfamiliar to professional translators. For instance, assessments designed for aphasia and language-specific evaluations cannot be directly translated but require adaptation to linguistic features of the target language (Ivanova & Hallowell, 2013). Established guidelines exist for such cases, as illustrated by the phonemic discrimination subtest in the Bilingual Aphasia Test (Paradis & Libben, 1987). This subtest involves selecting the correct image from a set of options and necessitates careful selection of words with initial consonantal sound differences. A pure semantic translation would undermine the subtest's purpose.

Tasks involving verbal working memory and mental calculations are often used when evaluating attentional functioning. Cross-cultural and cross-linguistic studies reveal the influence of factors like word length, speech rates, and literacy on performance in verbal working memory tasks (Chan & Elliott, 2011; Chincotta & Underwood, 1996; Rosselli & Ardila, 2003). As such, adapting digit and mental math tasks to different languages requires considering both digit universality and phonemic, visual, and cultural distinctions. Adapting verbal memory tasks across languages is challenging due to stimuli familiarity and word frequency variations (Nell, 2000). Several strategies have been proposed for adapting neuropsychological assessments for culturally diverse groups, including substituting culturally appropriate items and ensuring ecological relevance. The goal is to have materials and tasks that are understandable in the target culture, cover an appropriate psychometric range, evaluate the intended cognitive concepts, and maintain clinical relevance (Vlahou et al., 2013). Factors like word length, material, and task familiarity are essential considerations in this process (Messinis et al., 2016).

Once these goals of the adaptation have been clearly identified, appropriate translation designs and procedures can be selected. The ITC Guidelines recommend using multiple designs, discussing the advantages and disadvantages of double translation and reconciliation, translation from more than one language, back translation, simultaneous development of multiple language versions, and designing a source version that minimizes translation problems. Hambleton and Zenisky (2010) list 25 empirically validated features recommended for translated tests. Back translation, although still used for some very technical and literal

translation, is now considered obsolete for most test translation purposes, in large part because back translation is errorful and favors semantic equivalence over cultural, functional, linguistic, construct, and psychometric equivalence (Colina et al., 2017; DuBay et al., 2022).

***TD-3 (guideline 6): provide evidence that the test instructions and item content have similar meanings for intended populations***

Developing culturally-congruent test instructions is a key priority during the translation and adaptation process. A simple translation of instructions may not suffice for equivalence in measurement and construct. Consider the example of the clock drawing test, where the phrase "10 after 11" in one setting might be expressed as "10 past 11" in another, as in Botswana. While this difference may seem subtle, it could lead some test-takers to misinterpret it as "10 minutes to 11" rather than the intended "10 after 11." In another example, nine of the twelve naming items on the Addenbrooke's Cognitive Examination (Bak & Mioshi, 2007) from the United Kingdom were found to be insufficiently familiar with the Arabic translation for Saudi Arabia (Al Salman, 2013). Test instructions and other item content should be pilot-tested and refined with the adaptation team until a consensus is reached.

To enhance the validity and cultural appropriateness of neuropsychological tests, stimuli, and test instructions, it is recommended that test developers involve individuals native to the local culture and proficient in the language of the test being translated and adapted, as well as bilingual individuals, to solicit feedback and insights into test equivalence. Customizing test administrations that align instructions with cultural and linguistic preferences also enhances acceptability and validity for diverse populations. Conducting interviews with participants and test administrators following the test administration will offer valuable qualitative insights into any potential discrepancies between intended and perceived meanings, thereby enhancing the overall validity of the assessments.

***TD-4 (guideline 7): provide evidence that the item formats, rating scales, scoring categories, test conventions, and modes of administration, and other procedures are suitable for intended populations***

This guideline emphasizes the necessity of familiarity with test items and administration methods to ensure unbiased testing results; however, in neuropsychological assessment, ensuring test suitability for all intended populations encompasses more than mere familiarity. Establishing semantic equivalence may not suffice in some instances. For example, Franzen et al., (2019) found systematic performance variations among culturally diverse individuals in the Netherlands on the Visual Association Test (Lindeboom & Schmand, 2003), noting that memory performance differed depending on whether participants were presented with black-and-white line drawings or pictures. Differences in neuropsychological performance can stem from the individual's approach influenced by cultural strategies. Given the significance of speed and reaction time in numerous neuropsychological assessments, wherein cultures prioritize performance speed differently (e.g., Ardila, 2005), additional attention might be necessary in crafting test instructions. Similarly, thoroughly assessing the scoring criteria during the adaptation phase is necessary. In the evaluation of the suitability of the Clock Drawing Test for Bengali-speaking individuals in India, researchers

observed that low-educated, healthy participants frequently combined English and Bengali when marking numbers on the clock face, prompting the implementation of an alternative scoring system (Crombie *et al.*, 2023).

In practical application, closely monitoring how test adaptation procedures may impact construct measurement is recommended. Key considerations include whether the adapted test maintains functional suitability and ensuring that alterations to test items and administration tools align with the original purpose of assessing the operational definition of the construct. Additionally, test developers should examine whether response categories accurately reflect the intended constructs, particularly regarding Likert scales, to ensure that high scores consistently represent the same values across diverse populations. Qualitative approaches like error analyses and feedback surveys can provide holistic insights during the piloting phase. Practice items play a crucial role in familiarizing test-takers with unfamiliar tasks without inducing practice effects, especially pertinent for domains like executive functions. Clear instructions for testers are imperative to gauge test-takers' understanding. Lastly, it is prudent to consider amending scoring criteria to optimize test validity in new linguistic or cultural contexts, drawing upon data from validity studies and replication efforts for informed decisions.

#### **TD-5 (guideline 8): collecting pilot data**

Before initiating the extensive data collection and test adaptation phases, it is advisable to employ pilot qualitative and quantitative procedures that include item analysis, reliability assessment, and small-scale validity studies. Qualitative piloting involves assessing comprehension, item suitability, and method appropriateness by engaging with test-takers and administrators to gain insights into potential validity and reliability issues. Following this, quantitative piloting delves deeper into evaluating method and item suitability, collecting extensive qualitative data on errors, the test-taking experience, and overall reception. The pilot sample should closely resemble the target standardization/normative population and the intended test users, such as individuals with dementia. Initial qualitative piloting may require adjustments to test instructions or items, while further quantitative piloting helps identify and adapt problematic items and procedures. Examples of detailed piloting techniques can be found in the Canadian Indigenous Cognitive Assessment development project (Jacklin *et al.*, 2020) and the procedures demonstrated by Franzen *et al.* (2022).

In this context, a comprehensive approach involving both qualitative and quantitative piloting is proposed. Piloting should involve individuals representing the intended test populations across various relevant dimensions, including age, education, cultural background, language proficiency, cognitive abilities, and so forth. Additionally, test administrators involved in piloting should mirror the intended demographic in terms of language proficiency, educational background, and testing experience.

Qualitative piloting should involve exploring the comprehension of standardized instructions and test materials, along with incorporating open-ended observation and reporting of testing experiences and challenges from test-takers and administrators through interviews, surveys, focus groups, and/or other appropriate techniques. This can allow for iterative adaptations and targeted repiloting of problematic procedures and materials. For example, during the adaptation of the Addenbrooke's Cognitive Examination III for the Bengali-speaking population in Kolkata,

India, it was observed that phonemic verbal fluency instructions were poorly understood. Adding any vowel form - aa, i, u, e, OI, O, OU - and others to the phoneme "Pa" made the instruction clear to both healthy participants and patients with dementia and optimized test performance (Dutt *et al.*, 2024).

Quantitative piloting can be done when iterative qualitative piloting and adaptations have clarified common misunderstandings. This piloting evaluates the psychometric characteristics, including item analysis, reliability, and validity of the scores obtained from the adapted test. Based on the results, necessary revisions should be made to the final version of the test. The pilot sample size must be adequate to conduct the necessary statistical analysis for the study.

#### **Confirmation (C) guidelines**

##### **C-1 (guideline 9): select a sample relevant to the intended use of the test and sufficient size for analysis**

A successful adaptation process for neuropsychological tests for a particular language or culture should involve a strategic plan to recruit participants for validation studies who are representative of the population with whom the test is intended to be used. It is essential that test adapters consider the elements that characterize the target population and could impact both test performance and language. Ensuring the presence of these characteristics in the validation sample is of particular significance, especially considering the cultural diversity within a language group and the influence of sociocultural factors on test performance.

Factors such as bilingualism/multilingualism, socioeconomic status, sex, ethnicity, nationality, quantity or quality of education, and familiarity with tests should be examined in relation to neuropsychological test performance. Examiner-related factors, such as stereotype threat and examiner bias, can also impact test performance and should be considered in developing studies to validate neuropsychological tests empirically (Thames *et al.*, 2013). To reduce the likelihood of stereotype threat impacting test performance, strategies during the test adaptation process can involve developing test items that represent diverse contexts, cultures, and identities. Before finalizing test items, pilot them with a diverse sample of individuals to assess whether any items evoke stereotype threat or exhibit biases. Furthermore, validity studies should be implemented to assess whether the test effectively predicts performance outcomes without being confounded by stereotype threat effects. Collecting data from large samples may not always be feasible, particularly in resource-limited settings where skilled examiners must individually administer tests. As such, analyses should be limited to those appropriate for the sample size that can be obtained, and the limitations of using small samples must be documented.

##### **C-2 (guideline 10): provide relevant statistical evidence about the construct equivalence, method equivalence, and item equivalence**

A main concern when translating or adapting a test to a new language and/or cultural group is to ensure the test's validity for its intended purpose. For instance, verifying if the same factor structure is essential when translating or adapting the Wechsler Adult Intelligence Scale (WAIS) – Fourth Edition (IV). This ensures that the test battery assesses identical constructs across various languages and cultural groups and validates the

appropriateness of computing and reporting the four index scores. Thus, Cockcroft et al. (2015) compared a multilingual, low socioeconomic group of black South African students with a predominantly white, British, monolingual, higher socioeconomic group and found that subtest scores loaded differently for the two groups. Exploratory factor analysis showed that a four-factor structure was most suitable for the South African data, albeit with Arithmetic loading more on the verbal comprehension factor than the working memory factor. Still, a three-factor structure better suited the British data, although interestingly, the same four-factor structure present in the United States standardization sample was also evident in the sample used for the United Kingdom WAIS-III standardization. Another recent example is the work of Staios et al. (2023), who conducted a confirmatory factor analysis of normative data from an elderly sample of Greek Australians on the Greek adaptation of the WAIS-IV and showed a good fit for the same four-factor solution as the original version.

Test validity can be examined by examining convergent and discriminant validity, and item equivalence can be assessed to ensure that all items are useful in the new language and/or cultural group. For example, a naming test may introduce method bias as some cultures may be less familiar with line-drawing representations of objects (e.g., Reis et al., 2006). Therefore, it is important to assess item equivalence between the original and new language and/or cultural group for all items, assessing both adopted words and object exemplars and, if necessary, replacing items. A broad construct measured by the original test may be equivalent in the new language and/or cultural group. Still, some or all the items in the test may not be equivalent in both the original and target languages. The relevance is especially notable in neuropsychological tests where items are arranged by increasing difficulty and discontinuation rules are in effect. Differential item functioning analysis should be carried out to examine whether items function differently in different samples.

**C-3 (guideline 11): provide evidence supporting the norms, reliability, and validity of adapted version**

To ensure the usefulness of neuropsychological tests, the quality of the normative data utilized for interpreting test performance plays a pivotal role. Test developers must assess the adapted test's reliability in its intended context, including internal consistency, test-retest, inter-rater, and parallel-version reliability. When adapting a test, assess the suitability of the original norms or consider collecting new normative data. Avoid assuming that a national identity always provides the ideal basis for norms, as cultural diversity can influence norm applicability (Guàrdia-Olmos et al., 2015). For example, cultural heterogeneity may challenge the generalizability of the norms developed within a specific cultural context (e.g., Bengali-speaking population residing in Kolkata, in Eastern India) compared to another context (e.g., Malayalam speaking population residing in Bangalore, in Southern India) within the same country (Das et al., 2006; Mathuranath et al., 2003). Factors such as acculturation often significantly influence test outcomes (Tan et al., 2021), leading to the non-uniform application of norms within a language or cultural group across different levels of acculturation and generations of immigrants. This includes international migration and migration within a country, particularly rural-to-urban movement. Ideally, acculturation should be assessed in diverse normative samples, covering individuals from various cultural backgrounds who have immigrated.

**C-4 (guideline 12): use an appropriate equating design and data analysis procedures when linking scores between language versions**

Complex cultural factors affecting neuropsychological test performance make it challenging to compare scores across different test versions (Casaletto & Heaton, 2017), as the primary purpose of these tests is not to compare scores across language populations directly but to serve similar functions in each population. The adaptation process should focus on whether the test measures a similar construct across language groups and whether the scores derived from a given population have strong psychometric properties. It is important to exercise caution when interpreting scores derived from different cultural or language groups. Neuropsychological scores are closely tied to the normative sample from which they were developed. Tests may function well clinically within a population when referencing appropriate population norms. Test adaptors should select the appropriate equating design and data analysis procedures if they wish to equate data and comparisons across populations.

**Administration (A) guidelines**

**A-1 (guideline 13): prepare administration materials and instructions to minimize any culture- and language-related influences on test administration**

The nature and purpose of neuropsychological testing may be a foreign concept for test-takers without prior assessment exposure, which can lead to misunderstanding and potentially inaccurate results. Such misunderstanding may be mistaken for cognitive impairment, test anxiety, poor test effort, malingering, or other concerns. To minimize these effects, neuropsychologists must take into consideration the sociocultural, demographic, and functional context of the intended population in both the research aspects of the test (e.g., construction, validation, and standardization) and their clinical applicability (e.g., providing clear and culturally appropriate test description, explanations to clients/patients to obtain their consent for testing, and administration procedures). Thus, the true scope of the guideline presented here must encompass the adaptation of instructions specific to a given test and the entire testing context. Practical considerations are presented in Table 3.

**A-2 (guideline 14): specify testing conditions to be followed in all populations of interest**

In the context of neuropsychological testing, it is essential to strike a delicate balance between adhering to standardized procedures and adapting tests to suit various clinical settings and populations. Consideration of factors includes distinctions such as inpatient versus outpatient settings, teleneuropsychology, and the needs of individuals with disabilities or specific cultural requirements. When adjusting tests for environments divergent from their original validation context, it is essential to consider the test-specific components and the aspects of the testing environment that could impact results or the construct being measured. These adaptations require detailed documentation of changes made to the test structure, materials, and administration. In inpatient settings, where testing conditions can be constrained by patient posture, limited space, potential distractions, and frequent interruptions, adaptations may need to alter test length, timing, practice items, or instruction to accommodate these challenges. When outpatient tests are used in inpatient settings without formal

**Table 3.** Administration guidelines: practical considerations

Cultural Considerations in Test Administration	Recommendations to Accommodate Culture- and Language-related Influences on Test Administration
<ul style="list-style-type: none"> <li>• <i>Orientation to Test-Takers:</i> Tailor the information provided to test-takers based on their cultural understanding of assessments and the evaluation process.</li> </ul>	<p>Provide test-takers with clear and culturally sensitive information about the evaluation process. This involves explaining the rationale, materials, methods, procedures, and expectations, often called “informed consent.” Adapting this explanation based on the intended culture’s understanding of assessment measures and neuropsychology can help reduce potential confounding factors associated with language, power differentials, and performance anxiety.</p>
<ul style="list-style-type: none"> <li>• <i>Clarification of Purpose and Roles:</i> Explain the role of neuropsychologists and the assessment’s purpose in a way that suits the cultural context. Clarify misconceptions and roles in the care system.</li> </ul>	<p>Clarifying the roles and goals of the assessment within the healthcare system is crucial, especially when test-takers may have varying levels of familiarity with the specific procedures and nature of neuropsychological evaluation. This step can involve distinguishing between healthcare providers and addressing mental health and intelligence evaluation misconceptions.</p>
<ul style="list-style-type: none"> <li>• <i>Patient-Provider Relationship:</i> Build rapport and understand the cultural factors that affect test-taker comfort and motivation.</li> </ul>	<p>The patient-provider relationship is another vital consideration, as establishing a rapport and understanding individuals’ cultural factors affecting test performance is essential. Cultural values, such as “familismo” and “personalism,” can influence the level of trust and comfort test-takers have in the evaluation process. Tailoring the approach to individualism, respect for authority, and personal space preferences is also vital to fostering a positive patient-provider relationship.</p>
<ul style="list-style-type: none"> <li>• <i>Task Explanation:</i> Clearly describe assessment tasks and provide instructions that accommodate the cultural healthcare model. Consider how instructions are framed and the level of clarity needed.</li> </ul>	<p>A clear explanation of the task is crucial to obtaining accurate data, including instructions that accommodate culturally accepted healthcare models and social dynamics. Considerations must be made regarding the clarity of instructions, the level of effort encouraged, and how instructions are adapted to the intended culture’s communication norms.</p>
<ul style="list-style-type: none"> <li>• <i>Practice Items:</i> Determine whether practice items are necessary, considering the type of test, its purpose, and population characteristics. Adapt practice items as needed.</li> </ul>	<p>The use of practice items should also be adjusted to align with the familiarity of the task in the culture and the expected level of comfort for independent task completion. While extra practice may improve test performance in some cultures, it must not compromise the construct and functionality of the test.</p>
<ul style="list-style-type: none"> <li>• <i>Cueing and Corrections:</i> Use culturally appropriate forms of prompting and feedback during task administration, considering the test’s construct validity.</li> </ul>	<p>Cueing and correction procedures should consider culturally appropriate forms of prompting and feedback. For instance, tasks requiring correction of errors or direct feedback may induce anxiety in cultures where shame-based approaches are prevalent. Providing clear explanations and normalizing errors can mitigate this.</p>
<ul style="list-style-type: none"> <li>• <i>Time/Discontinuation Procedures:</i> Explain time limits in a culturally sensitive way, considering the cultural perception of urgency.</li> </ul>	<p>Time limits and discontinuation procedures need to be culturally sensitive, considering the intended culture’s perception of time urgency and the explanation of time-limited tasks. The design should balance assessment time limits with cultural considerations to minimize anxiety or frustration.</p>
<ul style="list-style-type: none"> <li>• <i>Formatting of Test Materials:</i> Ensure test materials are accessible and culturally appropriate, considering factors like dyslexia or familiarity with electronic devices.</li> </ul>	<p>The formatting of test materials, including answer sheets, should account for potential difficulties, such as those faced by individuals with dyslexia. Electronic devices may limit access in some populations, so adaptability to paper-based formats should be considered.</p>
<ul style="list-style-type: none"> <li>• <i>Test Selection:</i> Address the challenges of assessing bilingual or bicultural individuals, considering language dominance, cultural factors, and normative data.</li> </ul>	<p>Test selection for bilingual or bicultural individuals requires assessing language dominance, bilingualism, and available normative data for selected tests. Cultural suitability, construct validity, and administration recommendations should guide the selection process.</p>

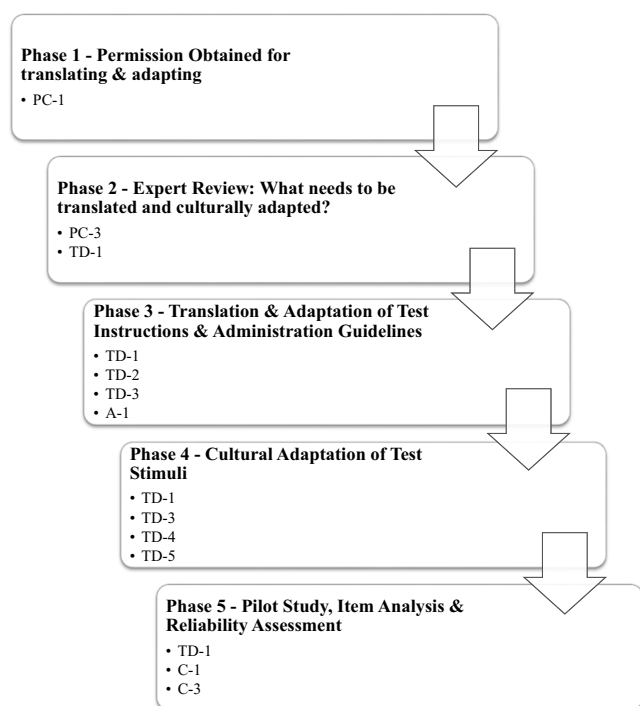
adaptations, administrators should consider the impact of the hospital environment on patient stress levels, privacy, and performance. For teleneuropsychology, where assessments are conducted remotely, additional considerations include potential equipment failures, display and audio issues, and internet connectivity problems. The test administrator should also be aware of distractions, interruptions, and privacy concerns in the test-taker’s home environment. Ensuring the security of test materials and addressing issues related to cultural and technological familiarity is crucial. Lastly, it is imperative to adapt tests for individuals with specific needs, such as those with hearing, visual, or motor impairments and those with lower levels of formal education. These adaptations may involve modifying tasks to accommodate verbal or visual mediation, adjusting practice item allowances, and offering appropriate feedback.

### Score scales and interpretation (SSI) guidelines

#### *SSI-1 (guideline 15): interpret group score differences with reference to all relevant available information*

Understanding and accounting for cultural factors is essential for accurately interpreting neuropsychological test performance. For example, educational systems can significantly differ between countries and cultures. Factors such as the importance placed on teaching syntax rules, geographic regions (urban vs. rural), gender roles, generational differences, and variations in educational facilities (private vs. public schools) can all impact an individual’s test performance. Secondly, levels and types of literacy and writing systems should be considered, as they influence an individual’s familiarity with test materials and strategies. Additionally, attitudes, motivations, expectations, and strategies related to





**Figure 1.** Example of a systematic approach based on test adaptation guidelines. Adapting a neuropsychological test involves several phases, starting with obtaining permission from the copyright holder and conducting an expert review. This is followed by translating or adapting the test, conducting a pilot study, performing item analysis, assessing reliability, collecting normative data, and producing an administrative manual. An example of this process can be seen here in the neuropsychological adaptation of the International Test Commission Guidelines. This example illustrates how the relevant guidelines can be applied throughout the different phases of the adaptation process. Here, Dutt et al. (2022) utilized a systematic approach for adapting the naming test from the Addenbrooke's Cognitive Examination III for the Bengali-speaking population in Kolkata, India. This figure is adapted from Dutt et al. (2024).

testing and attitudes toward timed tests and test procedures can significantly affect test outcomes. Immigration patterns and policies between countries, such as refugees or selection criteria based on skills, language, ethnicity, and age, can impact acculturation patterns and the representativeness of norms from one's home country. The level of acculturation in immigrant groups is a crucial factor to consider, as it influences how individuals navigate and respond to neuropsychological assessments. The phenomenon of stereotype threat in marginalized groups, which can lead to underperformance due to the fear of conforming to negative stereotypes, should also be considered. Moreover, other social determinants of health or non-medical factors that affect disease, treatment, and outcomes may play a significant role in test performance.

**SSI-2 (guideline 16): only compare scores across populations when scale invariance has been established**

Interpreting neuropsychological test results requires careful consideration of the available normative data, especially when assessing individuals from diverse cultural or linguistic backgrounds. Directly comparing raw scores among individuals with varying demographic characteristics, including cultural and language factors, may not yield valid results. Take, for example, the WAIS test, which maintains a common reporting scale but utilizes different raw scores depending on the country-specific or

cultural normative data applied. Even within the same language versions of the test, differences in normative data can lead to varying scaled scores, profiles, and diagnostic classifications across countries, as seen in the disparities between American, Canadian, Colombian, Mexican, and Spanish WAIS-IV norms (Duggan et al., 2019).

Due to the lack of content equivalence, substantial sample sizes across multiple populations, and research necessary for scalar or full score equivalence, direct score comparisons across populations should generally be avoided in neuropsychological testing. The limitation extends to comparing an individual's performance to a population they do not belong to, emphasizing the importance of considering the available validity evidence when interpreting test results. It is particularly challenging for multicultural and multilingual individuals, including older migrants, as neither the normative data from their country of origin nor the new country may accurately represent this population (Dutt et al., 2022; Plitas et al., 2009; Staios et al., 2023).

These concerns might not be relevant in criterion-based testing scenarios, where the assessment focuses on specific competencies or adaptive behaviors. In such cases, using scores and norms from a population that does not perfectly match the individual being evaluated could be deemed acceptable. For instance, in evaluations of driving safety or worker qualifications, it may be permissible to employ norms that do not precisely align with the individual's demographic background. However, even in these situations, it is prudent to proceed with caution. Considering alternative, context-specific assessments like on-road driving tests can ensure a more accurate evaluation of the individual's capabilities.

**Documentation (Doc) guidelines**

**Doc-1 (guideline 17): provide technical documentation of changes, including evidence of equivalence, when a test is adapted to a different population**

A comprehensive technical document and test manual are essential in neuropsychological test adaptation. This document should offer detailed insights into the adaptation process, providing evidence of the adapted test's reliability and validity within its intended new context. It should encompass a description of the normative data collection process, use of interpreters, characteristics of the normative sample, and the metrics used for test performance evaluation, such as scaled scores or T scores. Recommendations include presenting this technical documentation as a manual accompanying the test or a technical paper in a journal article. The document should address crucial aspects, including the adapted test's purpose in the new context, the relevance of the construct measured, the process of item translation or adaptation, ensuring familiarity of test formats for the target population, addressing factors influencing test performance, evidence from initial piloting, interpreter qualifications and roles, and thorough reliability and validity assessments. Additionally, it should detail the normative sample and data collection processes, offering demographic characteristics and outlining applicability limitations. The choice of metrics, analysis of normative data, rationale for norm provision, and user-friendly presentation of normative data should also be included. It is important to note that the specific content may vary depending on the test being adapted, and these guidelines offer a comprehensive framework for ensuring the integrity of the adaptation process.

**Table 4.** Neuropsychological application of International Test Commission Guidelines: criteria for evaluative checklist

Guidelines	Assessment Criteria	Not Applicable	Not Present	Mentioned but not adequately described (or not acceptable)	Well-described and empirical evidence presented
<b>Pre-condition</b>					
PC-1: Obtaining permission from the intellectual property rights owner.	<ul style="list-style-type: none"> <li>• Permission is obtained from the intellectual property rights owner before translating and adapting.</li> <li>• Neuropsychologists practicing in various countries are encouraged to reference copyright laws and agreements governing their standard practices.</li> </ul>				
PC-2: Evaluating overlap between test construct and item content.	<ul style="list-style-type: none"> <li>• Determine the degree to which the item content is consistent with the test construct in the population of interest.</li> <li>• Ensure that the construct exists in the vocabulary of the target population and examine the item-content equivalence of the test items.</li> <li>• Explore how culture influences test constructs without assuming universality.</li> </ul>				
PC-3: Minimize the influence of cultural and linguistic differences irrelevant to intended test uses in target populations.	<ul style="list-style-type: none"> <li>• To minimize the impact of cultural/linguistic differences on test performance, consult the full range of intended test users early and address these differences through piloting.</li> <li>• Review, survey, pilot, and debrief test-takers, administrators, and focus groups.</li> </ul>				
<b>Test development</b>					
TD-1: Consult with content experts to consider linguistic, psychological, and cultural differences.	<ul style="list-style-type: none"> <li>• Involve individuals with cultural, content, and testing expertise in the target language/population as part of the translation and localization process.</li> </ul>				
TD-2: Maximize test adaptation suitability for target populations through appropriate translation designs and procedures.	<ul style="list-style-type: none"> <li>• Have materials and tasks that are understandable in the target culture, cover an appropriate psychometric range, evaluate the intended cognitive concepts, and maintain clinical relevance.</li> </ul>				
TD-3: Provide evidence that the test instructions and item content have similar meanings for intended populations.	<ul style="list-style-type: none"> <li>• Pilot test and then refine the test instructions and other item content until a consensus is reached.</li> <li>• Conduct interviews with participants and test administrators following the test administration to solicit feedback on potential discrepancies.</li> </ul>				
TD-4: Provide evidence that the item formats, rating scales, scoring categories, test conventions, modes of administration, and other procedures are suitable for intended populations.	<ul style="list-style-type: none"> <li>• Ensure that test design, administration and scoring procedures, and modes of test interpretation are not biased against the intended population.</li> <li>• Minimize the potential for individuals to perform differently as a result of engaging strategies influenced by their culture.</li> </ul>				
TD-5: Collecting pilot data.	<ul style="list-style-type: none"> <li>• Engage in qualitative and quantitative piloting procedures.</li> <li>• Use a pilot sample closely resembling the target standardization/normative population and the intended test users.</li> </ul>				
<b>Confirmation</b>					
C-1: Select a sample relevant to the test's intended use and sufficient size for analysis.	<ul style="list-style-type: none"> <li>• Recruit participants for validation studies who are representative of the population with whom the test is intended to be used.</li> <li>• Representation should be according to variables known or suspected to significantly influence scores for the intended population and function measured (e.g., age, education, language)</li> </ul>				
C-2: Provide relevant statistical evidence about the construct equivalence, method equivalence, and item equivalence.	<ul style="list-style-type: none"> <li>• Examine and provide statistical data related to test properties in the target population.</li> <li>• Specifically consider whether all items are useful in the new language and/or cultural group and whether items might function differently in different samples.</li> </ul>				

(Continued)

Table 4. (Continued)

Guidelines	Assessment Criteria	Not Applicable	Not Present	Mentioned but not adequately described (or not acceptable)	Well-described and empirical evidence presented
C-3: Provide evidence supporting the norms, reliability, and validity of adapted version.	<ul style="list-style-type: none"> <li>Consider collecting new normative data or assessing the suitability of the original norms.</li> <li>Avoid assuming that a national identity always provides the ideal basis for norms, as cultural diversity can influence norm applicability.</li> </ul>				
C-4: Use an appropriate equating design and data analysis procedures when linking scores between language versions.	<ul style="list-style-type: none"> <li>Exercise caution when interpreting scores derived from different cultural/language groups.</li> <li>In the adaptation process, focus on whether the test measures a similar construct across language groups and whether the scores derived from a given population have strong psychometric properties.</li> </ul>				
<b>Administration</b>					
A-1: Prepare administration materials and instructions to minimize any culture- and language-related influences on test administration.	<ul style="list-style-type: none"> <li>Provide clear and culturally appropriate administration materials and instructions that take sociocultural, demographic, and functional context of the intended population into consideration.</li> </ul>				
A-2: Specify testing conditions to be followed in all populations of interest.	<ul style="list-style-type: none"> <li>Determine which testing conditions must be followed in all populations of interest to maintain test integrity.</li> <li>Consider test-specific elements and aspects of the testing environment that might influence results or the measured construct.</li> <li>Adapt tests for individuals with specific needs, such as hearing, visual, or motor impairments, and those with lower levels of formal education.</li> <li>Document any changes made to the testing protocol and consider ways these changes may impact the results obtained.</li> </ul>				
<b>Score scales and interpretation</b>					
SSI-1: Interpret group score differences with reference to all relevant available information.	<ul style="list-style-type: none"> <li>Understanding and accounting for cultural factors is essential for accurately interpreting neuropsychological test performance.</li> </ul>				
SSI-2: Only compare scores across populations when scale invariance has been established.	<ul style="list-style-type: none"> <li>Generally avoid direct score comparisons across populations due to the lack of content equivalence, substantial sample sizes across multiple populations, and research necessary for scalar or full score equivalence.</li> </ul>				
<b>Documentation</b>					
Doc-1: Provide technical documentation of changes, including evidence of equivalence, when a test is adapted to a different population.	<ul style="list-style-type: none"> <li>Create a comprehensive technical document and test manual that contains detailed information about the adaptation process and provides evidence of the adapted test's reliability and validity within its intended new context.</li> </ul>				
Doc-2: Provide documentation to support good practice in use of an adapted test in the target population.	<ul style="list-style-type: none"> <li>Create a detailed user manual to ensure proper administration, scoring, and interpretation of the test results in the new language or cultural group.</li> </ul>				

**Doc-2 (guideline 18): provide documentation to support good practice in the use of an adapted test in the target population**

A user manual is essential for all neuropsychological tests to ensure proper administration, scoring, and interpretation of the test results. It can serve as a vital resource, elucidating the rationale behind the test's original and new language or cultural group adaptation, substantiating the test's reliability and validity, furnishing precise administration and scoring instructions, and

offering guidance on score interpretation. The user manual should encompass several key elements, including an explanation of the original test's purpose and relevance in the new linguistic or cultural context, a concise overview of the adaptation process, and compelling evidence supporting the test's reliability and validity within the new context. It should also delve into detailed instructions for test administration, accounting for any context-specific nuances or differences from the original test, elucidate the scoring process, and provide insight into score interpretation tailored to the new language or cultural group. Additionally, the

manual should delineate whether the adapted version allows for direct population comparisons with the original test, specifying the basis for such a determination.

## Conclusion

Establishing a set of specific neuropsychological test translation, adaptation, and development guidelines is critical to ensure that such processes produce reliable and valid psychometric measures. The neuropsychological adaptation of the ITC guidelines can function as a practical guide to translate and adapt tests for use with a broad range of target populations. Following a systematic approach based on test adaptation guidelines is essential for several reasons. First, it helps assess the suitability of a test within a specific cultural context, even within individual nations. Additionally, it aids in preventing biases related to the test's construct, methods, and sampling. Adhering to a standardized adaptation process can significantly reduce the potential for these biases. A systematic and standardized adaptation process is necessary for developing robust and culturally sensitive assessment tools. A visual depiction of how this process may look is presented in Figure 1.

Although the suggested recommendations are intended for future test translation and adaptation procedures, neuropsychologists often rely on non-verbal tests (such as WAIS-IV Matrix and Block Design, Dot Counting) in their current practices with culturally diverse patients. While test translations may be available, securing these translations from the original authors may also be difficult. Other challenges may include dilemmas when working with interpreters unwilling to engage in on-spot translations. Typically, neuropsychologists face challenges accessing sufficient translations. Current practitioners are directed to the proposed systematic approach outlined in Figure 1. For example, during Phase 2, the question arises if a test needs to be translated or culturally adapted. If the translation test materials and stimuli are indicated, a recommended approach is for the practitioner to consult and collaborate with organizations such as the International Neuropsychological Society's Cultural Neuropsychology Special Interests Group throughout the translation process. Other strategies for leveraging currently available international data to estimate premorbid functioning may be helpful in providing further clinical context, and additional resources for working with culturally diverse individuals are available elsewhere (see Fujii, 2017).

Navigating the evolving landscape of cognitive assessment and cross-cultural neuropsychology involves acknowledging the iterative nature of the test adaptation process. Continuous refinement is vital to promoting systematic and accountable adaptation practices to progress the field. It is recommended that translators, adaptors, developers, reviewers, and editors actively employ these guidelines. Additionally, journal editors and grant review agencies are encouraged to consider including a self-accountability component in submission requirements for authors adapting tests, as detailed in Table 4. Practitioners, researchers, and authors are encouraged to demonstrate their commitment to these guidelines. This enhances the rigor of the adaptation process and ensures transparency and accountability, ultimately improving the reliability and validity of adapted tests.

Future iterations of the neuropsychology adaptation of the ITC guidelines can focus on ethical implications and the impact of technology on test development adaptation. Ethical considerations extend beyond scientific and clinical domains and include fairness, cultural sensitivity, and the avoidance of biases when creating or adapting tests. Addressing cultural biases is crucial to avoid unjust

outcomes and disparities. Transparency in test development and adaptation is essential for trust, and ethical test development requires a commitment to rigor, inclusivity, and continuous assessment. The impact of technology on neuropsychological test development and adaptation shows promise for revolutionizing various aspects of test creation. Computerized and digital assessments offer advantages like increased standardization, real-time data collection, and adaptive testing. Teleneuropsychology and remote assessments provide accessibility to individuals with geographical or physical limitations. Ethical considerations, such as data privacy, algorithmic biases, and maintaining the human element in assessment, remain critical as technology evolves. Researchers, practitioners, and policymakers must adapt to navigate the ethical, practical, and scientific implications to harness the full benefits of these technological advancements.

**Acknowledgments.** We gratefully acknowledge the thoughtful comments of Elaine Ballard, Jamie Berry, Kate Cockcroft, Andy Dawes, Sanne Franzen, Kristin Jacklin, Michel Paradis, Porrselvi A.P., Parisuth Sumranasub, and Sharon Truter, that have been helpful to our process.

**Funding statement.** The authors of this paper are members of the Cultural Neuropsychology Special Interest Group (CN-SIG), Assessment Workgroup of the International Neuropsychological Society (INS). We would like to express our appreciation for the inspiration and support provided by the many neuropsychologists and allied professionals working to translate, adapt, and develop neuropsychological tests, assessments, and other tools to serve unreached populations; the CN-SIG as a whole; INS administration, and especially the ITC for their pioneering work in developing and updating the ITC Guidelines for Translating and Adapting Tests, and for providing feedback on the current document. The views expressed in this paper are those of the authors and may not necessarily represent the views of the INS.

**Competing interests.** None.

## References

- Al Salman, A. S. A. (2013) The Saudi Arabian Adaptation of the Addenbrooke's Cognitive Examination – Revised (Arabic ACER). PhD thesis. <http://theses.gla.ac.uk/4706>
- Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review*, 15(4), 185–195. <https://doi.org/10.1007/s11065-005-9180-y>
- Bak, T. H., & Mioshi, E. (2007). A cognitive bedside assessment beyond the MMSE: the Addenbrooke's Cognitive Examination. *Practical Neurology*, 7(4), 245–249.
- Casaleto, K. B., & Heaton, R. K. (2017). Neuropsychological assessment: Past and future. *Journal of the International Neuropsychological Society*, 23(9–10), 778–790. <https://doi.org/10.1017/S1355617717001060>
- Chan A. S. (2006). *Hong Kong List Learning Test*. 2nd Edn. Department of Psychological and Integrative Neuropsychological Rehabilitation Center.
- Chan A. S., Shum D., & Cheung R. W. (2003). Recent development of cognitive and neuropsychological assessment in Asian countries. *Psychological Assessment*, 15(3), 257–267. <https://doi.org/10.1037/1040-3590.15.3.257>
- Chan, M. E., & Elliott, J. M. (2011). Cross-linguistic differences in digit memory span. *Australian Psychologist*, 46(1), 25–30.
- Chincotta, D., & Underwood, G. (1996). Mother tongue, language of schooling and bilingual digit span. *British Journal of Psychology*, 87(2), 193–208.
- Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). A cross-cultural comparison between south African and british students on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III). *Frontiers in Psychology*, 6, 297. <https://doi.org/10.3389/fpsyg.2015.00297>
- Colina, S., Marrone, N., Ingram, M., & Sánchez, D. (2017). Translation quality assessment in health research: A functionalist alternative to back-translation. *Evaluation & The Health Professions*, 40(3), 267–293. <https://doi.org/10.1177/0163278716648191>



- Agranovich, A. V., Panter, A. T., Puente, A. E., & Touradji, P. (2011). The culture of time in neuropsychological assessment exploring the effects of culture-specific time attitudes on timed test performance in Russian and American samples. *Journal of the International Neuropsychological Society*, 17(4), 692–701. <https://doi.org/10.1017/S1355617711000592>
- Crombie, M., Dutt, A., Dey, P., Nandi, R., & Evans, J. (2023). Examination of the validity of the 'Papadum test': An alternative to the Clock Drawing Test for people with low levels of education. *The Clinical Neuropsychologist*, 37(5), 1025–1042. <https://doi.org/10.1080/13854046.2022.2047789>
- Das, S. K., Banerjee, T. K., Mukherjee, C. S., Bose, P., Hazra, A., Dutt, A., Das, S., Chaudhuri, A., & Raut, D. K. (2006). An urban community-based study of cognitive function among non-demented elderly population in India. *Neurology Asia*, 11, 37–48.
- Daugherty, J. C., Puente, A. E., Fasfous, A. F., Hidalgo-Ruzzante, N., & Pérez-García, M. (2017). Diagnostic mistakes of culturally diverse individuals when using north American neuropsychological tests. *Applied Neuropsychology: Adult*, 24(1), 16–22. <https://doi.org/10.1080/23279095.2015.1036992>
- DuBay, M., Sideris, J., & Rouch, E. (2022). Is traditional back translation enough? Comparison of translation methodology for an ASD screening tool. *Autism research : official journal of the International Society for Autism Research*, 15(10), 1868–1882. <https://doi.org/10.1002/aur.2783>
- Duggan, E. C., Awakon, L. M., Loaliza, C. C., & Garcia-Barrera, M. A. (2019). Contributing towards a cultural neuropsychology assessment decision-making framework: Comparison of WAIS-IV norms from Colombia, Chile, Mexico, Spain, United States, and Canada. *Archives of Clinical Neuropsychology*, 34(5), 657–681. <https://doi.org/10.1093/arclin/acy074>
- Dutt, A., Evans, J. J., & Fernandez, A. L. (2022). Challenges for neuropsychology in the global context. In *Understanding cross-cultural neuropsychology, science, testing and challenges (current issues in neuropsychology)*. Routledge | Taylor & Francis Group.
- Dutt, A., Nandi, R., Venkatesh, R. K., Rao, P. S., Bhargava, P., Gopkrishnan, S., Bose, A., Ghosh, A., & Evans, J. J. (2022). A systematic approach to reduce cultural bias: An illustration from the adaptation of the Addenbrooke's Cognitive Examination III for the Bengali speaking population in India. *Alzheimer's & Dementia*, 18(S7), e067325. <https://doi.org/10.1002/alz.067325>
- Dutt, A., Nandi, R., Venkatesh, R. K., Rao, P. S., Bhargava, P., Gopkrishnan, S., Bose, A., Ghosh, A., & Evans, J. J. Application of the International Test Commission (ITC) guidelines in reducing bias in test adaptation: An illustration from the Addenbrooke's Cognitive Examination III for the Bengali speaking population in India. In: *Paper presented at the 52nd annual North American meeting of the international neuropsychological society*. 2024.
- Franzen, S., van den Berg, E., Bossenbroek, W., Kranenburg, J., Scheffers, E. A., van Hout, M., van de Wiel, L., Goudsmit, M., van Bruchem-Visser, R.I., van Hemmen, J., Jiskoot, L. C., & Papma, J. M. (2022). Neuropsychological assessment in the multicultural memory clinic: Development and feasibility of the TULIPA battery. *The Clinical Neuropsychologist*, 37(1), 60–80. <https://doi.org/10.1080/13854046.2022.2043447>
- Franzen, S., van den Berg, E., Kalkisim, Y., van de Wiel, L., Harkes, M., van Bruchem-Visser, R. L., de Jong, F. J., Jiskoot, L. C., & Papma, J. M. (2019). Assessment of visual association memory in low-educated, non-Western immigrants with the modified visual association test. *Dementia and Geriatric Cognitive Disorders*, 47(4–6), 345–354. <https://doi.org/10.1159/000501151>
- Fujii, D. (2017). *Conducting a culturally informed neuropsychological evaluation*. American Psychological Association. <https://doi.org/10.1037/15958-000>
- Grigorenko, E. L., Geissler, P. W., Prince, R., Okatcha, F., Nokes, C., Kenny, D. A., Bundy, D. A., & Sternberg, R. J. (2001). The organisation of Luo conceptions of intelligence: A study of implicit theories in a Kenyan village. *International Journal of Behavioral Development*, 25(4), 367–378. <https://doi.org/10.1080/01650250042000348>
- Guardia-Olmos, J., Peró-Cebollero, M., Rivera, D., & Arango-Lasprilla, J. C. (2015). Methodology for the development of normative data for ten Spanish-language neuropsychological tests in eleven Latin American countries. *Neuro Rehabilitation*, 37(4), 493–499. <https://doi.org/10.3233/NRE-151277>
- Hambleton, R. K., & Zenisky, A. (2010). Translating and adapting tests for cross-cultural assessment. In D. Matsumoto, & F. van de Vijver (Eds.), *Cross-cultural research methods* (pp. 46–74). Cambridge University Press.
- Harris, S. (2022). *Translation vs. localization vs. transcreation: Is there a difference?*. Argos Multilingual. <https://www.argosmultilingual.com/blog/translation-localization-difference>
- Heaton, R. K., Taylor, M. J., & Manly, J. J. (2003). Demographic effects and use of demographically corrected norms for the WAIS-III and WMS-III. In: D. Tulsky, D. Saklofske, G. Chlune, R. Heaton, R. Ivnik, R. Bornstein, A. Prifitera, & M. Ledbetter (Eds.), *Clinical Interpretation of the WAIS-III and WMS-III (Practical Resources for the Mental Health Professional)* (pp. 183–210). Academic Press.
- Hernández, A., Hidalgo, M. D., Hambleton, R. K., & Gómez-Benito, J. (2020). International Test Commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(3), 390–398. <https://doi.org/10.7334/psicothema.2019.306>
- International Test Commission (2017). *The ITC Guidelines for translating and adapting tests* (Second Edition). Retrieved June 20, 2024, from [https://www.intestcom.org/files/guideline\\_test\\_adaptation\\_2ed.pdf](https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf).
- 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>
- Jacklin, K., Pitawanakwat, K., Blind, M., O'Connell, M. E., Walker, J., Lemieux, A. M., Warry, W., & Phinney, A. (2020). Developing the Canadian Indigenous Cognitive Assessment for use with indigenous older Anishinaabe adults in Ontario, Canada. *Innovation in Aging*, 4(4), igaa038. <https://doi.org/10.1093/geroni/igaa038>
- Judd, T., Colon, J., Dutt, A., Evans, J., Hammond, A., Hendriks, M., Kgoro, T., Marquine, M., Mbakile-Mahlanza, I., Nielsen, R., Nguyen, C., Rampa, S., Serrano, Y., Staios, M., Zapparoli, B., & Zhou, E. (2024). *Neuropsychological application of the International Test Commission's (ITC) guidelines for translating and adapting tests*. <https://the-ins.org/wp-content/uploads/2024/01/INS-SIG-Assessment-Workgroup-2023-ITC-Guidelines-Neuropsychology-Application.pdf>.
- Kosmidis, M. H., Bozikas, V. P., & Vlahou, C. H. (2012). *Neuropsychological test battery*. Lab of Cognitive Neuroscience, School of Psychology, Aristotle University of Thessaloniki.
- Lindeboom, J., & Schmand, B. (2003). *Visual Association Test*. PITS.
- Malda, M., van de Vijver, F. J. R., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, 50(4), 451–468.
- Mathuranath, P. S., George, A., Cherian, P. J., Alexander, A. L., Sarma, S. G., & Sarma, P. S. (2003). Effects of age, education and gender on verbal fluency. *Journal of Clinical and Experimental Neuropsychology*, 25(8), 1057–1064. <https://doi.org/10.1076/jcen.25.8.1057.16736>
- Messinis, L., Malegiannaki, A.-C., Christodoulou, T., Panagiotopoulos, V., & Papatanasopoulos, P. (2011). Color Trails Test: Normative data and criterion validity for the Greek adult population. *Archives of Clinical Neuropsychology*, 26(4), 322–330. <https://doi.org/10.1093/arclin/acr027>
- Messinis, L., Nasios, G., Mougias, A., Politis, A., Zampakis, P., Tsiamaki, E., Malefaki, S., Gourzis, P., & Papatanasopoulos, P. (2016). Age and education adjusted normative data and discriminative validity for Rey's Auditory Verbal Learning Test in the elderly Greek population. *Journal of Clinical and Experimental Neuropsychology*, 38(1), 23–39. <https://doi.org/10.1080/13803395.2015.1085496>
- Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice*. Lawrence Erlbaum Associates Publishers.
- Nielsen, T. R., & Jørgensen, K. (2013). Visuoconstructional abilities in cognitively healthy illiterate Turkish immigrants: A quantitative and qualitative investigation. *The Clinical Neuropsychologist*, 27(4), 681–692. <http://doi.org/10.1080/13854046.2013.767379>
- Nielsen, T. R., Segers, K., Vanderaspolden, V., Beinhoff, U., Minthon, L., Pissioti, A., Bekkhus-Wetterberg, P., Bjørkløf, G. H., Tsolaki, M., Gkioka, M., & Waldemar, G. (2019). Validation of a European Cross-Cultural Neuropsychological Test Battery (CNTB) for evaluation of dementia. *International Journal of Geriatric Psychiatry*, 34(1), 144–152. <https://doi.org/10.1002/gps.5002>

- Norman, M. A., Moore, D. J., Taylor, M., Franklin D. Jr, Cysique, L., Ake, C., Lazzarotto, D., Vaida, F., Heaton, R. K., & the HNRC Group (2011). Demographically corrected norms for African Americans and caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. *Journal of Clinical and Experimental Neuropsychology*, 33(7), 793–804.
- Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Psychology Press.
- Patricacou, A., Psallida, E., Pring, T., & Dipper, L. (2007). The Boston Naming Test in Greek: Normative data and the effects of age and education on naming. *Aphasiology*, 21(12), 1157–1170. <https://doi.org/10.1080/02687030600670643>
- Plitas, A., Tucker, A., Kritikos, A., Walters, I., & Bardenhagen, F. (2009). Comparative study of the cognitive performance of Greek Australian and Greek national elderly: Implications for neuropsychological practice. *Australian Psychologist*, 44(1), 27–39. <https://doi.org/10.1080/00050060802587694>
- Reis, A., Fáisca, L., Ingvar, M., & Petersson, K. M. (2006). Color makes a difference: Two-Dimensional object naming in literate and illiterate subjects. *Brain and Cognition*, 60(1), 49–54.
- Rivera Mindt, M., Marquine, M. J., Aghvinian, M., Paredes, A. M., Kamalyan, L., Suárez, P., & Cherner, M. (2020). The neuropsychological norms for the U.S.-Mexico border region in Spanish (NP-NUMBRs) project: Overview and considerations for life span research and evidence-based practice. *The Clinical Neuropsychologist*, 35(2), 1–15. <https://doi.org/10.1080/13854046.2020.1794>
- Rock, D., & Price, I. R. (2019). Identifying culturally acceptable cognitive tests for use in remote northern Australia. *BMC Psychol*, 7(1), 62. <https://doi.org/10.1186/s40359-019-0335-7>
- Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, 52(3), 326–333. [https://doi.org/10.1016/s0278-2626\(03\)00170-2](https://doi.org/10.1016/s0278-2626(03)00170-2)
- Shuttleworth-Edwards, A. B. (2016). Generally representative is representative of none: Commentary on the pitfalls of IQ test standardization in multicultural settings. *The Clinical Neuropsychologist*, 30(7), 975–998. <https://doi.org/10.1080/13854046.2016.120401>
- Staios, M., Kosmidis, M. H., Nielsen, T. R., Kokkinis, N., Stogoannidou, A., March, E., & Stolwyk, R. J. (2023). *The Wechsler Adult Intelligence Scale-Fourth Edition, Greek adaptation (WAIS-IV GR): Confirmatory factor analysis and specific reference*. Group Normative Data for Greek Australian Older Adults. Australian Psychologists. <https://doi.org/10.1080/00050067.2023.2179387>
- Tan, Y. W., Burgess, G. H., & Green, R. J. (2021). The effects of acculturation on neuropsychological test performance: A systematic literature review. *The Clinical Neuropsychologist*, 35(3), 541–571. <https://doi.org/10.1080/13854046.2020.1714740>
- Thames, A. D., Hinkin, C. H., Byrd, D. A., Bilder, R. M., Duff, K. J., Mindt, M. R., Arentoft, A., & Streiff, V. (2013). Effects of stereotype threat, perceived discrimination, and examiner race on neuropsychological performance: Simple as black and white? *Journal of the International Neuropsychological Society*, 19(5), 583–593. <https://doi.org/10.1017/s1355617713000076>
- Vlahou, C. H., Kosmidis, M. H., Dardagani, A., Tsotsi, S., Giannakou, M., Giakoulidou, A., Zervoudakis, E., & Pontikakis, N. (2013). Development of the Greek Verbal Learning Test: Reliability, construct validity, and normative standards. *Archives of Clinical Neuropsychology*, 28(1), 52–64. <https://doi.org/10.1093/arclin/acs099>
- Walker, A. J., Batchelor, J., & Shores, A. (2009). Effects of education and cultural background on performance on WAIS-III, WMS-III, WAIS-R and WMS-R measures: Systematic review. *Australian Psychologist*, 44(4), 216–223. <https://doi.org/10.1080/00050060902833469>
- Wang, R., Hempton, B., Dugan, J. P., & Komives, S. R. (2008). Cultural differences: Why do Asians avoid extreme responses? *Survey Practice* 1(3). <https://doi.org/10.29115/SP-2008-0011>