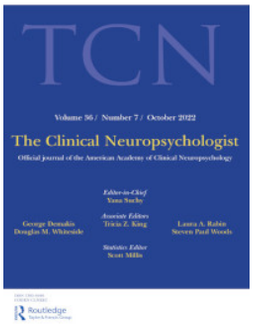

The Clinical Neuropsychologist

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ntcn20

Neuropsychological assessment in the multicultural memory clinic: Development and feasibility of the TULIPA battery

Sanne Franzen, Esther van den Berg, Willemijn Bossenbroek, Judi Kranenburg, Esther A. Scheffers, Moniek van Hout, Lotte van de Wiel, Miriam Goudsmit, Rozemarijn L. van Bruchem-Visser, Judy van Hemmen, Lize C. Jiskoot & Janne M. Papma

To cite this article: Sanne Franzen, Esther van den Berg, Willemijn Bossenbroek, Judi Kranenburg, Esther A. Scheffers, Moniek van Hout, Lotte van de Wiel, Miriam Goudsmit, Rozemarijn L. van Bruchem-Visser, Judy van Hemmen, Lize C. Jiskoot & Janne M. Papma (2022): Neuropsychological assessment in the multicultural memory clinic: Development and feasibility of the TULIPA battery, The Clinical Neuropsychologist, DOI: 10.1080/13854046.2022.2043447

To link to this article: https://doi.org/10.1080/13854046.2022.2043447

Published online: 27 Feb 2022.

THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2022.2043447

Neuropsychological assessment in the multicultural memory clinic: Development and feasibility of the TULIPA battery

Sanne Franzen[a], Esther van den Berg[a], Willemijn Bossenbroek[a,b], Judi Kranenburg[c], Esther A. Scheffers[c], Moniek van Hout[b], Lotte van de Wiel[d], Miriam Goudsmit[e], Rozemarijn L. van Bruchem-Visser[f], Judy van Hemmen[a], Lize C. Jiskoot[a], and Janne M. Papma[a]

[a]Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands; [b]Department of Medical Psychology, Medisch Spectrum Twente, Enschede, The Netherlands; [c]Department of Internal Medicine, Haaglanden Medical Center, The Hague, The Netherlands; [d]Department of Medical Psychology, Maasstad Ziekenhuis, Rotterdam, The Netherlands; [e]Department of Psychiatry and Medical Psychology, OLVG, Amsterdam, The Netherlands; [f]Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands

ABSTRACT

Objective: Neuropsychological assessment of culturally diverse populations is hindered by barriers in language, culture, education, and a lack of suitable tests. Furthermore, individuals from diverse backgrounds are often unfamiliar with being cognitively tested. The aim of this study was to develop a new neuropsychological test battery and study its feasibility in multicultural memory clinics.

Method: Composition of the TULIPA battery (Towards a Universal Language: Intervention and Psychodiagnostic Assessment) entailed a literature review and consultation with experts and individuals from diverse backgrounds. Feasibility was investigated by examining administration and completion rates and the frequency of factors complicating neuropsychological assessment in 345 patients from 37 countries visiting four multicultural memory clinics in the Netherlands.

Results: The test battery included existing tests such as the Cross-Cultural Dementia screening (CCD), Rowland Universal Dementia Assessment Scale (RUDAS), tests from the European Cross-Cultural Neuropsychological Test Battery, and newly developed tests. Completion rates for the test battery were generally high (82%–100%), except for CCD Dots subtest B (58%). Although tests of the "core" TULIPA battery were administered often (median: 6 of 7, IQR: 5–7), supplementary tests were administered less frequently (median: 1 of 9; IQR: 0–3). The number of administered tests correlated with disease severity (RUDAS, ρ=.33, adjusted $p < .001$), but not with other patient characteristics. Complicating factors were observed frequently, e.g. suboptimal effort (29%–50%), fatigue (29%), depression (37%–57%).

ARTICLE HISTORY

Received 15 October 2021
Accepted 13 February 2022
Published online 28 February 2022

KEYWORDS

Cultural diversity; literacy; education; neurodegenerative disease; cultural competency

CONTACT Sanne Franzen s.franzen@erasmusmc.nl Department of Neurology, Erasmus MC University Medical Center, Nf-331, Dr. Molewaterplein 40, Rotterdam, 3015 GD, The Netherlands.

Supplemental data for this article is available online at https://doi.org/10.1080/13854046.2022.2043447.

Conclusions: The TULIPA test battery is a promising new battery to assess culturally diverse populations in a feasible way, provided that complicating factors are taken into account.

Introduction

Over the past decades, Europe has become increasingly diverse. Many individuals from culturally, educationally, and linguistically diverse backgrounds living in Europe—in particular the "guest workers" who came to Europe as labor immigrants from Turkey and North Africa between 1950–1974—are at a higher risk of cognitive impairment, due to a higher prevalence of age-related medical conditions such as diabetes mellitus (Kunst et al., 2011), stroke (Kunst et al., 2011), and dementia (Selten et al., 2020). Neuropsychologists in Europe will therefore increasingly encounter such individuals from diverse backgrounds in their clinical practice.

The cognitive assessment of individuals from culturally, educationally, and linguistically diverse backgrounds in memory clinics can be hindered by several factors. First, communication can be hampered by language barriers and differences in communication styles, such as the level of directness or differences in perceptions of when it is considered (in)appropriate to speak openly (Fujii, 2018). Assessment with an interpreter is often necessary, but formal interpreters are inconsistently used across Europe, particularly due to a lack of funding (Franzen et al., 2021). The use of informal interpreters (particularly relatives) may be problematic due to the exclusion of the patient from the conversation, an interpreter's lack of familiarity with medical terminology, difficulties in assessing a patient's level of insight, and shame/embarrassment in discussing sensitive topics (Kilian et al., 2014; Manly & Espino, 2004; Zendedel et al., 2018a, 2018b). Second, differences in culture can impact perceptions of what is considered relevant information or what is considered "good" performance, as well as whether individuals are familiar with the stimuli used in tests (Ardila, 2007; Chen et al., 2014). Third, education—particularly literacy—influences processes such as abstract thinking/reasoning skills, perception, the ability to name black-and-white line drawings, and performance on tasks that require participants to draw, read, or count (Ardila et al., 2010; Nielsen & Jørgensen, 2013; Ostrosky-Solis et al., 1998; Reis et al., 2001, 2006).

The abovementioned barriers to neuropsychological testing may coincide with a lack of experience with being tested. This may result in incorrect expectations about neuropsychological assessment in general (e.g. length, content), a lack of understanding of the examiner's role, or (disproportionate) nervousness or fear to look "stupid" (Aghvinian et al., 2021). Patients with a diverse background may not be familiar with "best performance" or speed tests (Ardila, 2007). They may experience distress when the examiner points out errors or stops the test after the pre-set time limit has been passed (Aghvinian et al., 2021). In diverse populations, it is therefore even more important than usual to consider the patient's understanding of neuropsychological testing in general and of each individual test specifically, and to provide additional explanations if needed (Aghvinian et al., 2021; Franzen on behalf of the European Consortium on Cross-Cultural Neuropsychology, 2021). Additional practice items may need to be provided (Rock & Price, 2019).

Given the strong influence of diversity-related factors such as education, culture, and language on the performance on traditional neuropsychological tests, more suitable alternative tests are needed to assess culturally, educationally, and linguistically diverse populations. However, there currently is a lack of appropriate cognitive tests and normative data (Franzen, van den Berg, et al., 2020; Franzen et al., 2021; Nielsen et al., 2011; Ponsford, 2017). Several European initiatives have therefore unfolded in parallel over the past few years, including the development and validation of the European Cross-Cultural Neuropsychological Test Battery (CNTB; Nielsen et al., 2018) and the Cross-Cultural Dementia Screening (CCD; Goudsmit et al., 2017), as well as European validation studies of the Rowland Universal Dementia Assessment Scale (RUDAS; Storey et al., 2004; see also Nielsen & Jørgensen, 2020). The RUDAS and CCD are appropriate for screening purposes, whereas the CNTB thus far is the only large test battery available for diverse populations in Europe that can provide a more in-depth analysis of individual cognitive domains. Although the CNTB includes several promising tests, it also contains a number of tests that are less suitable for patients who are illiterate, because they require skills learned in the educational system, such as graphomotor figure copy tests and the Color Trails Test (Nielsen, 2019). Last, some cognitive domains, such as language (naming) and working memory, as well as performance validity are not or insufficiently covered by the CNTB. Moreover, the validity and feasibility of this battery has not been examined in diverse populations in the Netherlands.

Given the expected rise in the number of individuals from culturally, educationally, and linguistically diverse backgrounds visiting memory clinics, there is an urgent need for a cognitive test battery that is suitable for this diverse population, taking into account individuals' limited experience with being tested. In this study, we used a two-step process to address this need. First, we aimed to develop a new neuropsychological test protocol—consisting of promising existing tests and/or newly developed tests—by examining the scientific literature and consulting with experts. It is vital that such a test protocol has demonstrated feasibility, e.g. in terms of administration time, user friendliness, and completion rates, and that the test results reflect a patient's optimal performance. To this end, potential secondary influences on neuropsychological test performance that could complicate the assessment should also be taken into consideration, such as suboptimal effort/malingering, depression, (moderate to severe) anxiety, fatigue, pain, and motor and/or sensory impairments (Arnett, 2013). The second aim of this study was therefore to examine the feasibility of this neuropsychological test protocol in a culturally, educationally, and linguistically diverse memory clinic setting.

Part 1: development of the TULIPA test battery and optimization of procedures

Methods

In the following paragraphs we first describe the development of the TULIPA test battery (Towards a Universal Language: Intervention and Psychodiagnostic Assessment). This multi-stage process included a literature review, consultation with European

experts, and focus groups with Dutch specialists in cross-cultural neuropsychology. Second, we present the tests included in the battery. Third, we describe the steps towards implementation in clinical practice, including consultation with individuals from a diverse background and streamlining of interpreter-mediated assessment.

To determine which tests should be included in the neuropsychological test battery, we consulted the relevant international literature through a systematic review (Franzen, van den Berg, et al., 2020). In addition, we carried out a Delphi expert study across European Union-15 countries to determine which tests/practices are currently used in cross-cultural neuropsychological assessment in countries with similar populations (published in full elsewhere, for more detail see Franzen et al., 2021).

Next, we organized three focus groups with neuropsychologists experienced in assessing diverse patient populations (neuropsychologists present per focus group: 6–9, total number of neuropsychologists that contributed to these focus groups: 12). The goal was to identify relevant barriers and facilitators and select appropriate tests. The participants were recruited from academic and non-academic memory clinics in the three most populous and diverse cities in the Netherlands, as well as from two organizations specializing in research or care for older diverse populations (an organization for intercultural psychiatry and an organization promoting cognitive health in underrepresented populations). One participant was recruited in a more rural area in the Netherlands. All participants were invited by email and received financial compensation paid to their organization for participation and travel expenses. The participants were predominantly female (92%), reflecting the underrepresentation of men in the workforce of psychologists in the Netherlands. The face-to-face focus groups lasted 2 hours on average and included two short breaks. In the first focus group, participants were asked through open-ended questions (1) which barriers they experienced in the neuropsychological assessment of diverse individuals; (2) which aspects facilitated these assessments; and (3) where they saw areas of need. The focus group leader facilitated the discussion of each of these topics and subsequently ensured all participants' perspectives were identified and clarified where needed. Group discussion was encouraged. In the second focus group, the neuropsychologists were first presented with the available international instruments; they then (1) discussed which of the available instruments they considered suitable candidates for the test battery and (2) identified the need for the development of new tests and/or questionnaires. In the third focus group, the participants finalized their selection for the test battery. All session were videorecorded with consent of the participants and were transcribed verbatim. The focus groups were analyzed and coded by the first author through a combined deductive/inductive approach to thematic analysis (Roberts et al., 2019).

Subsequently, we organized a two-hour consultation with ten community-dwelling individuals from culturally and educationally diverse backgrounds recruited by community liaisons through a local network of diverse, faith-based community organizations (including both male and female participants). Some participants had prior experience with dementia in their personal network or through their occupation; one participant had previously been cognitively assessed. Given the potential mistrust in research (Gilmore-Bykovskyi et al., 2019), we prioritized trust-building in this meeting, and therefore decided to not record the personal information of the participants nor did we make any formal audio or video recordings during the meeting. All information

provided by participants was recorded through extensive note-taking and analyzed by one coder using a general inductive approach. The community liaison was present during the entire meeting. The aims of the consultation were (1) identifying how diverse individuals perceive the TULIPA tests, stimuli, and procedure and (2) determining which additional instructions are needed to use the tests in clinical practice. In three subgroups, the participants were first asked the following questions: "What do you think of these tests/what are your impressions of these tests?" and "What do you think we want to measure with these tests?" Afterwards, the purpose and instructions of the test were explained in Dutch by the discussion leader (SF, native Dutch background), while two bilingual, bicultural research assistants aided in case of a language barrier. Participants were then invited to share their opinions, thoughts, and emotions about the tests and assessment in general. All participants received a gift certificate as a token of appreciation for participating and received the summary of the meeting's findings by email from the community liaison.

Last, to ensure smooth implementation, we attempted to standardize interpreter-mediated assessment through a group discussion with a team of bilingual, bicultural interpreters with a background in medicine, (neuro)psychology, or paramedical disciplines. Furthermore, a manual for neuropsychological assessment with the TULIPA battery was written, which included guidelines for history taking, as well as administration, scoring, and interpretation of tests.

Results

Development of the TULIPA test battery

In the review and Delphi study, we found that memory was relatively well-studied in culturally and educationally diverse populations, whereas suitable tests for some other cognitive domains, such as language (e.g. naming) were urgently needed (for more detail, see Franzen et al., 2021; Franzen, van den Berg, et al., 2020). The available tests and norms identified in these studies were presented in the subsequent focus groups with neuropsychologists. The experts in the Delphi study strongly recommended assessment using formal interpreters where possible and also provided recommendations how to carry out such an assessment (see also *Implementation of the test battery*).

In the focus groups with neuropsychologists, several barriers to cognitive testing were identified. These barriers largely reflect those presented in the international literature, such as issues with working with interpreters, a lack of available tests and norms, specific test elements that are less suitable to culturally and educationally diverse populations (e.g. black-and-white line drawings, graphomotor tests), and challenges in determining whether a patient performs optimally. It was agreed in the second focus group that the battery at a minimum needed to cover the cognitive domains of memory, language, visuoconstruction, mental speed, attention, working memory, and executive functioning. These domains were selected because they are often impaired in individuals with cognitive impairment due to neurodegenerative disease. These tests should make it possible to determine a profile of impaired and intact cognitive functions that can aid in the differential diagnosis. In the third focus group, the neuropsychologists reached a consensus on tests to be included in the

TULIPA battery. The test battery consisted of several core tests already validated in culturally, educationally, and linguistically diverse populations in the Netherlands and a number of supplementary tests from the international literature. The neuropsychologists agreed that two new tests should be developed to cover aspects that could not be measured in a valid and reliable way with existing tests. First, the focus group highlighted the need for a new naming test—in line with findings of the Delphi study. Second, the neuropsychologists in the focus group identified a need for a test to examine academic achievement/quality of education by means of a literacy screening test; for example, one participant suggested the development of a literacy screening tool based on the Adult Literacy Supplemental Assessment of the National Assessment of Adult Literacy (National Center for Education Statistics, 2003).

The TULIPA test battery

The neuropsychological tests that were ultimately included in the TULIPA battery are displayed in Table 1. The core battery, administered as the "gold standard" to all patients, consisted of the RUDAS, which was validated in the Netherlands by Goudsmit et al. (2018), the CCD (Goudsmit et al., 2017), the modified Visual Association Test (Franzen et al., 2019), and semantic verbal fluency (animals and foods). The CCD consists of three tests, the Objects test (subtest A and B) for memory, as well as the Sun-Moon test (subtest A and B) and the Dots test (subtest A and B) measuring mental speed/attention and executive functioning. The modified Visual Association Test is a visual-associative memory test validated in diverse populations in the Netherlands which uses colored photographs as stimuli, instead of the black-and-white line drawings in the original test (Lindeboom & Schmand, 2003). The supplementary battery contained two tests of visuospatial functioning: The Clock Reading Test from the CNTB (Nielsen et al., 2018) and the Stick Design Test (Baiyewu et al., 2005); the latter was selected as it does not require any graphomotor drawing skills. In the domains of attention/mental speed/executive functioning, we included the Five Digit Test (Sedó, 2004) and a Turkish version of the Stroop test (Karakas et al., 1999), to be administered only to Turkish-speaking patients who are literate. The Corsi Block Tapping Test (Corsi, 1972) was added as a measure of (visual) working memory as the more commonly used digit span is heavily influenced by language of administration (Franzen, van den Berg, et al., 2020). The supplementary battery contained one additional memory test, the Recall of Pictures Test of the CNTB (Nielsen et al., 2018). The Coin-in-the-Hand Test (Kapur, 1994) was used to detect suboptimal performance. The Naming Assessment in Multicultural Europe (NAME: Franzen et al., 2022) was developed and validated over the course of 2018–2019. It is a 60-item naming test using colored photographs as stimuli as opposed to black-and-white line drawings. A second instrument that was developed was a literacy screening tool to capture educational quality/academic achievement (unpublished); an experimental version was developed for Dutch, Turkish, and Moroccan-Arabic.

In addition to neuropsychological tests, several questionnaires were used such as the short Informant Questionnaire on Cognitive Decline (IQCODE; Goudsmit et al.,

Table 1. Neuropsychological and achievement tests in the TULIPA test battery.

Core battery	Cognitive domain	Description	Approximate administration time
RUDAS	General cognitive functioning	Screening test measuring global cognition by assessing memory, body orientation, praxis, drawing, judgment, and language	20 min
CCD Objects test (subtest A & B)	Memory	Immediate and delayed recognition of photographs presented in a grid alongside distractors	30 min (full CCD)
CCD Sun Moon subtest A	Mental speed & attention	Sequentially name suns and moons printed on paper; score based on time corrected for errors	
CCD Sun Moon subtest B	Executive functioning	Sequentially name antonym for printed suns and moons (i.e. sun = moon); score based on time corrected for errors	
CCD Dots subtest A	Mental speed & attention	Connect rectangles containing an increasing number of dots in increasing order with a pencil	
CCD Dots subtest B	Executive functioning	Similar to A, but patients need to switch between black and white items	
Animal and foods fluency	Language & executive functioning	Patients name as many animals and edible things in 1 minute	2 min
Modified Visual Association Test	Memory	Patients are required to remember visually presented stimuli that are not usually associated with the stimulus they are paired with	5–10 min
Supplementary tests			
Literacy screener total	Reading/writing	Screening tool that examines academic achievement through the assessment of phonological awareness, receptive language abilities, and language production	5–10 min
Five Digit Test Reading	Mental speed & attention	Requires patients to name a series of 50 printed digits (digits between 1–5)	10 min (full FDT)
Five Digit Test Counting	Mental speed & attention	Requires patients to count a series of 50 printed asterisks (number of asterisks per square between 1–5)	
Five Digit Test Choosing	Executive functioning	Requires patients to count an incongruent number of digits, e.g. three fives	
Five Digit Test Shifting	Executive functioning	Similar to Choosing, but switch to reading instead of counting if item is highlighted	
Stroop Cards 1, 3, 4 (Turkish)	Mental speed & attention	Read colors printed in black (1); name colored dots (3); name color of neutral, non-color words printed in color (4)	10 min (full Stroop)
Stroop Cards 2, 5 (Turkish)	Executive functioning	Read words printed in an incongruent color (2); name the color of incongruent color-words (5)	
Recall of Pictures Test—naming	Language	Naming of 10 items presented as colored line drawings	1 min
Recall of Pictures Test—memory	Memory	Incidental, immediate, and delayed recall, as well as recognition, of 10 colored line drawings	15 min (incl. delay)
Corsi Block Tapping Test	Working memory	Reproduce block-tapping sequences of increasing length; includes a forward and backward condition	5–10 min
Coin in the Hand Test	Performance validity	10-item performance validity test in which patient has to remember in which hand a coin is placed while performing a mock distraction task	5 min
Stick Design Test	Visuospatial & construction	Copying of a matchstick configuration with four matches	5 min
Clock Reading Test	Visuospatial & construction	Read the time on 12 clock faces without numbers	5 min
Naming Assessment in Multicultural Europe	Language	Naming of 60 colored photographs	5–20 min

2021; Jorm & Jacomb, 1989) and adapted versions of the Geriatric Depression Scale (GDS; Uysal-Bozkir et al., 2016; Yesavage et al., 1982). Acculturation was measured with a shortened, adapted Short Acculturation Scale for Hispanics (SASH; Marin et al., 1987).

Implementation of the test battery and optimization of test procedures

In the consultation with individuals with a diverse background, we found that the goal of each of the individual tests and the relationship with everyday cognitive functioning was often unclear to the participants—in line with the findings by Aghvinian et al. (2021). In some cases, participants assumed aspects had meaning beyond the original intention of the test; for example, one participant thought that the Stick Design Test was meant to induce a perceptual illusion (see Supplementary Table 1 for example quotes and how these findings were subsequently used). Participants provided several comments on the large number of items or length of the tests. Furthermore, they reported their first (emotional) reactions to the stimuli, such as feeling nervous or overwhelmed, particularly when faced with time pressure. After having been explained what the tests were supposed to measure, the participants provided feedback on the best ways to instruct patients. Participants recommended neuropsychologists to provide more extensive information about the assessment before the actual appointment, or even to invite the caregiver for a separate session before the assessment to explain the procedure. The participants also provided advice how to ensure poor performance was indicative of cognitive impairment and not caused by other factors. For example, they recommended neuropsychologist to verify whether patients had been able to tell the time before administering the Clock Reading Test.

These recommendations provided by the individuals with diverse backgrounds on the instructions during the consultation session were incorporated into the manual. The manual also included the recommendations for interpreter-mediated assessment described in more detail in the Delphi study (see Franzen et al., 2021). Two follow-up meetings with the participating neuropsychologists were organized after data collection had started to share experiences and ensure test administration was comparable across centers.

In the discussion with interpreters, some aspects of interpreting during neuropsychological assessment were identified as problematic; for example, it proved particularly challenging to translate questions relating to sustained and divided attention, as well as mental speed—these terms often had to be explained using examples and longer sentences because adequate terminology capturing these terms was not available in all languages. In addition, regional variations/dialects made interpretation challenging for some populations; for example, four interpreters speaking Tamazight, a Moroccan language family, often used regionally appropriate terminology that was unfamiliar to the interpreters from the other regions. Similarly, one of the neuropsychologist who participated in the focus group was made aware by a certified interpreter that it was impossible to translate the patient's words literally because he/she was speaking in metaphors, the meaning of which would be lost if translated literally.

Part 2: feasibility of the TULIPA battery

Methods

Participants

For the feasibility study, we enrolled 345 patients at four Dutch memory clinics specializing in the assessment of culturally, educationally, and linguistically diverse populations: The Erasmus MC University Medical Center in Rotterdam (hereafter: "Rotterdam 1"), the Maasstad Ziekenhuis in Rotterdam ("Rotterdam 2"), the Haaglanden Medical Center in The Hague, and Medisch Spectrum Twente in Enschede (see Table 2). In these multicultural memory clinics, services are tailored specifically to diverse populations; for example, staff members (1) provide patients with culturally and linguistically appropriate information about cognitive impairment and subsequent cognitive assessment, (2) often use tools such as a cultural (formulation) interview and/or "teach-back" methods (Kripalani et al., 2008) to facilitate communication, and/or (3) may collaborate intensively with culture-sensitive care providers to offer suitable care after a diagnosis.

The Rotterdam 1, the Hague, and Enschede cohorts were enrolled consecutively, whereas the Rotterdam 2 cohort consisted of a subset of patients referred specifically for more extensive neuropsychological assessment after completing initial screening tests from the core battery (e.g. RUDAS). Patients were enrolled between January 2019 and May 2021. The NAME and literacy screener were introduced to the battery after their development was complete (October 2019). The majority of patients were immigrants from Turkey (n = 115, 33%), Morocco (n = 67, 19%), and Suriname (n = 57, 17%); all included Cape Verdean patients (n = 16, 5%) lived in Rotterdam, while Syrian patients (often with a Syriac-Orthodox background) were often seen in Enschede (n = 13 out of 16, 5%). In total, we included patients originating from 37 countries.

Procedure

All patients underwent neuropsychological testing with the TULIPA test battery as part of their routine clinical visit. The maximum duration of the neuropsychological assessment including history taking was 180 minutes. Neuropsychologists were free to select tests from the list of supplementary tests after completing the core battery. All neuropsychologist received the TULIPA test manual including scoring and administration guidelines. The two Rotterdam sites used formal interpreters for their assessments, while no formal interpreters were generally used in Enschede or the Hague, where assessments were mostly conducted with an informal interpreter or in Dutch (e.g. for Surinamese patients proficient in Dutch). The formal interpreters were either hired from a nationwide interpreter agency or hired and trained directly by one of the participating multicultural memory clinics. In all centers, the diagnostic workup consisted of a comprehensive clinical evaluation, with history taking by a geriatrician or neurologist, a neuropsychological assessment with the TULIPA test battery, and standard laboratory screening; structural brain imaging was performed in a subset of patients (n = 234, 67%). Clinical diagnoses were determined in multidisciplinary consensus meetings with (at a minimum) a neuropsychologist and geriatrician or neurologist present, based on all the available clinical information and using the diagnostic research criteria for subjective cognitive impairment (Jessen et al., 2020), mild cognitive

Table 2. Demographic characteristics of the full sample[a].

	Rotterdam 1 (n = 177)	Rotterdam 2 (n = 22)	Enschede (n = 48)	The Hague (n = 98)
Age	66.6 (12.6)	70.3 (9.7)	69.7 (9.3)	74.0 (7.6)
Education n(%):	42 (24%)	0 (0%)	16 (33%)	29 (30%)
• Zero years of education				
• >0 but < completed primary education	30 (17%)	6 (27%)	7 (15%)	22 (22%)
• Primary education	33 (19%)	5 (23%)	13 (27%)	26 (27%)
• Higher than primary education	70 (41%)	11 (50%)	12 (25%)	21 (21%)
• Sex (n(%) male)	83 (47%)	12 (55%)	20 (42%)	38 (39%)
• Years in the Netherlands	37.6 (14.0)	28.6 (16.3)	34.2 (15.3)	39.7 (12.8)
• RUDAS[b]	21.8 (5.1; n = 148)	21.2 (5.8; n = 17)	20.2 (6.2; n = 41)	19.2 (6.1; n = 75)
• Number of core tests administered[c]	6.0 (5.0–7.0)	7.0 (6.0–7.0)	7.0 (5.0–7.0)	6.0 (3.0–7.0)
• Supplementary tests administered[c]	2.0 (1.0–4.0)	2.0 (1.8–3.0)	1.0 (1.0–2.0)	0.0 (0.0–1.0)
Interpreters				
• Formal interpreter present (%)	148 (84%)	22 (100%)	0 (0%)	6 (6%)
• Informal interpreter present (%)	7 (4%)	0 (0%)	42 (88%)	52 (55%)
• No interpreter present (%)	22 (12%)	0 (0%)	6 (12%)	40 (41%)
Diagnosis n(%)				
• Subjective cognitive impairment	36 (20%)	5 (23%)	7 (15%)	12 (12%)
• Mild cognitive impairment	21 (12%)	6 (28%)	3 (6%)	14 (14%)
• Dementia	44 (25%)	7 (32%)	12 (25%)	49 (50%)
• Psychiatric disorder	40 (23%)	3 (14%)	16 (33%)	9 (9%)
• Cognitive disorder due to other known medical condition	11 (6%)	0 (0%)	3 (6%)	2 (2%)
• Could not be determined	23 (14%)	0 (0%)	7 (15%)	12 (12%)

RUDAS = Rowland Universal Dementia Assessment Scale.
Values are displayed as mean (standard deviation) unless otherwise specified.
[a]A number of cases are missing for education and years in the Netherlands because patients were asked but were unable to report it.
[b]The maximum score for the RUDAS is 30, with a cut-off score of <22 in culturally, educationally, and linguistically diverse populations in the Netherlands.
[c]Median (first quartile-third quartile).

impairment (Albert et al., 2011), and dementia subtypes (e.g. McKhann et al., 2011; Román et al., 1993), and the DSM-V for primary psychiatric disorders (American Psychiatric Association, 2013).

Feasibility was operationalized in two ways. First, we recorded the number of times a test was administered and the number of times the test was completed. Second, we collected data on the presence of complicating factors (or "secondary influences," Arnett, 2013) in neuropsychological assessment; these included suboptimal effort, depressive symptoms, anxiety, pain, other somatic complaints that may interfere with testing, fatigue, motor impairments, and sensory impairments. We collected this information retrospectively by analyzing the neuropsychological reports (see Supplementary Table 2 for example codes). We included both complicating factors that were self-reported during the assessment and subsequently written down in the patient records, as well as observations made by the neuropsychologist. The coding and analysis were done by one author (SF); cases of doubt were resolved in consensus with a second author (JMP). For the analyses of the complicating factors, we only had data available from the Rotterdam 1, the Hague, and Enschede cohorts, as the complete patient records including the observations were not available for the Rotterdam 2 site (n = 22) due to local privacy regulations. Ethical approval for the study was obtained from the institutional review board of the Erasmus Medical Center (MEC-2019-0036); additionally, local approval was obtained from the (scientific) boards of all participating centers. All procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Statistical analysis

Statistical analyses were carried out using IBM SPSS Statistics version 25. Descriptive analyses were used to examine administration and completion rates. We used Spearman correlations to examine the association between administration rates, demographic characteristics (sex, age, education, number of years in The Netherlands), and indices of disease severity (RUDAS, short IQCODE). We corrected for multiple testing using False Discovery Rates (FDR) based on Benjamini-Hochberg adjusted *p*-values. To investigate the influence of site and interpreter presence, we compared the number of tests administered at each study site using a Kruskal-Wallis test and compared administration rates with and without a formal interpreter present using a Mann-Whitney U test. Second, we calculated how often factors complicating the neuropsychological assessment were present, and subsequently examined the association between the number of complicating factors, the administration rate, demographic characteristic, and disease severity (RUDAS, IQCODE) with Spearman correlations corrected for FDR. In addition, we quantified depressive symptoms using the GDS and suboptimal performance using the Coin-in-the-Hand test.

Results

Feasibility of the TULIPA neuropsychological test battery

Table 3 shows the administration and completion rates of the TULIPA battery tests (see Supplementary Table 3 for the rates by study site). The median number of core

Table 3. Number of times TULIPA tests were administered and subsequently not completed.

	Administered (of n = 345)	Not completed (%)
Core battery		
RUDAS	290 (84%)	6 (2%)
CCD Objects test A	298 (86%)	7 (2%)
CCD Objects test B	284 (82%)	10 (4%)
CCD Sun Moon test A	290 (84%)	8 (3%)
CCD Sun Moon test B	281 (81%)	34 (12%)
CCD Dots test A	275 (80%)	49 (18%)
CCD Dots test B	230 (67%)	97 (42%)
Animal fluency	295 (86%)	2 (1%)
Food fluency; supermarket fluency[a]	186; 35 (54%; 10%)	0 (0%)
Modified Visual Association Test (short or long)	227 (66%)	3 (1%)
Supplementary tests		
Literacy screener total	71 (21%)	1 (1%)
Five Digit Test Reading and Counting	51; 51 (15%)	0 (0%)
Five Digit Test Choosing and Shifting	50; 39 (14%; 11%)	1; 4 (2%–10%)
Turkish Stroop Cards 1; 3; 4 (Attention/speed)	17; 17; 17 (5%)	0; 0; 1 (0%–6%)
Turkish Stroop Cards 2; 5 (Executive)	17; 16 (5%)	0; 0 (0%)
Recall of Pictures Test—naming subtest	90 (26%)	0 (0%)
Recall of Pictures Test—memory subtests	86 (25%)	0 (0%)
Corsi Block Tapping Test	66 (19%)	0 (0%)
Coin in the Hand Test	112 (32%)	3 (3%)
Stick Design Test	72 (21%)	2 (1%)
Clock Reading Test	85 (25%)	7 (8%)
Naming Assessment in Multicultural Europe	95 (28%)	3 (3%)

RUDAS = Rowland Universal Dementia Assessment Scale, CCD = Cross Cultural Dementia Screening.
[a]Supermarket fluency, which is traditionally recommended in the Netherlands for the assessment of low educated individuals, was administered in The Hague instead of food fluency in some cases.

tests administered across the sample was six out of seven (inter quartile range (IQR): 5 to 7). This number differed significantly by study site (H (3) = 13.25, p = .004; see Table 1 for medians) and depending on whether a formal interpreter was present (U = 18,257.50, p < .001). Most tests, including the CCD Objects test, RUDAS, and animal fluency showed high administration and completion rates. The CCD Dots subtest B was administered less often than the other tests of the CCD; this can partly be explained by the number of individuals who could not complete subtest A (and as a result were not administered part B). The CCD Dots subtest B frequently was not completed (42%). It was sometimes observed that patients counted the number of dots presented in each of the items. Many of the patients that completed the Dots subtest B needed one or more hints (e.g. ≥1 hint in 78%; ≥2 in 63%, and ≥5 in 28%).

The supplementary tests were used less often than the core battery (Table 3, bottom half); a median of one test from the list of supplementary tests was administered per patient (IQR: 0 to 3/9 tests). The number of administered supplementary tests differed by study site (H (3) = 83.79, p < .001; see Table 1 for medians) and depending on whether a formal interpreter was present (U = 21,949.50, p < .001). A subset of patients were administered a more substantial number of supplementary tests (e.g. ≥5/9 in 13%). Supplementary tests showed high completion rates (between 90%–100%). A test that was administered relatively infrequently was the Turkish version of the Stroop test (n = 17), which was likely due to the limited number of literate Turkish patients in the sample (n = 61; assessment rate in this group 28%). A larger number of tests that was administered was associated with better overall cognitive performance as measured by the RUDAS (ρ=.33, adjusted p < .001). We did not find a significant

Table 4. Presence of complicating factors in the assessments.

Complicating factor	Measure	Times observed (%)
Suboptimal effort/motivation	Suboptimal effort observed	92/314 (29%)
	Suboptimal effort on Coin-in-the-Hand	54/109 (50%)
	2–4 errors	31/109 (28%)
	≥5 errors (chance level and below)	23/109 (21%)
	Patient refuses to continue with testing	28/317 (9%)
Depression	Depressive symptoms observed during testing	98/262 (37%)
	Depression on GDS-15 (score ≥ 6)	138/243 (57%)
Anxiety	Anxiety observed/reported during testing	44/260 (17%)
Fatigue	Fatigue observed/reported during testing	103/317 (32%)
Pain	Pain observed/reported during testing	38/317 (12%)
	Other physical symptoms that hinder testing	19/317 (6%)
Motor impairment	Motor impairments that hinder testing	21/317 (7%)
Sensory impairment	Sensory impairments that hinder testing	41/317 (11%)

correlation with any other patient characteristics (i.e. sex, age, education level, years in the Netherlands, short IQCODE [n = 96]). A total of 28 patients (9%) at some point refused to continue with testing; a median of 5.5 tests (IQR: 3.0 to 7.8 tests) of the core and supplementary batteries had been administered before testing stopped.

Presence of complicating factors and relationship with demographics, disease severity, and number of tests administered

Table 4 shows the frequency of complicating factors observed during the neuropsychological assessment (coded according to the system in Supplementary Table 2). Depressive symptoms (37% of the sample), suboptimal effort (29%), and fatigue (32%) were observed frequently. The number of patients who showed symptoms of depression was even higher when formally measured with the GDS (57%). In cases where neuropsychologists decided to formally test effort using the Coin-in-the-Hand test (32% of all cases), close to half of the tests were indicative of possible suboptimal performance. A larger number of complicating factors was present in patients who were younger (ρ=-.23, adjusted p = .001) and female (ρ=.21, adjusted p = .003). We did not find any significant correlations with other patient characteristics (education level, years in the Netherlands, RUDAS score, IQCODE score). Although complicating factors were observed to some degree in patients with all types of diagnoses, they were observed slightly more often in patients who were ultimately diagnosed with psychiatric illness (e.g. depression, post-traumatic stress disorder; see Supplementary Figure 1 for a plot showing the distribution of complicating factors across diagnostic groups). There was no significant correlation between the number of administered tests and the number of complicating factors present during neuropsychological testing.

Discussion

Few neuropsychological tests are available that are suitable for culturally, linguistically, and educationally diverse populations unfamiliar with undergoing formal tests. Our aims were therefore (1) to compose a test battery specifically for such a population, and (2) to examine the feasibility of this battery in a multicultural memory clinic

setting. The TULIPA test battery was composed after a literature review, consultation with European experts, and focus groups, and the implementation phase included consultations with individuals from diverse backgrounds and streamlining of interpreter-mediated assessment. The newly composed TULIPA test battery included tests such as the CCD, RUDAS, mVAT, and several subtests of the CNTB, as well as newly developed tests to assess language (NAME) and a literacy screener (as an academic achievement test). Our results indicated that, with the exception of the Dots subtest B of the CCD, administration and completion rates of the core test protocol were high, indicating that the core battery is feasible. A limited number of supplementary tests were administered per patient, but when used, completion rates were similarly high. The number of tests that could be administered was associated with disease severity as measured by the RUDAS, but not with other patient characteristics. Factors complicating the neuropsychological assessment that may impact feasibility were observed frequently, in particular suboptimal effort/motivation, fatigue, and depressive symptoms. Last, our consultations with interpreters highlighted that neuropsychologists should be aware that interpreters may (need to) deviate from translating literally during interpreter-mediated cross-cultural assessments and that communication difficulties may arise if interpreters and patients speak (slightly) different dialects.

Unsurprisingly, we found that fewer TULIPA tests were administered in patients with more objective cognitive impairment. The lack of association with any other patient characteristics, such as age or number of years living in the Netherlands, makes this a promising battery for the assessment of diverse populations. Although the current study does not allow for a formal comparison of the feasibility of different approaches to the assessment of diverse populations—e.g. the use of the TULIPA battery versus simple translations of traditional Dutch tests—it seems likely that the TULIPA battery represents an improvement in feasibility, given the issues identified in past research in the assessment of diverse populations with traditional test batteries in memory clinics (Nielsen et al., 2011). The TULIPA battery incorporates some of the psychometrically sound elements of the CNTB (Nielsen et al., 2018)—the only battery available for European diverse populations thus far (Franzen et al., 2021)—while also tailoring to very low-educated individuals and covering several (additional) cognitive functions (naming, non-graphomotor visuoconstruction, working memory), performance validity, and quality of education. Before assessment with the TULIPA battery can become recommended practice, however, diagnostic accuracy studies should be carried out to determine the validity of the individual tests in the TULIPA supplementary test battery. For example, although the first international diagnostic accuracy studies of the Stick Design Test were promising (e.g. Baiyewu et al., 2005; de Paula et al., 2013), diagnostic accuracy was poor in a later study (Ortega et al., 2021). Diagnostic accuracy studies in patient populations with different diagnoses may also result in clinical guidelines to decide which tests to prioritize for which patient. In addition to these diagnostic accuracy studies, the knowledge and skills relevant for cross-cultural neuropsychological assessment in Europe as identified in the Delphi study (Franzen et al., 2021) should be transformed into guidelines to help neuropsychologists determine whether they possess the necessary competencies to assess patient with diverse backgrounds in Europe.

Factors complicating the neuropsychological assessment, in particular depressive symptoms, fatigue, and suboptimal effort (likely often related to fatigue), occurred in between a quarter and half of all patients. The number of complicating factors observed was not associated with the number of tests that was administered; that is, patients with pain, fatigue, or depressive symptoms were not administered fewer tests because of these symptoms. Interestingly, these complicating factors were observed more frequently in women and in younger individuals. This might be explained by the large number of complicating factors in patients with psychiatric diagnoses, who in this study were often relatively young and in whom symptoms of depression or anxiety (by definition) are common. In clinical practice, factors such as fatigue should be monitored during the assessment, e.g. by frequently asking the patient if they are tired and/or need a break. Although studies investigating the influence of fatigue on cognitive test performance show that fatigue may not universally impair performance on objective measures of cognitive functioning (e.g. Johnson et al., 1997), it may impact the willingness to undergo (additional) tests and the overall experience of neuropsychological testing.

It is worthwhile to note that no studies have been carried out comparing performance on the Coin-in-the-Hand test between individuals from culturally, educationally, and linguistically diverse backgrounds with objective memory impairment and individuals with feigned memory problems. It is widely established that persons with dementia in particular can fail performance validity tests due to objective cognitive impairment (Dean et al., 2009); the finding that a large number of individuals obtained a score below the cutoff on the Coin-in-the-Hand test should therefore be interpreted with caution. Few if any alternatives to the Coin-in-the-Hand test are currently available to detect suboptimal performance in the diverse populations assessed in European memory clinics. Studies suggest that false-positive results on performance validity tests may occur more frequently in diverse populations when traditional tests such as the Test of Memory Malingering or Rey-15 are used (Nijdam-Jones et al., 2017; Nijdam-Jones & Rosenfeld, 2017). For example, one quarter of the healthy adults tested with the Test of Memory Malingering in Paraguay were misclassified as displaying insufficient effort (Nijdam-Jones et al., 2017). Although the Amsterdam Short-Term Memory test (Nijdam-Jones & Rosenfeld, 2017) showed more promising sensitivity and specificity, this test cannot be administered to low-educated populations because it requires participants to read and calculate. Last, it is challenging to derive embedded measures of performance validity from TULIPA test scores, such as from the animal fluency score. Although such measures are increasingly recommended (e.g. Sugarman & Axelrod, 2015), separate cut-offs would likely be required for each language given the substantial influence of language on the number of words generated during animal fluency (Kempler et al., 1998).

Some limitations should be acknowledged. First, feasibility can be investigated in a number of ways, and only a select number of indicators were investigated here. Previous feasibility studies in neuropsychology have also looked into (1) experiences of the patients undergoing the tests (e.g. Hildebrand et al., 2004; Spreij et al., 2020), (2) how often participants required breaks (Spreij et al., 2020), and (3) test-specific feasibility aspects (e.g. visibility of stimuli). A study of these other indicators of feasibility can provide an even more in-depth perspective on feasibility of the TULIPA

battery. Second, some centers administered "traditional" neuropsychological tests that are not part of the TULIPA protocol to some of their patients, instead of the supplementary TULIPA subtests; for example, several higher educated Surinamese individuals proficient in Dutch underwent tests (e.g. a Dutch auditory verbal learning test) not included in the sum score for the total number of tests administered. Therefore, it may have been possible to administer more TULIPA tests had the neuropsychologist selected those. Other site-specific factors, such as the type of patient population and referrals, as well as the availability and use of formal interpreter services may also have influenced the number of tests administered at each site. Third, the feasibility study was carried out in a clinical setting, in which the clinicians were allowed to choose how many and which tests from the list of supplementary tests they felt necessary and worthwhile to administer. This leads to a selection bias—that is, we cannot ascertain the feasibility of the tests in individuals in which they were not administered. Last, all coding was performed by one person, precluding analyses from multiple perspectives and/or analyses of intercoder reliability/consistency.

This study has several strengths. First, the test protocol was developed based on a thorough review of the available international tests and practices and was decided upon in consensus with neuropsychologists who often assess culturally, educationally, and linguistically diverse populations. Second, individuals from diverse backgrounds were actively consulted in the development stages of the battery and their feedback was incorporated in the implementation phase. Third, the data were collected in multicultural memory clinics that have ample experience in assessing culturally, educationally, and linguistically diverse populations. Last, we were able to include a large sample of patients who were extremely diverse in terms of country of origin, language, and years of education, which is reflective of the remarkable diversity in Europe itself.

This study provides several points of departure for future research, in addition to the need for diagnostic accuracy studies. First, future studies might examine ways to improve the feasibility of neuropsychological testing. Both the international literature and the individuals from diverse backgrounds that were consulted stress the importance of providing patients from culturally, educationally, and linguistically diverse backgrounds and their caregivers with sufficient information about the purpose of, need for, and rationale behind the assessment and the individual tests (Aghvinian et al., 2021; Franzen on behalf of the European Consortium on Cross-Cultural Neuropsychology, 2021; Rock & Price, 2019). Although this need is in no way unique to diverse populations (see e.g. Gruters et al., 2021), it may be especially important in this population given the limited experience with formal testing that characterizes (low educated) diverse populations. Extra information that can be provided may include, for example, explanations before the assessment how seemingly abstract tests—such as the Five Digit Test and Sun-Moon test—are used to make inferences about a patient's everyday functioning in the domains of attention and executive functioning. In addition, it may be necessary to explain how findings on the neuropsychological assessment reflect changes in different regions of the brain and how the assessment, combined with neuroimaging biomarkers, can contribute to the overall diagnosis. In some cases, providing explicit examples of impaired performance, such as hemispatial neglect, during testing may help patients

understand why they need to undergo specific tests. Given the number of individuals who at some point refused to continue with testing in our sample (slightly under one in ten), such explanations may encourage patients to deliver an optimal performance. A second approach to make the TULIPA battery more feasible is by shortening the individual tests, such as by administering only half of the items of the Five Digit Test or by eliminating less sensitive items of the NAME based on an item analysis. Third, future research may investigate whether current procedures to provide feedback on suboptimal performance such as those by Carone et al. (2010) are culturally appropriate and effective in diverse populations. Fourth, both the TULIPA battery and CNTB rely mostly on visually presented stimuli; this may pose problems in the assessment of patients with visual impairment, as well as in patients without visual impairment by resulting in interference from one visual test to the other. Language-specific verbal tests are likely needed and should be examined in future studies. Last, some cognitive domains that are not routinely assessed in all patients in every memory clinic, such as praxis or social cognition, were not included in the battery. It remains to be seen whether it is possible to develop suitable, cross-cultural tests for social cognition, a cognitive function that is substantially influenced by culture (Quesque et al., 2020).

In conclusion, the TULIPA battery is a promising new battery for neuropsychological assessment of culturally, educationally, and linguistically diverse populations unfamiliar with undergoing formal tests. Assessment with TULIPA tests is feasible, as long as a selection is made from the available core and supplementary tests. Given that factors complicating neuropsychological testing were observed frequently in our sample, the influence of these factors should be well-monitored and taken into consideration.

Acknowledgements

The authors would like to thank Caroline Jurgens, Wilma Smith-Spijkerboer, Lothar van Hoogdalem, Özgül Uysal-Bozkir, Charlotte Schreuder, Jennifer van den Broeke, Yuled Kalkisim, Swastie Doekhie, and Bregje Appels for their contributions to the focus groups. In addition, we would like to thank the geriatricians and neurologists in all the participating centers for their contribution to the TULIPA-project.

Disclosure statement

The authors report no conflicts of interest.

Funding

This work was supported by the Netherlands Organisation for Health Research and Development (grant number: 733050834).

ORCID

Sanne Franzen http://orcid.org/0000-0002-4461-5109
Miriam Goudsmit http://orcid.org/0000-0001-6876-0469
Rozemarijn L. van Bruchem-Visser http://orcid.org/0000-0002-6833-7641

References

Aghvinian, M., Santoro, A. F., Gouse, H., Joska, J. A., Linda, T., Thomas, K., & Robbins, R. N. (2021). Taking the test: A qualitative analysis of cultural and contextual factors impacting neuropsychological assessment of Xhosa-speaking South Africans. *Archives of Clinical Neuropsychology*, *36*(6), 976–980. https://doi.org/10.1093/arclin/acaa115

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Ardila, A. (2007). The impact of culture on neuropsychological test performance. In B. P. Uzzell, M. Ponton, & A. Ardila (Eds.), *International handbook of cross-cultural neuropsychology* (pp. 23–44). Psychology Press.

Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., Matute, E., Nitrini, R., Ostrosky-Solis, F., & Rosselli, M. (2010). Illiteracy: The neuropsychology of cognition without reading. *Archives of Clinical Neuropsychology*, *25*(8), 689–712. https://doi.org/10.1093/arclin/acq079

Arnett, P. A. (Ed.). (2013). *Secondary influences on neuropsychological test performance: Research findings and practical applications* (1st ed.). Oxford University Press.

Baiyewu, O., Unverzagt, F. W., Lane, K. A., Gureje, O., Ogunniyi, A., Musick, B., Gao, S., Hall, K. S., & Hendrie, H. C. (2005). The Stick Design test: A new measure of visuoconstructional ability. *Journal of the International Neuropsychological Society: JINS*, *11*(5), 598–605. https://doi.org/10.1017/S135561770505071X

Carone, D. A., Iverson, G. L., & Bush, S. S. (2010). A Model to approaching and providing feedback to patients regarding invalid test performance in clinical neuropsychological evaluations. *The Clinical Neuropsychologist*, *24*(5), 759–778. https://doi.org/10.1080/13854041003712951

Chen, T. B., Lin, C. Y., Lin, K. N., Yeh, Y. C., Chen, W. T., Wang, K. S., & Wang, P. N. (2014). Culture qualitatively but not quantitatively influences performance in the Boston naming test in a Chinese-speaking population. *Dementia and Geriatric Cognitive Disorders Extra*, *4*(1), 86–94. https://doi.org/10.1159/000360695

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International*, *34*, 819B.

de Paula, J. J., Costa, M. V., Bocardi, M. B., Cortezzi, M., De Moraes, E. N., & Malloy-Diniz, L. F. (2013). The Stick Design Test on the assessment of older adults with low formal education: Evidences of construct, criterion-related and ecological validity. *International Psychogeriatrics*, *25*(12), 2057–2065. https://doi.org/10.1017/S1041610213001282

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, *23*(1), 133–152. https://doi.org/10.1080/13854040701819050

Franzen, S., Papma, J. M., van den Berg, E., & Nielsen, T. R. (2021). Cross-cultural neuropsychological assessment in the European Union: A Delphi expert study. *Archives of Clinical Neuropsychology*, *36*(5), 815–830. https://doi.org/10.1093/arclin/acaa083

Franzen, S., van den Berg, E., Goudsmit, M., Jurgens, C. K., van de Wiel, L., Kalkisim, Y., Uysal-Bozkir, O., Ayhan, Y., Nielsen, T. R., & Papma, J. M. (2020). A systematic review of neuropsychological tests for the assessment of dementia in non-Western, low-educated or illiterate populations. *Journal of the International Neuropsychological Society: JINS*, *26*(3), 331–351. https://doi.org/10.1017/S1355617719000894

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

Franzen, S. on behalf of the European Consortium on Cross-Cultural Neuropsychology. (2021). Cross-cultural neuropsychological assessment in Europe: Position statement of the European Consortium on Cross-Cultural Neuropsychology (ECCroN). *The Clinical Neuropsychologist*, 1–12. https://doi.org/10.1080/13854046.2021.1981456

Franzen, S., van den Berg, E., Ayhan, Y., Satoer, D. D., Türkoğlu, Ö., Genç Akpulat, G. E., Visch-Brink, E. G., Scheffers, E. A., Kranenburg, J., Jiskoot, L. C., van Hemmen, J., & Papma, J. M. (2022). The Naming Assessment in Multicultural Europe (NAME): Development and validation in a memory clinic. *Journal of the International Neuropsychological Society*, 1–13. https://doi.org/10.1017/S135561772100148X

Fujii, D. E. M. (2018). Developing a cultural context for conducting a neuropsychological evaluation with a culturally diverse client: The ECLECTIC framework. *The Clinical Neuropsychologist*, *32*(8), 1356–1392. https://doi.org/10.1080/13854046.2018.1435826

Gilmore-Bykovskyi, A. L., Jin, Y., Gleason, C., Flowers-Benton, S., Block, L. M., Dilworth-Anderson, P., Barnes, L. L., Shah, M. N., & Zuelsdorff, M. (2019). Recruitment and retention of underrepresented populations in Alzheimer's disease research: A systematic review. *Alzheimer's & Dementia (New York, NY)*, *5*, 751–770. https://doi.org/10.1016/j.trci.2019.09.018

Goudsmit, M., Uysal-Bozkir, O., Parlevliet, J. L., van Campen, J. P. C. M., de Rooij, S. E., & Schmand, B. (2017). The Cross-Cultural Dementia Screening (CCD): A new neuropsychological screening instrument for dementia in elderly immigrants. *Journal of Clinical and Experimental Neuropsychology*, *39*(2), 163–172. https://doi.org/10.1080/13803395.2016.1209464

Goudsmit, M., van Campen, J., Franzen, S., van den Berg, E., Schilt, T., & Schmand, B. (2021). Dementia detection with a combination of informant-based and performance-based measures in low-educated and illiterate elderly migrants. *The Clinical Neuropsychologist*, *35*(3), 660–678. https://doi.org/10.1080/13854046.2020.1711967

Goudsmit, M., van Campen, J., Schilt, T., Hinnen, C., Franzen, S., & Schmand, B. (2018). One size does not fit all: Comparative diagnostic accuracy of the Rowland Universal Dementia Assessment Scale and the Mini Mental State Examination in a memory clinic population with very low education. *Dementia and Geriatric Cognitive Disorders Extra*, *8*(2), 290–305. https://doi.org/10.1159/000490174

Gruters, A., Christie, H. L., Ramakers, I., Verhey, F., Kessels, R., & de Vugt, M. E. (2021). Neuropsychological assessment and diagnostic disclosure at a memory clinic: A qualitative study of the experiences of patients and their family members. *The Clinical Neuropsychologist*, *35*(8), 1398–1414. https://doi.org/10.1080/13854046.2020.1749936

Hildebrand, R., Chow, H., Williams, C., Nelson, M., & Wass, P. (2004). Feasibility of neuropsychological testing of older adults via videoconference: Implications for assessing the capacity for independent living. *Journal of Telemedicine and Telecare*, *10*(3), 130–134. https://doi.org/10.1258/135763304323070751

Jessen, F., Amariglio, R. E., Buckley, R. F., van der Flier, W. M., Han, Y., Molinuevo, J. L., Rabin, L., Rentz, D. M., Rodriguez-Gomez, O., Saykin, A. J., Sikkes, S. A. M., Smart, C. M., Wolfsgruber, S., & Wagner, M. (2020). The characterisation of subjective cognitive decline. *The Lancet Neurology*, *19*(3), 271–278. https://doi.org/10.1016/S1474-4422(19)30368-0

Johnson, S. K., Lange, G., DeLuca, J., Korn, L. R., & Natelson, B. (1997). The effects of fatigue on neuropsychological performance in patients with chronic fatigue syndrome, multiple sclerosis, and depression. *Applied Neuropsychology*, *4*(3), 145–153. https://doi.org/10.1207/s15324826an0403_1

Jorm, A. F., & Jacomb, P. A. (1989). The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): Socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine*, *19*(4), 1015–1022. https://doi.org/10.1017/S0033291700005742

Kapur, N. (1994). The coin-in-the-hand test: A new "bed-side" test for the detection of malingering in patients with suspected memory disorder. *Journal of Neurology, Neurosurgery, and Psychiatry*, *57*(3), 385–386. https://doi.org/10.1136/jnnp.57.3.385

Karakas, S., Erdogan, E., Sak, L., Soysal, A. S., Ulusoy, T., Ulusoy, I. Y., & Alkan, S. (1999). Stroop Testi TBAG Formu: Türk Kültürüne Standardizasyon Çalışmaları, Güvenilirlik ve Geçerlilik. *Klinik Psikiyatri*, *2*, 75–88.

Kempler, D., Teng, E. L., Dick, M., Taussig, I. M., & Davis, D. S. (1998). The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society: JINS*, *4*(6), 531–538. https://doi.org/10.1017/s1355617798466013

Kilian, S., Swartz, L., Dowling, T., Dlali, M., & Chiliza, B. (2014). The potential consequences of informal interpreting practices for assessment of patients in a South African psychiatric hospital. *Social Science & Medicine (1982)*, *106*, 159–167. https://doi.org/10.1016/j.socscimed.2014.01.019

Kripalani, S., Bengtzen, R., Henderson, L. E., & Jacobson, T. A. (2008). Clinical research in low-literacy populations: Using teach-back to assess comprehension of informed consent and privacy information. *IRB*, *30*(2), 13–19.

Kunst, A. E., Stronks, K., & Agyemang, C. (2011). Non-communicable diseases. In R. Bernd, P. Mladovsky, W. Devillé, B. Rijks, R. Petrova-Benedict, & M. McKee (Eds.), *Migration and health in the European Union* (pp. 101–120). Open University Press.

Lindeboom, J., & Schmand, B. (2003). *Visual Association Test*. PITS.

Manly, J. J., & Espino, D. V. (2004). Cultural influences on dementia recognition and management. *Clinics in Geriatric Medicine*, *20*(1), 93–119. https://doi.org/10.1016/j.cger.2003.10.004

Marin, G., Sabogal, F., Marin, B. V., Otero-Sabogal, F., & Perez-Stable, E. J. (1987). Development of a short acculturation scale for Hispanics. *Hispanic Journal of Behavioral Sciences*, *9*(2), 183–205. https://doi.org/10.1177/07399863870092005

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

National Center for Education Statistics. (2003). Adult literacy supplemental assessment. https://nces.ed.gov/naal/pdf/AdultLiteracyFactSheet.pdf

Nielsen, T. R. (2019). Effects of illiteracy on the European Cross-Cultural Neuropsychological Test Battery (CNTB). *Archives of Clinical Neuropsychology*, *34*(5), 713–720. https://doi.org/10.1093/arclin/acy076

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. https://doi.org/10.1080/13854046.2013.767379

Nielsen, T. R., & Jørgensen, K. (2020). Cross-cultural dementia screening using the Rowland Universal Dementia Assessment Scale: A systematic review and meta-analysis. *International Psychogeriatrics*, *32*(9), 1031–1044. https://doi.org/10.1017/S1041610220000344

Nielsen, T. R., Vogel, A., Riepe, M. W., de Mendonça, A., Rodriguez, G., Nobili, F., Gade, A., & Waldemar, G. (2011). Assessment of dementia in ethnic minority patients in Europe: A European Alzheimer's Disease Consortium survey. *International Psychogeriatrics*, *23*(1), 86–95. https://doi.org/10.1017/S1041610210000955

Nielsen, T. R., Segers, K., Vanderaspoilden, V., Bekkhus-Wetterberg, P., Minthon, L., Pissiota, A., Bjorklof, G. H., Beinhoff, U., Tsolaki, M., Gkioka, M., & Waldemar, G. (2018). Performance of middle-aged and elderly European minority and majority populations on a Cross-Cultural Neuropsychological Test Battery (CNTB). *The Clinical Neuropsychologist*, *32*(8), 1411–1430. https://doi.org/10.1080/13854046.2018.1430256

Nijdam-Jones, A., Rivera, D., Rosenfeld, B., & Arango-Lasprilla, J. C. (2017). A cross-cultural analysis of the Test of Memory Malingering among Latin American Spanish-speaking adults. *Law and Human Behavior*, *41*(5), 422–428. https://doi.org/10.1037/lhb0000250

Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*, *29*(11), 1321–1336. https://doi.org/10.1037/pas0000438

Ortega, L. V., Aprahamian, I., Martinelli, J. E., Cecchini, M. A., Cação, J. C., & Yassuda, M. S. (2021). Diagnostic accuracy of usual cognitive screening tests versus appropriate tests for lower

education to identify Alzheimer disease. *Journal of Geriatric Psychiatry and Neurology*, *34*(3), 222–231. https://doi.org/10.1177/0891988720958542

Ostrosky-Solis, F., Ardila, A., Rosselli, M., Lopez-Arango, G., & Uriel-Mendoza, V. (1998). Neuropsychological test performance in illiterate subjects. *Archives of Clinical Neuropsychology*, *13*(7), 645–660.

Quesque, F., Coutrot, A., Cruz de Souza, L., Baez, S., Cardona, J. F., Neely-Prado, A., Clarens, M. F., Trujillo, C., Grisales, J. S., Fittipaldi, S., Magrath Guimet, N., Calandri, I. L., Crivelli, L. L. S., Garcia, A. M., Fermin, M., Indakoetxea, B., Benussi, A., Brandao Moura, M. V., … Bertoux, M. (2020). Culture shapes our understanding of others' thoughts and emotions: An investigation across 12 countries. https://hal.archives-ouvertes.fr/hal-02995977

Ponsford, J. (2017). International growth of neuropsychology. *Neuropsychology*, *31*(8), 921–933. https://doi.org/10.1037/neu0000415

Reis, A., Faisca, 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. https://doi.org/10.1016/j.bandc.2005.09.012

Reis, A., Petersson, K. M., Castro-Caldas, A., & Ingvar, M. (2001). Formal schooling influences two- but not three-dimensional naming skills. *Brain and Cognition*, *47*(3), 397–411. https://doi.org/10.1006/brcg.2001.1316

Roberts, K., Dowell, A., & Nie, J. B. (2019). Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development. *BMC Medical Research Methodology*, *19*(1), 66. https://doi.org/10.1186/s12874-019-0707-y

Rock, D., & Price, I. R. (2019). Identifying culturally acceptable cognitive tests for use in remote northern Australia. *BMC Psychology*, *7*(1), 62. https://doi.org/10.1186/s40359-019-0335-7

Román, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., Amaducci, L., Orgogozo, J. M., Brun, A., & Hofman, A. (1993). Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*, *43*(2), 250–260. https://doi.org/10.1212/wnl.43.2.250

Sedó, M. A. (2004). Test de las cinco cifras: Una alternativa multilingue y no lectora al test de Stroop ['5 digit test': A multilinguistic non-reading alternative to the Stroop test]. *Revista de Neurologia*, *38*(9), 824–828.

Selten, J.-P., Termorshuizen, F., van Sonsbeek, M., Bogers, J., & Schmand, B. (2020). Migration and dementia: A meta-analysis of epidemiological studies in Europe. *Psychological Medicine*, *50*(2), 303–313. https://doi.org/10.1017/S0033291720000586

Spreij, L. A., Gosselt, I. K., Visser-Meily, J., & Nijboer, T. (2020). Digital neuropsychological assessment: Feasibility and applicability in patients with acquired brain injury. *Journal of Clinical and Experimental Neuropsychology*, *42*(8), 781–793. https://doi.org/10.1080/13803395.2020.1808595

Storey, J. E., Rowland, J. T., Basic, D., Conforti, D. A., & Dickson, H. G. (2004). The Rowland Universal Dementia Assessment Scale (RUDAS): A multicultural cognitive assessment scale. *International Psychogeriatrics*, *16*(1), 13–31.

Sugarman, M. A., & Axelrod, B. N. (2015). Embedded measures of performance validity using verbal fluency tests in a clinical sample. *Applied Neuropsychology. Adult*, *22*(2), 141–146. https://doi.org/10.1080/23279095.2013.873439

Uysal-Bozkir, O., Hoopman, R., & De Rooij, S. E. (2016). *Translation and validation of the short Geriatric Depression Scale (GDS-15) among Turkish, Moroccan and Surinamese older migrants in the Netherlands* [Unpublished dissertation]. Health status of older migrants in the Netherlands.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, O. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*, *17*(1), 37–49.

Zendedel, R., Schouten, B. C., van Weert, J. C. M., & van den Putte, B. (2018a). Informal interpreting in general practice: The migrant patient's voice. *Ethnicity & Health*, *23*(2), 158–173. https://doi.org/10.1080/13557858.2016.1246939

Zendedel, R., Schouten, B. C., van Weert, J. C. M., & van den Putte, B. (2018b). Informal interpreting in general practice: Are interpreters' roles related to perceived control, trust, and satisfaction? *Patient Education and Counseling*, *101*(6), 1058–1065. https://doi.org/10.1016/j.pec.2018.01.012