NEUROPSYCHOLOGICAL ASSESSMENT IN THE MULTICULTURAL MEMORY CLINIC

Sanne Franzen

Neuropsychological Assessment in the Multicultural Memory Clinic

Sanne Franzen

Colophon

The research described in this PhD dissertation was supported by the Netherlands Organisation for Health Research and Development (ZonMw Memorabel) and Alzheimer Nederland. The printing of this thesis was kindly supported by Alzheimer Nederland and Erasmus University Rotterdam.

Cover image and design: Sanne Franzen & Pauw Vos
Printing: ProefschriftMaken.nl

ISBN 978-94-6423-664-4

Neuropsychological Assessment in the Multicultural Memory Clinic

Neuropsychologisch onderzoek in de
multiculturele geheugenpolikliniek

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

prof. dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 29 maart 2022 om 15.30 uur

door
Sanne Franzen
geboren te De Bilt

Erasmus University Rotterdam

Promotiecommissie

Promotor:	Prof. dr. J.C. van Swieten
Overige leden:	Prof. dr. F.U.S. Mattace Raso
	Prof. dr. B.A. Schmand
	Prof. dr. J.J. Manly
Copromotoren:	Dr. J.M. Papma
	Dr. E. van den Berg

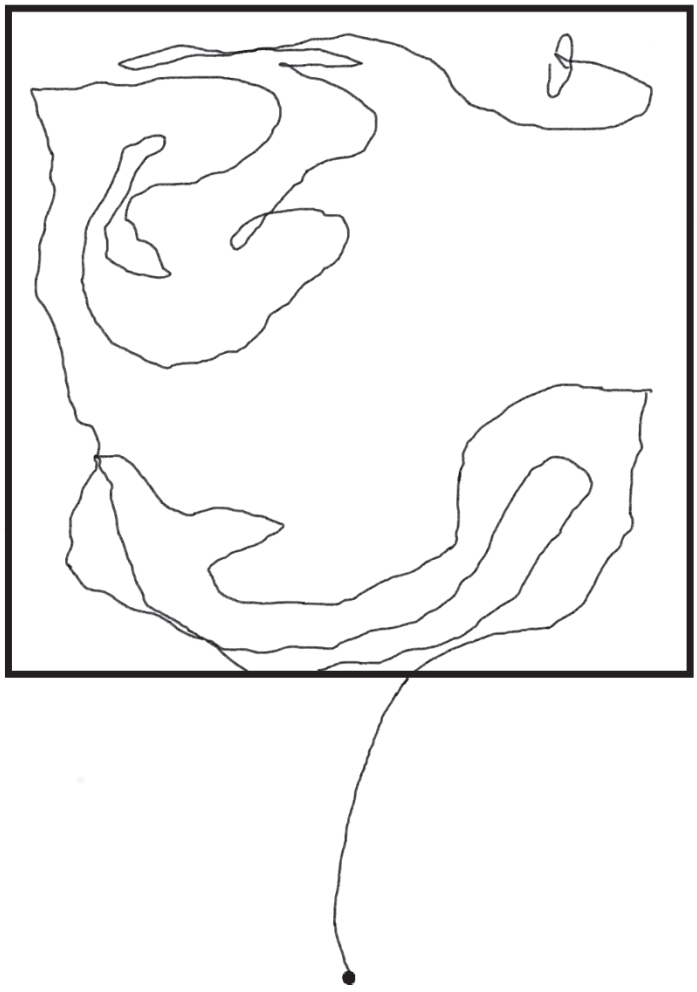
Contents

Author's note on terminology

For several decades there has been an ongoing national and international debate about the terminology we should use to describe diverse populations [1-3]. Some of the terms used in this dissertation are 'non-Western' or 'first-generation' 'immigrants', 'minority ethnic groups', and 'culturally, linguistically, and educationally diverse individuals'. Recent years have seen a trend in the United States towards terms that reflect historical inequalities, such as 'minoritized' or 'marginalized' groups, or 'historically excluded' groups. The field is in need of continued expert guidance on the most appropriate terminology to use. To this end, two workgroups have been initiated over the past three years: *Developing a Common Language and Glossary of Terms for Cultural/Cross-Cultural Neuropsychology* of the Cultural Special Interest Group of the International Neuropsychological Society and the *Diversity and Disparities Lexicon Workgroup* of the Alzheimer's Association. Neither of these glossaries/lexicons, however, were available at the start of my research project in 2017. Furthermore, these lexicons do not cover diverse populations in Europe—explaining the variation in terminology used throughout this dissertation. I ultimately chose to use 'culturally, educationally, and linguistically diverse', or a variation thereof, in the parts of this dissertation that I wrote last; this term at least for now seems to be relatively free of negative connotations. However, even if a list of 'correct' terminology were to be published, it remains to be seen how long these terms will remain relatively neutral, as long as exclusion, discrimination, and racism continue to exist. In the words of Esther Peeren, professor of Cultural Analysis at the University of Amsterdam (translated from Dutch): "All words come with a history of use, and the meaning of these words can change to a level beyond our control. We cannot simply 'clean up' language: we also need to expose the world views that underpin the words themselves." [4]

CHAPTER 1

GENERAL INTRODUCTION



1 General introduction

1.1 A brief history of cross-cultural neuropsychology

In his *Völkerpsychologie* (1900–1920), Wilhelm Wundt—often regarded as the founding father of experimental psychology—laid the groundwork for what was to become the study of cross-cultural psychology [5]. In this ten-volume work, he examines the individual as part of their external environment—language, customs, myth, culture, and history. Although cross-cultural neuropsychology as a separate discipline within psychology was not formally recognized until recent times, examples of issues in cross-cultural neuropsychological assessment were already recorded over a century ago. For example, the pioneering Binet-Simon IQ-test was developed in 1908, and Howard Andrew Knox recognized in the following years (1912–1916) that adaptations were necessary to make cognitive tests like these intelligence batteries suitable to test the immigrant populations entering the United States at Ellis Island [6]. Two decades later (1931–1933), Luria and Vygotsky made various expeditions to Uzbekistan, in which they observed firsthand how processes such as perception, problem-solving, and language development are substantially influenced by culture, literacy, and education [7]. Although the methodology and results of some of the work by Luria, such as his studies on optical illusions (using black-and-white line drawings), have been called into question [8], his work led to a novel focus on the development of so-called "culture-free" tests that could be applied across all cultures. At the time, researchers believed that limiting the number of verbal items and the need for verbal instructions would eliminate cultural effects, resulting in tests such as Cattell's Culture Fair Intelligence Test [9] containing nonverbal matrices, drawings, geometric figures/spatial arrangements, and symbols as stimuli. Later studies, however, showed that the assumption that non-verbal tests are culture-free was incorrect; some researchers actually found larger differences between cultural groups for non-verbal than for verbal tests [10,11]. Consequently a shift occurred, moving away from the concept of 'culture-free' tests towards 'culture-sensitive' tests. In the late 1990s, cross-cultural neuropsychology became an established field in (neuro)psychology through the efforts of pioneers such as the Colombian neuropsychologist Alfredo Ardila (1946–2021), one of the last students of Luria in Moscow. He was worried that "the evaluation of an alien cultural group via our neuropsychological instruments, procedures, and norms may result in serious conceptual errors" and stressed the need for insight into the values and norms underlying cognitive testing [12]. Ardila's work paved the way for studies examining the factors that influence neuropsychological testing in diverse individuals and spurred research focused on the development of novel neuropsychological tests applicable in cross-cultural contexts.

1.2 Factors that may influence cross-cultural neuropsychological assessment

1.2.1 The general context of a neuropsychological assessment

For a neuropsychological assessment to have any meaning, validity, and reliability, the patient must agree with several assumptions underlying the assessment. For example, Fujii [13] describes how a patient undergoing a neuropsychological assessment "must share Western assumptions that the universal unit for knowledge resides in the individual and not in the group"; that is, in some collectivist societies, it may be uncommon to solve problems based on individual decisions and knowledge, and patients from such groups may feel uncomfortable or unfamiliar with the individualistic nature of a neuropsychological assessment. Greenfield [14] adds that "there must be congruent expectations of test-taking

conditions including a) [the] purpose of asking questions b) what is relevant information c) decontextualized communication or talking about something that is not present, and d) comfort and acceptability in conversing with strangers". Only if these conditions are met, a neuropsychological assessment may be considered. For any cross-cultural neuropsychological assessment to succeed, however, neuropsychologists should be aware of the many contextual factors that can potentially influence the neuropsychological assessment. The ECLECTIC framework [13] describes such factors by means of eight overarching themes, which are summarized in Table 1.

Table 1. Summary of the ECLECTIC framework

Education	Level and quality of education, (il)literacy and its causes, such as geographical distance to educational facilities and limited financial means (a.k.a. social illiteracy), intellectual disabilities (a.k.a. personal illiteracy).
Culture and acculturation	How long has the patient lived in the country where the testing takes place? To what degree has the patient been immersed in/experienced the dominant culture, e.g. at work, in social life, at school? What is the cultural identity of the patient? To what degree has the person assimilated to the dominant culture?
Language	What is the native language of the patient? Does the patient speak other languages, and if so, to what degree? Is the patient bilingual or multilingual? How proficient is the patient in the majority language? Is an interpreter present?
Economic issues	This category capture the effects of socioeconomic status and poverty on performance on neuropsychological tests. This category includes feelings of discomfort with testing experienced by patients of low socioeconomic status in case of lower educational achievement due to family economic priorities, as well as limited exposure to certain stimuli in tests.
Communication style	Which style of communication does the patient use? This may dictate how, when and with whom information may be shared. This features different dimensions, such as direct versus indirect (low-context vs. high-context) communication and differences in idioms of distress.
Testing situation	The testing situation entails aspects such as comfort and motivation, and being familiar with being tested—also known as *test-wiseness*. Relevant aspects in this category may be whether it is considered appropriate for a man to be alone with a woman he has not met before, or the possible effects of ethnic matching between the neuropsychologist and the patient. Microaggressions are also mentioned in this category.
Intelligence	How is intelligence perceived across cultures? One important example is the concept of speed—in some countries, being able to respond swiftly is a sign of intelligence, but in other cultures, an intelligent response is associated with long deliberation.
Context of immigration	It is often a specific, select group within the general population of a country that immigrates to a different country—those with specific financial resources, a specific level of education, a certain age, a specific political status etc. Furthermore, it is relevant to look into traumas related to migration.

Although all of these overarching themes are important, the majority of studies have focused on the effects of culture/acculturation, language, and education; in the next paragraphs, these factors will therefore be examined in more detail.

1.2.2 Culture/acculturation & neuropsychological assessment

Before discussing the effects of culture, it is important to first define it. One such conceptualization is: "Culture consists of all those things that people have learned to do, believe, value, and enjoy in their history. It is the totality of ideals, beliefs, skills, tools,

customs, and institutions into which each member of the society is born" [15]. A cross-cultural clinical or research encounter specifically may be defined to occur: "when there are significant cultural or language differences between the examiner, examinee, informants, tests, and/or social context" [16]. Some of the most obvious effects of culture on test performance are found in tests containing items that are culture-specific [17], such as the igloo, beaver, and pretzel in the Boston Naming Test [18,19]. Perhaps more surprising effects of culture manifest in the cognitive domain of mental speed. For example, Agranovich et al. [20] found that healthy American control participants outscored their Russian counterparts on tests of mental speed due to cultural differences in familiarity with timed testing procedures. In addition, Al-Jawahiri and Nielsen [21] found that acculturation influenced scores on tests of mental speed and executive functioning, even when they used a test battery specifically composed for diverse individuals. It was hypothesized that these differences can be explained by speed being valued differently across cultures.

1.2.3. The effects of (quality of) education and literacy on the neuropsychological assessment

Education is a well-established factor influencing neuropsychological test performance [12], and it is commonly measured in years or highest completed level of education. In diverse populations, quality of education should also be considered. For example, in the USA, it was demonstrated that school quality—such as term length, class size, teacher qualifications, teacher-to-student ratio, rural vs. urban school location—is associated with cognitive functioning in late life [22,23]. Quality of education can also be measured using reading level tests, such as the National Adult Reading Test in the USA. Learning to read and write has a profound impact on the structure and functional organization of the brain [24,25], and unsurprisingly, not being able to read (well) has a substantial influence on a neuropsychological assessment. Illiteracy in particular is known to influence performance on neuropsychological tests, such as on 1) verbal memory tests [26,27], 2) language tests, particularly naming tests with black-and-white line drawings [28,29] and verbal fluency tests [30,31], 3) tests of visuoconstruction [32,33], and 4) tests of attention and executive functioning requiring literacy skills, such as the Trail Making Test (e.g. [34]).

1.2.4. The effects of language and interpretation on the neuropsychological assessment

Neuropsychological assessment of diverse individuals may be influenced by the level of proficiency in a patient's native language (or L1) as well as proficiency in languages learned later in life (L2) on cognitive test performance. Second, test performance may be influenced by the presence of an interpreter in the case of interpreter-mediated assessment. As this dissertation investigates a multicultural memory clinic setting in which interpreters are used to assess patients in L1, here I will focus on aspects of interpreter-mediated assessment.

The reliability and validity of neuropsychological assessments in which an informal interpreter is used is threatened by factors such as the exclusion of the patient from the conversation [35], problems with the adequate translation of medical terminology [36], obscuring of the patient's explanatory models, and difficulties in assessing the level of insight [37]. Although some of these problems can be overcome by using a formal interpreter, challenges may remain, especially for tests in which the interpreter needs to remember a large quantity of information or where the instructions are complex, as well as when the interpreters have received little formal training [38].

1.2.5. A word of caution

The relative contributions of language, culture, and education to test performance can be very hard to disentangle, and a risk of misinterpretation exists. For example, Manly et al. showed that differences in test performance that were previously (erroneously) attributed to race, could in fact be accounted for by differences in quality of education across different ethnoracial groups [39]. However, as data on language (20%), race/ethnicity (36%), socioeconomic status (13%), and acculturation (<1%) is infrequently reported in neuropsychological research [40], it may often be hard to ensure the interpretation of the results is valid. To improve cross-cultural neuropsychology as a science, it will be necessary to routinely measure and report relevant demographics.

1.3 Cross-cultural neuropsychology in Europe

1.3.1. An approach tailored to the diversity of Europe

To mitigate the abovementioned effects of culture, language, education, and other factors on neuropsychological test performance, several approaches have been used. A first approach—mainly applied in the USA—is to use the same neuropsychological tests for all populations, but develop (race-based) norms for different groups of individuals. This approach has been criticized in the field, however (see e.g. [41]). A second approach is to modify existing tests to better suit diverse populations. One example is the Color Trails Test [42] as a modification to the Trail Making Test. The feasibility of developing language- and culture-specific versions for all populations is likely limited, and issues of construct validity may remain in some cases. A third approach is the development of new neuropsychological tests that are more widely applicable. This seems to be the approach favored by researchers in Europe—befitting the diversity in Europe, where individuals of many different nationalities and cultural backgrounds are scattered across dozens of different countries.

1.3.2 Age-related cognitive impairment in diverse individuals in Europe

The level of diversity in Europe has steadily increased over the past 75 years; in the Netherlands specifically, this transition started with the influx of individuals from the former Dutch East Indies after gaining independence in 1949. In the period after the second world war until 1974—at which point the oil crisis hit the Netherlands—many unskilled labor immigrants came from countries such as Morocco and Turkey as 'guest workers'. In 1975, Suriname gained independence, and in the following five years almost half of the then population of Suriname immigrated to the Netherlands. Since the 1980s, other groups have found their way to the Netherlands, such as refugees from former Yugoslavia, Iraq, Iran, Somalia, Afghanistan, and Eritrea. In 1985, large groups of individuals from the Dutch Antilles moved to the mainland. Inter-EU migration, such as the influx of (seasonal) labor workers from Eastern Europe, contributed to diversity in the Netherlands in recent years. In 2019, the majority of the population of metropolises such as Rotterdam, Amsterdam, and The Hague consisted of people with a migration background [43].

Across Europe, the postwar labor immigrants and immigrants from former colonies are now reaching an age at which dementia and cognitive impairment due to other (age-related) medical conditions become more prevalent. The first European studies indicate a higher prevalence of dementia in immigrant populations [44]. In the Netherlands specifically, the number of diverse individuals with dementia was estimated to rise drastically over

the course of three decades, from 28.000 in 2014 to 38.000 in 2020 and 60.000 in 2030 [45]. The prevalence of dementia in adults over 55 was estimated at 14.8% in the Turkish community, 12.2% in the Moroccan-Arabic community, 11.3% in the Amazigh (Moroccan-Berber) community, and 12.6% in the Surinamese-Hindustani community—in contrast with 4.0% in the Surinamese-Creole and 3.5% in the native Dutch community [46]. These numbers show that dementia may be three to four times more prevalent in these diverse populations than in the native Dutch population. This increased risk is most likely due to the higher prevalence of risk factors for dementia, such as diabetes [47,48], cardiovascular disease [48], hypertension [49], depression [50], and lower education levels (e.g. [51]). Therefore, memory clinics across Europe—and in the Netherlands specifically—urgently need to reevaluate and adapt their diagnostic trajectories to suit this diverse population.

At the outset of this research project, memory clinics were unprepared for these rising numbers of diverse individuals. Several barriers to accessing dementia services can be present [52]. For example, diverse individuals may experience language barriers, a fear of discrimination, or a lack of familiarity with the health care system [53]. There can also be a lack of dementia awareness; symptoms of dementia may be explained as part of 'normal' aging or as a spiritual condition [54,55]. Patients that do find their way to the memory clinic are often faced with having to undergo a diagnostic trajectory that is insufficiently tailored to their cultural, educational, and linguistic background. For example, proficiency in Dutch can be limited in some groups, such as in the Moroccan and Turkish "guest worker" generation [51]. Similarly, a relatively larger share of this population has a low education level and/or is illiterate [51].

In 2015, together with colleagues at the former Havenziekenhuis, the Erasmus Medical Center therefore started a (pilot) multicultural memory clinic, inspired by the work of colleagues in the former Slotervaartziekenhuis in Amsterdam. In these multicultural memory clinics, dedicated services are provided for diverse individuals. Using trained bilingual and bicultural interpreters, a Dutch version of the Cultural Formulation Interview [56], and the test protocol delineated by Goudsmit et al. [52], a first attempt was made to improve services in our center. By providing a solution to the language barrier, we—somewhat naively—believed most issues would be solved. However, in reality, few of the instruments suggested by Goudsmit et al. [52] showed sufficient promise in a multicultural memory clinic setting, precluding a valid neuropsychological assessment. In essence, the Cross-Cultural Dementia Screening (CCD [57]) was the only suitable instrument that could be used to assess diverse individuals in 2017. As a screening instrument measuring aspects of memory, executive functioning, and attention/mental speed, this tool was insufficient to determine a profile of impaired and intact cognitive functions, which in turn contributes to determining the underlying etiology of the cognitive impairment.

1.4 Outline of this dissertation

In sum, there were several barriers to cross-cultural assessment at the memory clinic present at the outset of this project, with an emphasis on a lack of neuropsychological tests. The aim for this project was to take considerable steps towards a more sensitive neuropsychological assessment of diverse patients. The starting point was to investigate the major gaps in cross-cultural neuropsychology internationally and within Europe specifically. Subsequently, we aimed to make changes to the neuropsychological

assessment in multicultural memory clinics. Third, we focused on the clinical and research practices of tomorrow and the implementation of our findings in clinical practice and research.

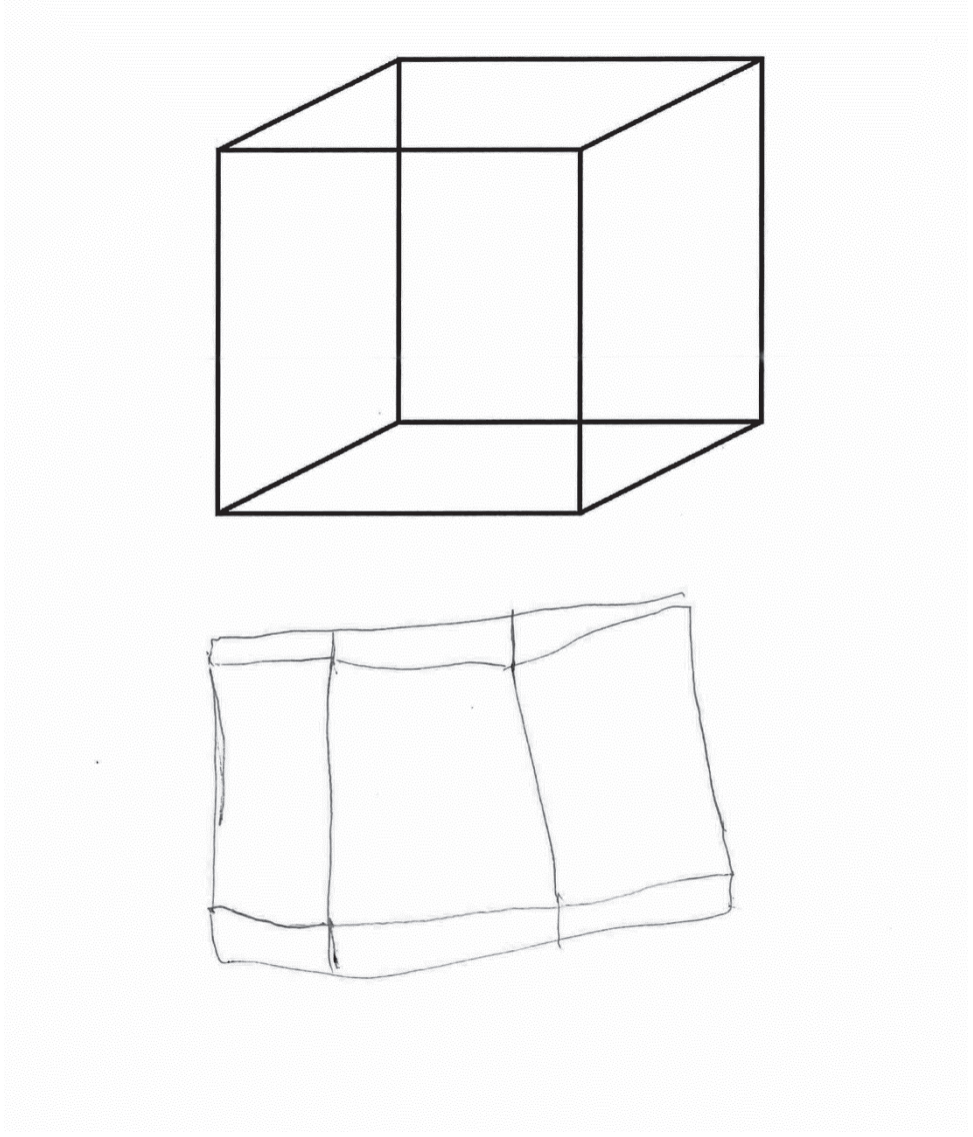
Chapter 2: State of the art of cross-cultural neuropsychological assessment
Chapter 2 of this dissertation starts out with a case study (chapter 2.1) that examines how the selection and reporting of the neuropsychological tests (available in 2017) may influence the outcomes of the assessment and have consequences for subsequent treatment. Chapter 2.2 reviews the neuropsychological tests available to diagnose dementia in low-educated, culturally diverse populations. Chapter 2.3 describes the practices in cross-cultural neuropsychological assessment across Europe and provides recommendations for future improvements to the field.

Chapter 3: Improvements to the field of cross-cultural neuropsychological assessment
A new test battery was composed based on the available international literature as identified in chapter 2, and its feasibility was studied (chapter 3.1). The next chapters zoom in on specific instruments included in this test battery. Chapter 3.2 describes the development and validation of a modified Visual Association Test (mVAT) and chapter 3.3 describes the development and validation of the Naming Assessment in Multicultural Europe (NAME), a test measuring confrontation naming in diverse individuals using photographs as stimuli. Last, chapter 3.4 focuses on culturally appropriate measurement of caregiver burden.

Chapter 4: Implementation: diversity in clinical practice and research
In chapter 4.1, I provide clinicians with recommendations on how to assess diverse individuals in a more sensitive way. In chapter 4.2, I describe the development, standpoints, and goals of the European Consortium on Cross-Cultural Neuropsychology (ECCroN) that I co-founded with the aim of improving cross-cultural neuropsychological assessment in Europe. Finally, in chapter 4.3, I review the eligibility criteria of Alzheimer's disease (AD) drug trials targeting Aβ and tau, investigating how such criteria may have impacted the inclusion of diverse participants. In this study, I also provide recommendations on how to broaden eligibility criteria to potentially make them more inclusive, with a specific focus on neuropsychological tests.

CHAPTER 2

STATE OF THE ART OF CROSS-CULTURAL NEUROPSYCHOLOGICAL ASSESSMENT



CHAPTER 2.1

Neuropsychological assessment of non-Western immigrants

Neuropsychologische diagnostiek bij niet-Westerse migranten

Sanne Franzen
Esther van den Berg
Janne M. Papma

Tijdschrift voor Neuropsychologie, 2018, 13(1), 59–67

Abstract

The number of non-Western elderly patients visiting memory clinics is rising. Cross-cultural dementia diagnosis is hindered by barriers in language and culture, as well as by low levels of education. In this article, the diagnostic trajectory of dementia—in particular the neuropsychological assessment—of the multicultural memory clinic of the Alzheimer Center Erasmus MC is described. As it stands, it can be concluded that few cross-cultural, adequately normed neuropsychological instruments are available to diagnose dementia. It is therefore of the utmost importance to pay close attention to the selection of test instruments and reporting of the results in non-Western patients.

1 Introduction

Over the next few years, the number of non-Western immigrant patients with dementia in the Netherlands is expected to rise dramatically, due to factors such as the higher prevalence of risk factors for dementia in this population [46]. This trend is already visible in the number of referrals of non-Western elderly patients to memory clinics. Neuropsychological assessment of these patients is hindered by a language barrier, by the fact that these patients are less familiar with being tested than native-born elderly, and by low education levels and/or illiteracy [52]. These barriers, combined with the fact that most neuropsychological tests that are used in the Netherlands have not been validated in immigrant populations and that no normative data is available, hinder dementia diagnostics.

Over the last few years, various cross-cultural screening measures of dementia have become available. For example, the Cross-Cultural Dementia screening (CCD [58]), a test that was specifically designed for immigrant populations, was published in 2014. This screening test contains instructions recorded in several languages: Turkish, Moroccan-Berber, Moroccan-Arabic, Dutch, Sranantongo, and Surinamese-Hindustani (Sarnami). Aside from these tests, various screening instruments are available from outside the Netherlands, such as the Mini-Mental State Examination for illiterate individuals (MMSE-I [59]) and the Rowland Universal Dementia Assessment Scale [60]. However, cross-cultural tests that measure specific cognitive domains are lacking. It therefore remains difficult to adequately assess and diagnose these patients. By means of this case study, we describe the challenges that arise in neuropsychological dementia diagnostics in non-Western immigrant patients.

2 Case description

Mr. A. is a 47-year-old man, who was diagnosed with Parkinson's disease at the age of 36. He is of Turkish descent and immigrated to the Netherlands in 1992. Mr. A. has completed primary school in Turkey and is able to read and write in Turkish, but his ability to speak Dutch is limited. Due to various somatic reasons, he quit his job in excavation work at construction sites in 2002.

As regular medication for Parkinson's disease was insufficiently effective and led to various side-effects, in 2016, the treatment team considered treating mr. A's motor symptoms with Deep Brain Stimulation (DBS). Considering the invasive nature of such a procedure, the cognitive functions of mr. A. were assessed by means of a neuropsychological assessment. The assessment, at which an informal interpreter was present, was described to have limited validity. Partly based on the conclusions from the neuropsychological assessment, it was decided that mr. A. was not eligible for DBS treatment. In 2017, mr. A. was referred to the Erasmus MC to investigate other options for treatment, such as treatment with a Duodopa pump. To explore the possibilities for treatment, the patient underwent another neuropsychological assessment, this time at the multicultural memory clinic of the Alzheimer Center of the Erasmus MC. In addition to the diagnostic trajectory at the multicultural memory clinic, a psychiatrist evaluated the patient.

2.1 Multicultural memory clinic

The interpreters of the multicultural memory clinic have a background in medicine or psychology and have been trained to interpret during the examinations of the geriatrician or neurologist and during the neuropsychological assessment. They are instructed to listen closely to several language characteristics and receive a checklist they can use for this purpose. During history taking, the cultural interview [56] is used, an instrument that can be used to increase cultural sensitivity in communication. Adapted test materials are used in the neuropsychological assessment, consisting of the CCD and other tests that are deemed to be the most suitable for non-Western elderly immigrants, as described by Goudsmit et al. [52].

2.2 Neuropsychological assessment with an interpreter

During the initial history taking interview, mr. A. mentioned having memory problems. He has difficulty finding his belongings and remembering a number of groceries. He also experiences a decline in his ability to concentrate, his mental speed, and his speed of movement. Regarding his mood he mentions that, over the last five years, he has started to become emotional and anxious more frequently, especially in social situations—i.e. when he has visitors, or when he visits the mosque.

During the interview, mr. A. provided clear and adequate responses. The interpreter noticed dysarthria and a fast pace of speaking. There were no noticeable word finding difficulties. During the neuropsychological assessment, mr. A. was cooperative and sufficiently motivated. To estimate his general cognitive functioning, three screening measures were used: the instructions of one were translated on the spot (RUDAS), whereas two others were already available in the Turkish language—the MMSE and Frontal Assessment Battery (FAB). The patient only scored below the cut-off on the FAB. Mr. A. had difficulty suppressing motor movements. As mr. A. was sufficiently able to understand and remember the instructions of the screening tests, the assessment was extended with test measuring the following cognitive domains: orientation, memory, mental speed, focused and divided attention, executive functioning, language, and visuoconstruction. Table 1 summarizes the tests that were administered, the test scores, and the norms that were used.

For nearly all tasks, mr. A. achieved scores that were equal to the expected performance based on his age and education level—largely based on Dutch norms. The assessment did not show any disorders in mental speed, focused and divided attention, memory, or orientation. Aside from the low score on the FAB, other executive functioning tasks showed unimpaired scores, such as the Sun-Moon test B of the CCD—measuring the ability to suppress cognitive interference, conceptually similar to the Stroop—and the Dots test B of the CCD—measuring divided attention, similar to the Trail Making Test. Mr. A. did not complete the Boston Naming Test as he indicated he did not recognize several of the (western) items, such as the harmonica and the stilts. Mr. A's results on often-used tasks of visuoconstruction were difficult to interpret; he was unable to copy a cube drawing, but he was able to correctly copy two overlapping pentagons (see Figure 1).

Last, some mood symptoms were measured at the neuropsychological assessment by means of the Beck Depression Inventory-II (BDI-II, Turkish version; for an overview of validation studies in different countries/languages, see Wang & Gorenstein [61]).

2.3 Psychiatric evaluation

At the psychiatric evaluation, it was concluded that mr. A. was suffering from chronic depressive symptoms. He was anxious about having hallucinations and feared what would happen if his medication would no longer work—a point of view that was easy to

Table 1. Neuropsychological tests, raw, scores, standardized scores, and normative data

Test	Raw score	Standardized score	Normative data
Screening tests			
MMSE (Turkish version)	28/30	Normal	Turkish [62]
RUDAS	27/30	Normal	Australian [60]
FAB (Turkish version)	12/18	Impaired	Turkish [63]/Dutch [64]
Neuropsychological assessment			
Language			
Boston Naming Test 60	23/34 (not completed)	Impaired	American [65]
Fluency animals/supermarket	22, 10	T = 54, T = 30	Dutch [66]
Attention, mental speed, and working memory			
Digit Span WAIS-IV	8/6/7 (span: 5/3/5)	ss = 6	Dutch (WAIS-IV-NL manual)
CCD Sun-Moon test A	16 sec	P ≥ 50	Turkish/Moroccan Verhage ≥1 [58]
CCD Dots test A	14 sec	P ≥ 50	Turkish/Moroccan Verhage ≥1 [58]
Executive functioning			
BADS key search	10	profile score = 2	British (manual)
CCD Sun-Moon test B	28 sec	P ≥ 50	Turkish/Moroccan Verhage ≥1 [58]
CCD Dots test B	97 sec	P = 10–50	Turkish/Moroccan Verhage ≥1 [58]
Letter fluency (K/A/B)	19 (7/4/8)	T = 42	Dutch [66]
Orientation & memory			
Orientation (MMSE)	Place: 4/5, time: 4/5	Normal	No norms available
Turkish RAVLT immediate recall*	37/75 (5-7-8-8-9)	T = 50	Dutch [66]
Turkish RAVLT delayed recall*	6/15	T = 47	Dutch [66]
Turkish RAVLT recognition	27/30	Normal	Dutch [66]
Turkish RAVLT delayed | immediate		T = 42	
VAT (long version)	23/24	P = 62	Dutch [67]
CCD Objects test A	122/122	Normal	Turkish/Moroccan Verhage ≥1 [58]
CCD Objects test B	122/122	Normal	Turkish/Moroccan Verhage ≥1 [58]
Visuoconstruction			
Cube drawing (RUDAS)	1/3	-	No norms available
Pentagons (MMSE)	1/1	-	No norms available
Questionnaires			
BDI-II-TR (Turkish)§	34/63	Severe depression	American [68]/Turkish [69]
CDR total; sum of boxes	0; 0.5	Normal	British [70]

Abbreviations: MMSE = Mini-Mental State Examination; RUDAS = Rowland Universal Dementia Assessment Scale; FAB = Frontal Assessment Battery; WAIS-IV = Wechsler Adult Intelligence Scale-IV; CCD = Cross-Cultural Dementia Screening; BADS = Behavioural Assessment of the Dysexecutive Syndrome; RAVLT = Rey Auditory Verbal Learning Test (*Vijftienwoordentest*); VAT = Visual Association Test; BDI-II-TR = Turkish Beck Depression Inventory-II; CDR = Clinical Dementia Rating

§ This Turkish version of the FAB has been validated in Turkey in a population of patients with schizophrenia and controls. Because the Turkish translation was not published in the paper by Güleç et al. [63], an unofficial translation has been used that has been compared with the Dutch original by two interpreters. The MMSE has been validated in patients with mild dementia in a Turkish population in Turkey; the translated version of the test is available in the paper by Gungen et al. [62]. The Dutch RUDAS is available via: www.nkop.nl/praktijk/meetinstrumenten/

* Together with three Turkish interpreters, a Turkish version of the RAVLT was developed, in which as many of the items from the existing versions of the RAVLT were used as possible, as long as they had a limited number of syllables in Turkish.

§ Validated in an adult Turkish population in Turkey. Because the translation was not published in the paper by Kapci et al. [69], we used an unofficial translation that has been compared to the Dutch original by an interpreter.

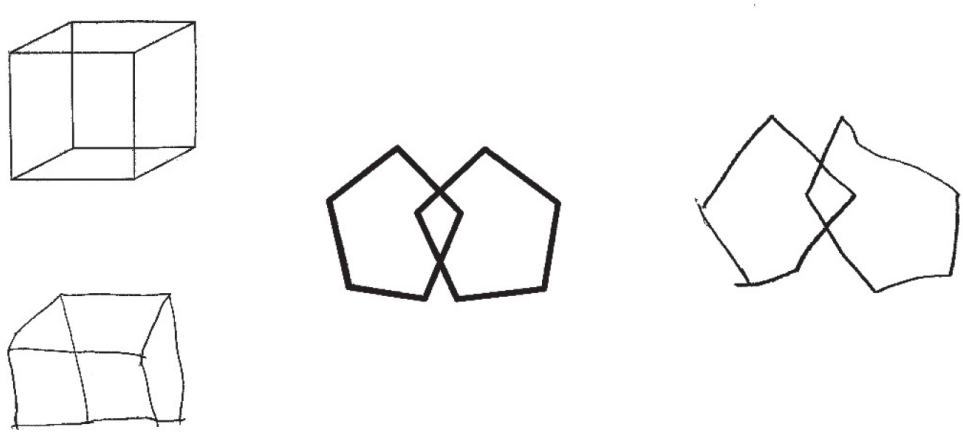


Figure 1. Cube drawing and overlapping pentagons of mr. A.

understand given his symptoms. These complaints were interpreted as, on the one hand, directly related to his Parkinson's disease and, on the other hand, stemming from the difficulty mr. A. had to accept the diagnosis and the symptoms associated with it. The psychiatrist concluded that the psychiatric symptoms were stable and expected his symptoms of anxiety and depression to decrease if he would find relief for the other symptoms of his Parkinson's disease.

3 Discussion

In this case study, we discussed a Turkish man who visited our multicultural memory clinic for a second opinion. This second neuropsychological assessment—in which an interpreter was present and the Cultural Interview and an adapted cognitive test protocol were used—did not show any disorders in cognitive functioning. Because, additionally, the (otherwise stable) mood and anxiety symptoms of the patient were expected to decrease if the motor symptoms were treated, mr. A. was found to ultimately be eligible for DBS treatment.

As in the assessment of mr. A., memory tends to be relatively easy to assess in non-Western immigrants, especially if culture-sensitive and clearly-depicted items are used, such as those in the CCD Objects test. Other domains are harder to assess. For example, immigrant patients may not be used to speed tests with a time limit [12], such as those often used to measure attention and mental speed. Language and visuoconstruction are similarly hard to measure. This is partly due to cultural differences—if Western items are used in tests [17]—and partly due to low education and illiteracy, when a patient has difficulty with 3-d drawing and angles in visuoconstructive tasks [32] or in case of difficulty with black-and-white line drawings [29]. The results on tasks of executive functioning are often influenced by education, e.g. because they require specific skills learned in the educational system or when they are abstract in nature. This is true even for the Dots test (CCD), a test that, even though it was specifically designed for this population, is not always feasible due to its dependence on skills learned in school. Aside from this lack of suitable instruments, the lack of adequate norms is also a problem.

The limited validity of the neuropsychological assessment can have consequences for the medical treatment of the patient. For example, in one institution, an Amazigh (Moroccan-Berber) man was tested with TMT-A and TMT-B, even though Tifinagh, the current Berber script, has only recently been implemented in Morocco—and is based on a non-Latin alphabet—and even though the patient had not been to school, had difficulty holding a pencil, and could not read Dutch. Another example is the incorrect diagnosis of naming impairment, when seemingly incorrect answers on the Boston Naming Test, such as 'korjaal' for canoe, are actually correct in the Surinamese dialect. Both of these examples show how the cognitive functioning of immigrant patients may be systematically underestimated. A dementia diagnosis may subsequently result in the patient being declared unfit to drive, or the patient being prescribed medication that will only lead to unnecessary side effects.

4 Conclusions

The aim of this case study was to show how difficult it can be to diagnose dementia in non-Western immigrants. Even if the assessment is carried out under optimal conditions—with an interpreter present, using adapted materials, and carried out by a culture-competent neuropsychologist—the underlying neurological or neuropsychological condition may still remain unclear. Currently, the Erasmus MC, the Maasstadziekenhuis, the MC Slotervaart, and the Haaglanden MC have joined forces in a project to develop and improve neuropsychological instruments for this specific group of patients. As long as these instruments are not available, however, several aspects are of the utmost importance: first, to use the most suitable instruments that are currently available for this population. Second, to continuously remain aware of the influence of the effects of language, culture, and education on test performance. Third, if no interpreter is present, it can be useful to stress why it is so important to interpret literally and to not help the patient, in case the relative is asked to interpret during the assessment. It is also advisable to continuously monitor whether the caregiver has actually understood the questions, as informal interpreters often misinterpret questions, or do not relay them to the patients at all [71]. Last, it is important to be careful in drawing conclusions for any assessment in which cultural or linguistic barriers were present, and/or when there was a low education level/illiteracy. In most cases, it is better to stress the limited validity of the assessment in the final concluding or summarizing sentences of the report, than in the main body of text. In our experience, having an in-person discussion with the referring physician about your conclusions—as well as of the hypotheses you could not test in your assessment—may help prevent misinterpretation and overgeneralization, and thereby reduce the likelihood of consequences that negatively impact the patient.

CHAPTER 2.2

A systematic review of neuropsychological tests for the assessment of dementia in non-Western, low educated or illiterate populations

Sanne Franzen
Esther van den Berg
Miriam Goudsmit
Caroline K. Jurgens
Lotte van de Wiel
Yuled Kalkisim
Özgül Uysal-Bozkir
Yavuz Ayhan
T. Rune Nielsen
Janne M. Papma

Journal of the International Neuropsychological Society: JINS, 2020, 26(3), 331–351

Abstract

Introduction:
Neuropsychological tests are important instruments to determine a cognitive profile, giving insight into the etiology of dementia; however, these tests cannot readily be used in culturally diverse, low-educated populations, due to their dependence upon (Western) culture, education, and literacy. In this review we aim to give an overview of studies investigating domain-specific cognitive tests used to assess dementia in non-Western, low-educated populations. The second aim was to examine the quality of these studies and of the adaptations for culturally, linguistically, and educationally diverse populations.

Methods:
A systematic review was performed using six databases, without restrictions on the year or language of publication.

Results:
Forty-four studies were included, stemming mainly from Brazil, Hong Kong, Korea, and considering Hispanics/Latinos residing in the USA. Most studies focused on Alzheimer's disease (n = 17) or unspecified dementia (n = 16). Memory (n = 18) was studied most often, using 14 different tests. The traditional Western tests in the domains of attention (n = 8) and construction (n = 15), were unsuitable for low-educated patients. There was little variety in instruments measuring executive functioning (two tests, n = 13), and language (n = 12, of which 10 were naming tests). Many studies did not report a thorough adaptation procedure (n = 39) or blinding procedures (n = 29).

Conclusions:
Various formats of memory tests seem suitable for low-educated, non-Western populations. Promising tasks in other cognitive domains are the Stick Design Test, Five Digit Test, and verbal fluency test. Further research is needed regarding cross-cultural instruments measuring executive functioning and language in low-educated people.

1 Introduction

Over the next decades, a dramatic increase is expected in the number of people living with dementia in developing regions compared to those living in developed regions [72,73], due to improvements in life expectancy and rapid population aging, especially in lower- and middle-income countries [74]. In addition, non-Western immigrant populations in Western countries, such as people from Turkey and Morocco who immigrated to Western Europe [46,75], or Hispanic people who immigrated to the USA [76], are reaching an age at which dementia is increasingly prevalent.

Most neuropsychological tests were developed to be used in (educated) Western populations. The work by Howard Andrew Knox in the early 1900s at Ellis Island already showed that adaptations are needed to make tests suitable for populations with diverse backgrounds [6]. It is now widely documented that neuropsychological test performance is substantially affected by factors such as culture, language, (quality of) education, and literacy [12,17,30,32,33,77,78]. The rising number of patients with dementia from low-educated and non-Western populations therefore calls for an increase in studies addressing the reliability, validity, and cross-cultural and cross-linguistic applicability of neuropsychological instruments used to assess dementia. Furthermore, these studies should include patients with dementia or mild cognitive impairment (MCI) in their sample to determine whether these tests are sufficiently sensitive and specific to dementia.

Recent studies have mostly focused on developing cognitive screening tests, and an excellent review is available of screening tests that can be used in people who are illiterate [79] and/or low educated [80], as well as reviews about screening tests for specific regions, such as Asia [81] and Brazil [82]. However, an overview of domain-specific cognitive tests and test batteries that are adapted to or developed for a non-Western, low-educated population is lacking. Domain-specific neuropsychological tests are essential to determine a profile of impaired and intact cognitive functions, providing insights into the underlying etiology of the dementia—something that is not possible with screening tests alone. Furthermore, a comprehensive assessment of the cognitive profile may result in more tailored, personalized care after a diagnosis [83].

The first aim of this review was to generate an overview of all studies investigating either 1) traditional neuropsychological measures, or adaptions of these measures in non-Western populations with low education levels, or 2) new, assembled neuropsychological tests developed for non-Western, low-educated populations. The second aim was to determine the quality of these studies, and to examine the validity and reliability of the current neuropsychological measures in each cognitive domain, as well as determine which could be applied cross-culturally and cross-linguistically.

2 Methods

2.1 Identification of studies

2.1.1 Search terms and databases

Studies were selected based on the title and the abstract. Medline, Embase, Web of Science, Cochrane, Psycinfo, and Google Scholar were used to identify relevant papers,

without restrictions on the year of publication or language (for a list of the search terms used, see Supplementary Material). Studies were included up until August 2018 (no start date). The papers were judged independently by two authors (SF and JMP) according to the inclusion criteria described later. In case of disagreement a consensus agreement was made together with EvdB.

2.1.2 Inclusion criteria

The inclusion criteria were as follows:

1. The study included patients with dementia and/or patients with MCI/Cognitive Impairment No Dementia (CIND).
2. The study was conducted in a non-Western country, or a non-Western population in a Western country. Western was defined as all EU/EEA countries (including Switzerland), Australia, New Zealand, Canada, and the USA. Hispanic/Latino populations in the USA were included in this review as a non-Western population, as this group likely encompasses people with heterogeneous immigration histories and diverse cultural and linguistic backgrounds [84].
3. The study described the instrument in sufficient detail for the authors to judge its applicability in a non-Western context, its validity and/or its reliability, that is, it was not merely mentioned as used during a diagnostic/research process, without any further elaboration.

2.1.3 Exclusion criteria

Studies that focused on medical conditions other than dementia were excluded. Screening tests—defined as tests covering multiple domains, but yielding a single total score without individually normed subscores—were also excluded, as some reviews of these already exist [79-82]. Intelligence tests were also excluded from the analysis, except when subtests (e.g. Digit Span) were used to assess dementia in combination with other neuropsychological tests and the study described the cross-cultural applicability. Unpublished dissertations and book chapters were excluded. Finally, studies that did not include low-educated people were excluded. This was operationalized as studies that did not describe the inclusion of low-educated or illiterate participants in the text, and did not include any education levels lower than primary school in their descriptive tables. An exception was made for studies of which the means and standard deviations of the years of education made it highly likely that low-educated participants were included, defined as a mean number of years of education that did not exceed primary school for the respective country by more than one standard deviation. Data from the UNESCO Institute for Statistics [85] were used to determine the length of primary school education for each country.

2.2 Data analysis

2.2.1 Quality assessment

The quality of the studies and the cross-cultural applicability of the instruments was assessed according to eight criteria. These criteria were developed specifically for this study to reflect important variables in the assessment of low-educated, non-Western persons. Any ambiguous cases with regard to the scoring were resolved in a consensus agreement.

The first criterion was whether any participants who are illiterate were included in the study ("Illiteracy"): 0 = no/not stated, 1 = yes. The second criterion was if the language in which

the test was administered was specified ("Language"): 0 = no, 1 = yes. The administration language can significantly influence performance on neuropsychological tests [86-88], and is especially important in the assessment of immigrants, or in countries where many languages are spoken, such as China [89]. Third, the cross-cultural adaptations were scored ("Adaptations"). For this criterion, a modification was made to the system by Beaton et al. [90] to capture the aspects relevant to neuropsychological test development: 0 = no procedures mentioned, 1 = translation (and/or back translation) or other changes to the form, but not the concept of the test, such as replacing letters with numbers or colors, 2 = an expert committee reviewed the (back)translation, or stimuli chosen by expert committee, 3 = all of the previous and pretesting, such as a pilot study in healthy controls. Assembled tests were scored either 0, 2, or 3, as no translation and back translation procedures would be required for assembled tests. The fourth criterion was whether the study reported qualitatively on the usefulness of the instrument for clinical practice, such as the acceptability of the material, acceptability of the duration of the test, and/or floor- or ceiling effects ("Feasibility"): 0 = no, 1 = yes. Illiterate people are known to be less test-wise than literate people, potentially affecting the feasibility of a test in this population [91]. Fifth, the study was scored on the availability of information on reliability and/or validity: 0 = absent, 1 = either validity or reliability data were described, 2 = both validity and reliability were described. Additionally, three criteria were proposed with regard to the final diagnosis. First, "Circularity" —whether the study described preventive measures against circularity, that is, blinding [similar to the domain "The Reference Standard" in the tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews [92]. This was scored: 0 = no/not stated, 1 = yes. Second, "Sources"—whether both neuropsychological and imaging data were used for the diagnosis, and whether a consensus meeting was held: 0 = not specified, 1 = only neuropsychological assessment or imaging, 2 = both neuropsychological assessment and imaging, and (C) for consensus meeting. As misdiagnoses are common in non-Western populations [75], it is important to rely on multiple sources of data to support the diagnosis. Third, "Criteria"—whether the study reported using subtype-specific dementia criteria: 0 = not specified, 1 = general criteria, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria [93-95] or the International Classification of Diseases and Related Health Problems (ICD) criteria, 2 = extensive clinical criteria, for example, the National Institute on Aging-Alzheimer's Association (NIA-AA) criteria [96] for Alzheimer's disease (AD) or the Petersen criteria [97] for MCI. Although a score of one point on any criterion does not necessarily directly equate with one point on any other criterion, sum scores of these eight quality criteria were calculated for each instrument to provide a general indicator of the quality of the study (with a higher score indicating a higher general quality).

In the following sections and tables, the studies are described by cognitive domain, as defined by cognitive theory and according to standard clinical practice [98]. Although neuropsychological tests often tap multiple cognitive functions, for example, verbal fluency is a sensitive measure of executive function, but also taps language and memory processes, tests are listed in only one primary cognitive domain. Studies investigating multiple cognitive instruments are described in multiple paragraphs if the tests belong to different cognitive domains. When both Western and non-Western populations are described, only the data for the non-Western group are shown. Discriminative validity is described with the Area Under the Curve (AUC), either for people with dementia versus

controls or people with MCI versus controls (when only people with MCI were included in the study). AUC classification follows the traditional academic point system (<.60 = fail, .60–.69 = poor, .70–.79 = fair, .80–.89 = good, .90–.99 = excellent). When multiple studies reported on the same (partial) study cohort, the study with the most detailed information, the largest study population and/or the most comprehensive dataset is described.

3 Results

The review process is summarized in Figure 1. The search identified 9869 citations. Furthermore, 23 citations were identified through the reference lists of included studies. After deduplication, 5071 citations remained; these citations were screened on title and abstract. If the topic of the abstract fell within the criteria, but there was insufficient information on the type of population and/or education level that was studied, the participants section and demographic tables in the full text were checked. A total of 81 studies were assessed for eligibility, of which 37 were excluded: 26 due to the fact that low-educated participants were not included in the study sample (see Figure 1).

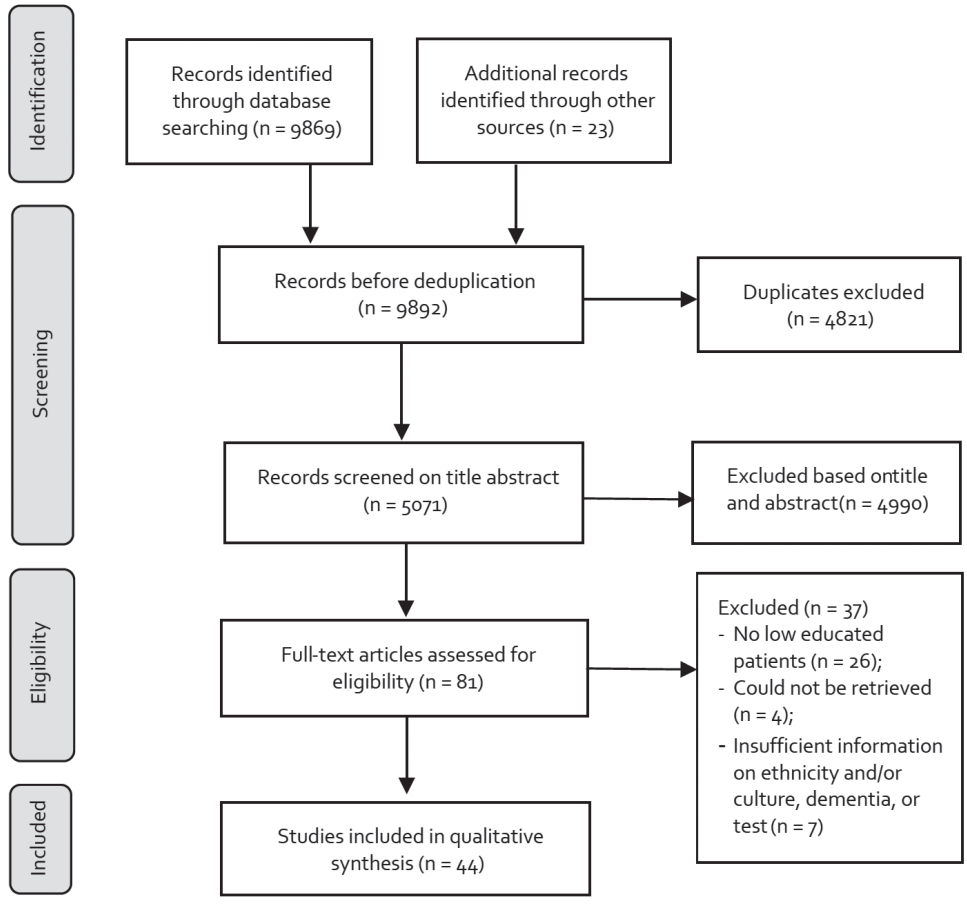


Figure 1. Results of database searches and selection process.

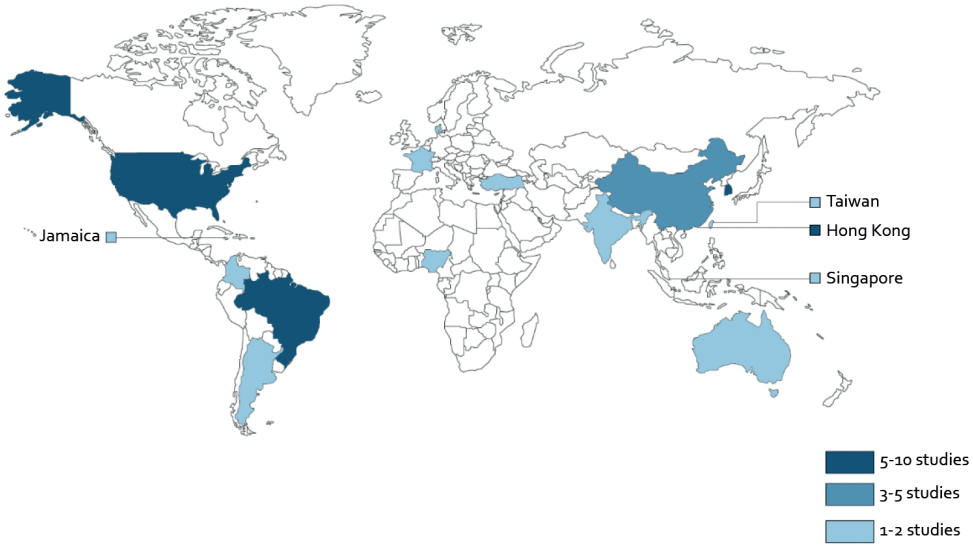


Figure 2. Number of studies per country.

A total of 44 studies were included in this review. As shown in Figure 2, most studies stemmed from Brazil, the USA (Hispanic/Latino population), Hong Kong, and Korea. Primary school education in these countries lasts 5.46 years on average (with a standard deviation of .74 years and range of 4–7 years). Seventeen studies specifically focused on a population of patients with AD, 16 studies investigated an unspecified dementia group or MCI only, and 11 studies investigated a mixed population (mostly AD and smaller groups of other dementias, or AD vs. a "non-AD" group). Of those 11 studies, only one study was specifically aimed at a type of dementia other than AD, that is, Parkinson's disease dementia (PDD).

Quality criteria scores are summarized in Supplementary Table 1. People who are illiterate were included in 26 of 44 studies. Regarding the tests that were used, 15 studies did not describe performing any translation procedures, and only five studies using an existing test described a complete adaptation procedure with translation, back translation (or other conceptual changes), review by an expert committee, and pretesting [99-103]. The language the test was administered in, or the fact that it was administered with an interpreter present, was specified in 32 studies. Aspects of the feasibility of the tests were mentioned in 25 studies. With regard to the reference standard, blinding procedures were described in 15 studies. Out of 44 studies, 14 studies made use of both imaging data and neuropsychological assessment to determine the diagnosis, 13 studies used either one of these two and 17 studies did not mention using either imaging data or a neuropsychological assessment to support the final diagnosis. Nearly all studies specified the criteria that were used to determine the diagnosis: the DSM or similar criteria were used in 15 studies, and 25 studies used specific clinical criteria. Out of 44 studies, 12 studies reported on both the reliability and the validity of the test.

Table 1. Attention.

Study, year	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
[104] 2010	Brazil	DS WAIS-III	32	20	17	AD	26	~18	22	70(5)	~75	71(6)	5 (4)	~4	4(4)	Population based	.72	4
		Corsi Blocks															.66	
[105] 2013	Brazil	DSF WAIS-III	96	93	85	AD	26	21	24	73(8)	75(7)	73(8)	5(4)	5(3)	5(4)	Other	.69	8
		DSB WAIS-III															.82	
[106] 2015	Brazil	FDT – Reading	40	40	0	–	–	–	–	76(8)[a]			5(4)[a]			Outpatient	.72	6
		FDT – Counting															.75	
		FDT – Choosing															.70	
		FDT – Shifting															.74	
[107] 2014	Brazil	DSF (unspecified)	202	21	22	–	–	–	–	70[b]	72[b]	70[b]	4[b]	2[b]	4[b]	Outpatient	.69	3
		DSB (unspecified)															.72	
[108] 2014	Korea	TMT Black and White	19	11	20	AD	28	20	26	63(6)	74(8)	69(7)	12(6)	8(7)	8(7)	Outpatient	–	6
[102] 1993	Cuban American (USA)	DS (WAIS-R)	0	38	0	AD	–	16	–	–	72(6)	–	–	9(5)	–	Outpatient	–	7
[109] 2016	China	DS (WAIS-R)	107 (PD)	33	0	PDD	27	19	–	63(9)	66(9)	–	10(4)	6(5)	–	Outpatient	.84	4
[110] 1995	China	TMT-A	67; 46[c]	61; 16[c]	0	–	23; 26[c]	17; 16[c]	–	74(8); 72(9)	78(7); 75(9)[c]	–	–	–	–	Population based	–	3
		DSST (WAIS-R)															–	3
		DS (WAIS-R)															.64–.56; .79–.46[c]	4

Abbreviations: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; AD = Alzheimer's Dementia; WAIS-R = Wechsler Adult Intelligence Scale-Revised; DS = Digit Span; DSF = Digit Span Forward; DSB = Digit Span Backward; FDT = Five Digit Test; TMT = Trail Making Test; DSST = Digit Symbol Substitution Test; PDD = Parkinson's Disease Dementia; PD = Parkinson's Disease
Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified.
– indicates no data available or not applicable.
[a] Group total. [b] Median instead of mean. [c] Entire dataset split into uneducated, educated respectively.

Table 2. Construction and perception.

Study, year	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
[111] 2010	Brazil	CDT	40	66	0	AD	23	15	-	78(7)	80(7)	-	0	0	-	Outpatient	0.83	8
[112] 2005	Nigeria	SDT CP (CERAD)	340	88	296	-	20	11	16	78(6)	80(8)	79(6)	88%	95%	93%	Population based	0.78 0.69	6
[99] 2002	Chinese (HK)	Olfactory Identification Test	12	12	0	AD	27	18	-	74(6)	76(5)	-	4(4)-	5(6)	-	Outpatient	83% classified correctly	8
[113] 2005	Chinese (HK)	CDT	34	51	0	AD, VaD, other	17[a]			78(7)	78(6)	-	3(3)	3(3)	-	Outpatient	0.81	8
[114] 2007	India	CP (CERAD)	634	0	111	-	29	-	~27	67	-	~68	8(5)	-	~6	Population based	-	8
[104] 2010	Brazil	CDT	32	20	17	AD	26	~18	22	70(5)	~75	71(6)	5(4)	~4	4(4)	Population based	0.79	4
[115] 2013	Brazil	SDT CDT	62	93	0	AD	28	21	-	75[b]	75[b]	-	4b	4b	-	Outpatient	0.76 0.84	6
[105] 2013	Brazil	SDT CDT	96	93	85	AD	26	21	24	73(8)	75(7)	73(8)	5(4)	5(3)	5(4)	Other	0.77 0.87	9
[107] 2014	Brazil	CDT	202	21	22	-	-	-	-	70[b]	72[b]	70[b]	4[b]	2[b]	4[b]	Outpatient	0.69-0.72	3
[116] 1998	Chinese (HK)	CDT/Clock Reading/Setting	53	53	0	AD, non-AD	-	-	-	74(7)	77(9)	-	4(4)[a]			Mixed sample	0.83-0.79	6
[102] 1993	Cuban American (USA)	Block Design (WAIS-R) Object Assembly (WAIS-R)	0	38	0	AD	-	16	-	-	72(6)	-	-	9(5)	-	Outpatient	- -	7
[117] 2018	Korea	CCSIT	15	20	78	AD	25[a]			72(8)[a]			9(5)[a]			Outpatient	-	3
[109] 2016	China	Block Design (WISC-III)	107 (PD)	33	0	PDD	27	19	-	63(9)	66(9)	-	10(4)	6(5)	-	Outpatient	0.91	4
[118] 2002	Chinese (SG)	CP (CERAD) Block Design (WAIS-R) Object Assembly (WAIS-R)	155	72	0	AD	24	16	-	26% ≥75	60% ≥75	-	54%	67%	-	Outpatient	- 0.78-0.91 0.89-0.74	10 9 9
[110] 1995[c]	China	CDT Block Design (WAIS-R)	113	77	0	-	23; 26[c]	17; 16[c]	-	74(8); 72(9)[c]	78(7); 75(9)[c]	-	-	-	-	Population based	- 0.66-0.64; 0.77-0.74[c]	5 4
[119] 2002	Mixed (Australia)	CDT	44	49	0	-	23[b]	14[b]	-	76(8)	80(7)	-	68%	61%	-	Outpatient	0.60 to 0.72	5
[120] 2007	Chinese (SG)	CDT	75	73	0	AD, AD with CVD, VaD	29	17	-	71(5)	78(6)	-	8(5)	5(4)	-	Outpatient	0.84 to 0.85	9

Abbreviations: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; CDT = Clock Drawing Test; AD = Alzheimer's Dementia; SDT = Stick Design Test; CP = Constructional Praxis; HK = Hong Kong; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; VaD = Vascular Dementia; WAIS-R = Wechsler Adult Intelligence Scale-Revised; CCSIT = Cross-Cultural Smell Identification Test; WISC-III = Wechsler Intelligence Scale for Children-III; SG = Singapore; PD = Parkinson's Disease; PDD = Parkinson's Disease Dementia; CVD = Cerebrovascular disease
Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified
- indicates no data available or not applicable
[a] Group total. [b] Median instead of mean. [c] Entire dataset split into uneducated, educated respectively.

Table 3. Executive functions.

Study, year	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
[111] 2010	Brazil	CVF animals	40	66	0	AD	23	15	-	78(7)	80(7)	-	0	0	-	Outpatient	0.79	8
[31] 2007	Brazil	CVF animals	117	88	0	AD	~25	~18	-	~76	~77	-	~4	~4	-	Outpatient	0.91-0.81 to 0.83–1	5
[121] 1997[a]	Chinese (HK)	CVF animals CVF fruits CVF veg.	53	56	0	AD, VaD, other	27	15	-	74(7)	77(9)	-	5(5)	3(4)	-	Instituti-onalized/ Outpatient	0.84-0.85 0.94-0.81 0.78-0.71	7
[114] 2007	India	CVF animals CVF fruits	634	0	111	-	29	-	~27	67	-	~68	8(5)	-	~6	Population based	-	8
[122] 2012	Brazil	Tower of London	60	60	60	AD	27	21	24	74(6)	76(7)	74(9)	7(3)	5(3)	6(4)	Outpatient	0.80 to 0.90	2
[104] 2010	Brazil	CVF animals	32	20	17	AD	26	~18	22	70(5)	~75	71(6)	5(4)	~4	4(4)	Population based	0.82	4
[105] 2013	Brazil	CVF animals CVF fruits Letter fluency (S)	96	93	85	AD	26	21	24	73(8)	75(7)	73(8)	5(4)	5(3)	5(4)	Other	0.92 0.87 0.85	8
[107] 2014	Brazil	CVF animals	202	21	22	-	-	-	-	70[b]	72[b]	70[b]	4[b]	2[b]	4[b]	Outpatient	0.78	3
[102] 1993	Cuban American (USA)	Letter fluency (COWAT)	0	38	0	AD	-	16	-	-	72(6)	-	-	9(5)	-	Outpatient	-	7
[123] 2004	Chinese (HK)	CVF animals CVF fruits CVF veg.	81	72	0	AD	26	17	-	75(5)	77(8)	-	4(5)	3(3)	-	Outpatient	0.87-0.93 to 0.88-0.93	6
[124] 2007	Brazil	CVF animals CVF fruit	33	24	17	AD, VaD, PDD	23	16	18	77(5)	79(5)	77(7)	2(3)	2(3)	0(1)	Outpatient	0.91 0.91	7
[118] 2002	Chinese (SG)	CVF animals	155	72	0	AD	24	16	-	26% ≥75	60% ≥75	-	54%	67%	-	Outpatient	0.81-0.90	9
[110] 1995[c]	China	CVF animals, fruits, veg. (combined)	113	77	0	-	23; 26[c]	17; 16[c]	-	74(8); 72(9)[c]	78(7); 75(9)[c]	-	-	-	-	Population based	0.67-0.70; 0.86-0.78[c]	4

Abbreviations: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; CVF = Category Verbal Fluency; AD = Alzheimer's Dementia; HK = Hong Kong; VaD = Vascular Dementia; veg. = Vegetables; COWAT = Controlled Oral Word Association Test; PDD = Parkinson's Disease Dementia; SG = Singapore
Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified
- indicates no data available or not applicable
[a] Two other fluency categories were described, but not used to assess validity. [b] Median instead of mean. [c] Entire dataset split into uneducated, educated respectively.

Table 4. Language.

Study, year	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
[114] 2007	India	Object Naming Test	634	0	111	-	29	-	~27	67	-	~68	8(5)	-	~6	Population based	-	8
[104] 2010	Brazil	Token Test	32	20	17	AD	26	~18	22	70(5)	~75	71(6)	5(4)	~4	4(4)	Population based	0.76	4
[105] 2013	Brazil	TN-LIN Token Test	96	93	85	AD	26	21	24	73(8)	75(7)	73(8)	5(4)-	5(3)	5(4)	Other	0.84/0.70/0.78 0.84/0.68	8 9
[125] 2013	Argentina	Cordoba Naming Test	26	23	0	AD	-	-	-	74(7)	76(9)	-	12(6)	13(4)	-	Outpatient	0.76	9
[107] 2014	Brazil	Naming Test (BCSB)	202	21	22	-	-	-	-	70[b]	72[b]	70[b]	4[b]	2[b]	4[b]	Outpatient	0.61	3
[126] 2017	Korea	BNT-Korean (CERAD)	452	268	0	-	21(5)[a]			74(7)[a]			6(5)[a]			Population based	-	9
[102] 1993	Cuban American (USA)	BNT Comprehension (WAIS-R)	0	38	0	AD	-	16	-	-	72(6)	-	-	9(5)	-	Outpatient	- -	7
[127] 2008 (also: [128])	Hispanic (USA)	TNT MBNT-S 15-item Spanish Naming Test	55	30	0	-	23	15	-	73(6)	78(7)	-	5[b]	1[b]	-	Outpatient	0.90 0.88 0.81	5 5
[128] 2009	Colombia	TNT MBNT-S BNT (CERAD)	20	36	0	-	27	17	-	69(10)	74(7)	-	9(4)	6(5)	-	Outpatient	- - -	5
[124] 2007	Brazil	BNT (CERAD) Naming Test (BCSB)	33	24	17	AD, VaD, PDD	23	16	18	77(5)	79(5)	77(7)	2(3)	2(3)	0(1)	Outpatient	0.76 0.88	7
[118] 2002	Chinese (SG)	BNT	155	72	0	AD	24	16	-	26% ≥75	60% ≥75	-	54%	67%	-	Outpatient	0.63-0.83	10
[110] 1995[c]	China	BNT Vocabulary (WAIS-R)	113	77	0	-	23; 26[c]	17; 16[c]	-	74(8); 72(9)[c]	78(7); 75(9)[c]	-	-	-	-	Population based	0.67-0.54; 0.80-0.59[c]	4

Abbreviations: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; AD = Alzheimer's Dementia; TN-LIN = The Neuropsychological Investigations Laboratory Naming Test; BCSB = Brief Cognitive Screening Battery; BNT = Boston Naming Test; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; WAIS-R = Wechsler Adult Intelligence Scale-Revised; TNT = Texas Spanish Naming Test; MBNT-S = Modified BNT Spanish; VaD = Vascular Dementia; PDD = Parkinson's Disease Dementia; SG = Singapore
Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified
- indicates no data available or not applicable
[a] Group total. [b] Median instead of mean. [c] Entire dataset split into uneducated, educated respectively.

Table 5. Memory.

Study, year	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC IR or SN-SP	AUC DR or SN-SP	AUC Rec or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI					
[129] 2012	Korea	Korean Story Recall Test Seoul Verbal Learning Test	53	72	127	AD	27	23	26	66(7)	73(6)	69(7)	12(5)	12(9)	11(5)	Outpatient	0.74 0.83	0.77 0.87	0.73 0.80	8
[130] 2010	Taiwan	CVVLT	217	185	0	AD	29	18	-	71(10)	79(7)	-	13(4)	10(5)	-	Outpatient	0.97	0.98	-	9
[131] 2009	Chinese (HK)	FOME	135	57	0	-	25	16	-	76(8)	79(7)	-	33%	47%	-	Outpatient	0.97	0.93	-	5
[114] 2007	India	Memory (word list)	634	0	111	-	29	-	~27	67	-	~68	8(5)	-	~6	Population based	-	-	-	10
[105] 2013	Brazil	RAVLT	96	93	85	AD	26	21	24	73(8)	75(7)	73(8)	5(4)	5(3)	5(4)	Other	0.93	0.93	0.93	8
[132] 2014	Latino (USA)	Picture FCSRT	88	24	0	-	27	24	-	72(5)	77(7)	-	8(4)	6(4)	-	Population based	0.86	-	-	6
[107] 2014	Brazil	List learning (BCSB; picture-based)	202	21	22	-	-	-	-	70b	72b	70b	4b	2b	4b	Outpatient	0.76	0.80	0.80	3
[102] 1993	Cuban American (USA)	Original WMS-LM Original WMS-VR	0	38	0	AD	-	16	-	-	72(6)	-	-	9(5)	-	Outpatient	- -	- -	- -	7
[133] 1995	Hispanics (USA)	FOME	23	27	0	AD	27	21	-	72(4)	72(8)	-	13(5)	10(5)	-	Outpatient	0.96	-	-	7
[134] 2017	Mixed (France)	TMA-93	376	94	0	AD	-	19	-	69(6)	78(7)	-	18%	20%	-	Outpatient	0.88-0.97	-	-	9
[135] 2016	Mixed (France)	TNI-93	282	87	0	-	20(5)[a]			70(7)[a]			12%[a]			Outpatient	0.87-0.96	-	-	7
[109] 2016	China	FOME	107 (PD)	33	0	PDD	27	19	-	63(9)	66(9)	-	10(4)	6(5)	-	Outpatient	0.73	-	-	4
[136] 2012	Latino (USA)	FOME	28	13	27	AD, VaD, other	21(5)[a]			79(6)[a]			5(4)[a]			Population based	0.92-0.93	-	-	5
[118] 2002	Chinese (SG))	Word list memory	155	72	0	AD	24	16	-	26% ≥75	60% ≥75	-	54%	67%	-	Outpatient	0.87-0.82	0.93-0.91	0.85-0.84	10
[137] 2006	Turkey	Enhanced cued recall	33	62	18	AD vs. non-AD	27	18/ 21	27	73(7)	74(6)/ 65(10)	69(8)	8(5)	7(5)/ 8(5)	8(5)	Outpatient	0.91	-	-	10
[110] 1995[c]	China	FOME	113	77	0	-	23; 26	17; 16;	-	74(8); 72(9)	78(7); 75(9)	-	-	-	-	Population based	0.47-0.63; 0.92-0.58	-	-	4
[138] 2006	Brazil	List Learning (CERAD) List learning (BCSB; picture-based)	51	50	0	AD, VaD, PDD, etc.	-	-	-	74(5); 74(6)	80(5); 81(7)	-	45%	43%	-	Population based	- -	0.85; 0.99 0.98; 0.98	- -	6
[139] 2012	India	PMIS	239	65	0	-	27	14	-	67(6)	72(7)	-	8(4)	7(4)	-	Outpatient	0.95-0.99	-	-	10

Abbreviations: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; CDT = Clock Drawing Test; AD = Alzheimer's Dementia; SDT = Stick Design Test; CP = Constructional Praxis; HK = Hong Kong; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; VaD = Vascular Dementia; WAIS-R = Wechsler Adult Intelligence Scale-Revised; CCSIT = Cross-Cultural Smell Identification Test; WISC-III = Wechsler Intelligence Scale for Children-III; SG = Singapore; PD = Parkinson's Disease; PDD = Parkinson's Disease Dementia; CVD = Cerebrovascular disease
Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified
- indicates no data available or not applicable
[a] Group total. [b] Median instead of mean. [c] Entire dataset split into uneducated, educated respectively.

Table 6. Test batteries.

Study, year	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
[101] 2002	Korea	CERAD	212	194	-	AD vs. non-AD	28	17	-	68(4)	70(8)	-	8(4)	6(5)	-	Outpatient	-	10
[140] 2018	Mixed (Western Europe)	CNTB	52	41	-	AD vs. non-AD	-	-	-	73(7)[a]		-	5(6)[a]		-	Outpatient	-	8
[103] 2015	Korea	LICA	634	0	128	-	26	-	22	72(6)	-	73(7)	7(5)	-	6(5)	Outpatient	0.83	11
[141] 1999	Jamaica	CERAD	72	20	-	-	23	14	-	79(6)	82(6)	-	6(3)	5(3)	-	Population based & Outpatient	-	6
[142] 2017	China	NLCA	50	-	50	-	-	-	-	40% ≥65	-	46% ≥65	44%	-	52%	Outpatient	0.94	7

Subtests of the test batteries

CERAD [101]		**CNTB [140]**		**LICA [103]**		**CERAD [141]**		**NLCA [142]**	
Subtest	**AUC**	**Subtest**	**AUC**	**Subtest**	**AUC**	**Subtest**	**AUC ***	**Subtest**	**AUC**
BNT	-	CDT	0.79	Digit Stroop	-	BNT	42%	Attention subtest	-
CP	-	CRT	0.77	Fluency	-	BNT Recall	-	Block Design ('Executive subtest')	-
CP Recall	-	CTT	~0.85	Naming	-	BNT Visual Recognition	-	Memory subtest	-
CVF animals	-	Copying semi-complex figure	0.67	Stick Construction	-	CP	25%	Reasoning subtest	-
Word List Memory	-	Copying simple figures	0.62	Story Recall	-	CVF animals	58%	Visuospatial function subtest	-
		ECR	0.96	Visual Recognition	-	IUTT	67%		
		Five Digit Test	~0.78	Visuospatial Span	-	Word List Memory	83%		
		CVF animals	0.90	Word List Memory	-				
		CVF supermarket	0.92						
		Picture Naming	0.65						
		RPT	~0.93						
		Recall semi-complex figure	0.93						

Abbreviations: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; AD = Alzheimer's Dementia; CNTB = European Cross-Cultural Neuropsychological Test Battery; LICA = Literacy Independent Cognitive Assessment; NLCA = Non-Language based Cognitive Assessment; BNT= Boston Naming Test; CDT = Clock Drawing Test; CP = Constructional Praxis; CRT = Clock Reading Test; CTT = Color Trails Test; CVF = Category Verbal Fluency; ECR = Enhanced Cued Recall; IUTT = Indiana University Token Test; RPT = Recall of Pictures Test
Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified
- indicates no data available or not applicable
[a] Group total. * correct classification rate of dementia patients.

3.1 Attention

Attention tests were described in eight studies, with a total of five different types of tests: the Five Digit Test, the Trail Making Test, the Digit Span subtest of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and WAIS-III, the Corsi Block-Tapping Task, and the WAIS-R Digit Symbol subtest (see Table 1). The Five Digit Test is a relatively new, Stroop-like test, in which participants are asked to either read or count the digits one through five, in congruent and incongruent conditions (e.g. counting two printed fives). With regard to the Trail Making Test, two studies reported on its feasibility. The traditional Trail Making Test could not be used in Chinese and Korean populations with low education levels, leading to "frustration" [110] and to a 100% failure rate, even in healthy controls [108]. An adapted version of Trail Making Test part B, in which participants had to switch between black and white numbers instead of numbers and letters, was completed by a higher percentage of both healthy controls and patients with dementia [108]. Generally, the AUCs in the domain of attention were variable, ranging from poor to good (.66–.84). In particular, the AUCs for the Digit Span test varied across studies (.69–.84).

3.2 Construction and perception

Construction tests were investigated in 15 studies, by means of five different instruments: the Clock Drawing Test, the Constructional Praxis Test of the neuropsychological test battery of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), the Stick Design Test, the Block Design subtest of the WAIS-R and of the Wechsler Intelligence Scale for Children-III (WISC-III), and the Object Assembly subtest of the WAIS-R (see Table 2). Of these tests, the Clock Drawing Test was studied most often (n = 10). The results with regard to construction tests were mixed. They were described as useful in four studies [111,113,116,120], whereas most of the others, such as Salmon et al. [110], describe this cognitive domain to be "particularly difficult for uneducated subjects" and that some patients "refused to continue because of frustration generated by the difficulty of the task". The Constructional Praxis Test was evaluated in three studies [112,114,118], and was compared with the Stick Design Test in one study [112]. In the Stick Design Test, participants are asked to use matchsticks to copy various printed designs that are similar in complexity to those of the Constructional Praxis Test. The Stick Design Test had lower failure rates (4% vs. 15%) and was also described as "more acceptable" and more sensitive than the Constructional Praxis Test [112]. Although a study by de Paula, et al. [115] also described the Stick Design Test as useful, "eliciting less negative emotional reactions [than the Constructional Praxis Test] and lowering anxiety levels", it showed ceiling effects in both healthy controls and patients, similar to the Clock Drawing Test. Generally, the Stick Design Test had fair AUCs of .76 to .79 [105,112,115]. AUCs for the Constructional Praxis were low [112], not reported [114], or left out of the report due to "low diagnostic ability" [118]. The AUCs were variable for the Clock Drawing Test, ranging from .60 to .87. The Block Design Test had lower sensitivity and specificity in the low educated than high-educated group in one study [110], and different cutoff scores for low and high education levels were recommended in a second study [118], as performance was highly influenced by education.

Perception was investigated in two studies, both focusing on olfactory processes. The study by Chan et al. [99] with the Olfactory Identification Test explicitly describes the adaptation procedure of the test. The authors did a pilot study of 16 odors specific to Hong Kong, and substituted some American items with the items that were most frequently identified as

correct in their pilot study. The correct classification rate of the test was 83%. The study by Park et al. [117] with the Cross-Cultural Smell Identification Test scored positively on only two of the quality criteria and did not provide any sensitivity/specificity data.

3.3 Executive functions

Measures of executive function were investigated in 13 studies (see Table 3), of which 12 studies used the verbal fluency test, mostly focusing on category fluency (i.e. animals, fruits, vegetables). AUCs were fair to excellent for the fluency test (between .79 and .94), although lower sensitivity and specificity were found for lower-educated participants than higher-educated participants in one study [110]. Of the six studies that included people who are illiterate (see Table 3), two observed different optimal cutoff scores for illiterate versus higher-educated groups [31,123]. Only one study investigated another measure of executive function, the Tower of London test, with low scores for the quality criteria [122]. The AUCs for the Tower of London test were good (.80–.90).

3.4 Language

Language tests were investigated in 12 studies, with a total of ten tests, or variations thereof (see Table 4). Of these ten tests, only three measured a language function other than naming: the Token Test, the Comprehension subtest of the WAIS-R, and the Vocabulary subtest of the WAIS-R. Information about the discriminative validity was not reported in three studies that used naming tests [100,102,114], as well as in all studies using the Comprehension and Vocabulary subtests of the WAIS-R [102,110]. The AUCs of the Token Test were fair (.76) in both studies [104,105]. The naming tests were frequently adapted from the Boston Naming Test, or similar types of tests making use of black-and-white line drawings. The AUCs of the naming tests varied, ranging from poor to excellent (.61–.90), with lower sensitivity and specificity for low educated than high-educated participants in one study [110].

3.5 Memory

A total of 14 memory tests were investigated in 18 studies, with stimuli presented to different modalities (visual, auditory, and tactile), and in various formats (cued vs. free recall; word lists vs. stories; see Table 5). Both adaptations of existing tests and some assembled tests were studied, such as a picture-based list learning test from Brazil [107,138] and picture-based cued recall tests in France [134,135]. AUCs were generally fair to excellent (.74–.99). Remarkably, more than half (n = 11) of the studies did not describe blinding procedures (see Table 5). With regard to specific tests, the Fuld Object Memory Evaluation (FOME), using common household objects as stimuli, was used in five studies [109,131,133,136], yielding high sensitivity and specificity rates in most studies, although one found lower sensitivity and specificity in the low-educated group [110]. However, the overall quality of the studies investigating this test was relatively low (see Table 5). Tests using a verbal list learning format [105,118,129,130,138] also had good to excellent AUCs (.80–.99). With regard to the modality the stimuli were presented to, one study [138] found that a picture-based memory test had better discriminative abilities than a verbal list learning test in the low educated, but not the higher-educated group.

3.6 Assessment batteries

Extensive test batteries were investigated in five studies (see Table 6). The studies by Lee et al. [101] and Unverzagt et al. [141] looked into versions of the CERAD neuropsychological test battery. The CERAD battery was specifically designed to create uniformity in assessment methods of AD worldwide [143] and contains category verbal fluency (animals), a 15-item version of the Boston Naming Test, the Mini-Mental State Examination, a word list learning task with immediate- and delayed recall, and recognition trials, and the Constructional Praxis Test, including a recall trial. The study by Lee et al. [101] extensively describes the difficulties in designing an equivalent version in Korean, most notably with regard to "word frequency, mental imagery, phonemic similarity and semantic or word length equivalence". In some cases, an adequate translation proved to be "impossible". Items that used reading and writing (MMSE) were replaced by items concerning judgment to better suit the illiterate population in Korea. The Trail Making Test was added in this study to assess vascular dementia (VaD) and PDD, but—similar to other studies in the domain of attention—less-educated controls had "great difficulties" completing parts A and B of this test. A second study investigated the CERAD in a Jamaican population [141]. Remarkably, eight out of 20 dementia patients were "not testable" with the CERAD battery. No further information was supplied as to the cause. The correct classification rates for the patients with dementia that did finish the battery were low (ranging from 25% to 67%)—except for the word list memory test (83%).

A study by Nielsen et al. [144] investigated the European Cross-Cultural Neuropsychological Test Battery (CNTB) in immigrants with dementia from a Turkish, Moroccan, former Yugoslav, Polish, or Pakistani/Indian background. The CNTB consists of the Rowland Universal Dementia Assessment Scale (RUDAS), the Recall of Pictures Test, Enhanced Cued Recall, the copying and recall of a semi-complex figure, copying of simple figures, the Clock Drawing Test, the Clock Reading Test, a picture naming test, category verbal fluency (animal and supermarket), the Color Trails Test, the Five Digit Test, and serial threes. The Color Trails Test and copy and recall of a semi-complex figure were not administered to participants with less than one year of education. The study showed excellent discriminative abilities for measures of memory—Enhanced Cued Recall, Recall of Pictures Test, and recall of a semi-complex figure—and category word fluency. Most of the AUCs for these tests were .90 or higher. Attention measures, that is, the Color Trails Test and Five Digit Test, had fair to good discriminative abilities, with AUCs of around .85 and .78, respectively. The diagnostic accuracy was poor for picture naming (AUC .65) and graphomotor construction tests (AUCs of .62 and .67).

A third battery was the Literacy Independent Cognitive Assessment, or LICA [103], a newly developed cognitive battery for people who are illiterate. Subtests include Story and Word Memory, Stick Construction (similar to, but more extensive than the Stick Design Test), a modified Corsi Block Tapping Task, Digit Stroop, category word fluency (animals), a Color and Object Recognition Test, and a naming test. Only the performance on Stick Construction and the Color and Object Recognition Test were not significantly different between controls and MCI patients. The AUC for the entire battery was good (.83) in both the group of people who were literate and the group of people who were illiterate, but no information was provided on the AUCs of the subtests.

The last battery was the Non-Language–based Cognitive Assessment [142], a battery primarily designed for aphasia patients, but also validated in Chinese MCI patients. It contains Judgment of Line Orientation, overlapping figures, a visual reasoning subtest, a visual memory test using stimuli chosen to match the Chinese culture, an attention task in a cross-out paradigm, and Block Design test. All demonstrations were nonverbal. The AUC was excellent (.94), but no information was available regarding the subtests.

4 Discussion

In this systematic review, an overview was provided of 44 studies investigating domain-specific neuropsychological tests used to assess dementia in non-Western populations with low education levels. The quality of these studies, the reliability, validity, and cross-cultural and/or cross-linguistic applicability were summarized. The studies stemmed mainly from Brazil, Hong Kong, and Korea, or concerned Hispanics/Latinos residing in the USA. Most studies focused on AD or unspecified dementia. Memory was studied most often, and various formats of memory tests seem suitable for low-educated, non-Western populations. The traditional Western tests in the domains of attention and construction were unsuitable for low-educated patients; instead, tests such as the Stick Design Test or Five Digit Test may be considered. There was little variety in instruments measuring executive functioning and language. More cross-cultural studies are needed to advance the assessment of these cognitive domains. With regard to the quality of the studies, the most remarkable findings were that many studies did not report a thorough adaptation procedure or blinding procedures.

A main finding of this review was that most studies investigated either patients with AD or a mixed or unspecified group of patients with dementia or MCI. In practice, this means that it remains unknown whether current domain-specific neuropsychological tests can be used to diagnose other types of dementia in non-Western, low-educated populations. Furthermore, only a third of the included studies described taking procedures against circularity of reasoning, such as blinding, potentially inflating the values for the AUCs. Only a third of the studies made use of both imaging and neuropsychological assessment to determine the reference standard. This can be problematic considering that misdiagnoses are likely to be more prevalent in a population in which barriers to dementia diagnostics in terms of culture, language, and education are present [75,145,146]. Another remarkable finding in this review was that only a handful of studies applied a rigorous adaptation procedure in which the instrument was translated, back translated, reviewed by an expert committee, and pilot-tested. These studies highlight the difficulty of developing a test that measures a cognitive construct in the same way as the original test in terms of the language used and the difficulty level. Abou-Mrad et al. [147] elegantly describe these difficulties and provide details for the interested reader about the way some of these issues were resolved in their study.

With regard to specific cognitive domains, the tests identified in this review that measured attention were the Trail Making Test, WAIS-R Digit Span, Corsi Block Tapping Task, WAIS-R Digit Symbol, and Five Digit Test. It was apparent that traditional Western paper-and-pencil tests (Trail Making Test, Digit Symbol) are hard for uneducated subjects [101,108,110]. It therefore seems unlikely that these types of tests will be useful in low-educated, non-

Western populations. With regard to Digit Span tests, previous studies have indicated that performance levels vary depending on the language of administration, for example, due to the way digits are ordered in Spanish versus English [148], or due to a short pronunciation time in Chinese [149]. This makes Digit Span less suitable as a measure for cross-linguistic evaluations in diverse populations. On the other hand, the Five Digit Test does not seem to suffer from this limitation: it is described by Sedó [150] as less influenced by differences in culture, language, and formal education, partially because it only makes use of the numbers one through five, that most illiterate people can identify and use correctly (according to Sedó).

Western instruments used to assess the domain construction, such as the Clock Drawing Test, led to frustration in multiple studies and had limited usefulness in the clinical practice with low-educated patients. This is in line with the finding by Nielsen and Jørgensen [32], that even healthy illiterate people may experience problems with graphomotor construction tasks. The Stick Design Test, that does not rely on graphomotor responses, was described as more acceptable for low-educated patients. Given the ceiling effects that were present in one study [115], as well as the differences in performance between the samples from Nigeria [112] and Brazil [115], further studies on this instrument are required.

Interestingly, no studies in the domain of Perception and Construction focused specifically on the assessment of visual agnosias, although a test of object recognition and a test with overlapping figures were included in two test batteries. As agnosia is included in the core clinical criteria of probable AD [96], it is important to have the appropriate instruments available to determine whether agnosia is present. The only tests measuring perception were two smell identification tasks [99,117]. In recent years, this topic has received more attention from cross-cultural researchers. Although olfactory identification is influenced by experience with specific odors [151], and tests would therefore have to be adapted to specific populations, deficits in olfactory perception have been described in the early stages of AD and PDD [152]. As this task might also be considered to be ecologically valid, it may be an interesting avenue for further research. The study by Chan et al. [99] with the Olfactory Identification Test explicitly describes the selection procedure of the scents used in the study, making it easy to adapt to other populations.

With regard to executive functioning, nearly all studies examined the verbal fluency test. In addition, the Tower of London test was examined in one study, and some subtests of attention tests tap aspects of executive functioning as well, such as the incongruent trial of the Five Digit Test or the Color Trails Test part 2. This relative lack of executive functioning tests poses significant problems to the diagnosis of Frontotemporal Dementia (FTD) and other dementias influencing frontal or frontostriatal pathways, such as PDD and dementia with Lewy Bodies (DLB) [153,154]. Although this review shows that a limited amount of research is available on lower-educated populations, studies in higher-educated populations have given some indication of the clinical usefulness of other types of executive functioning tests in non-Western populations. For example, Brazilian researchers [155,156] found the Rule Shift, Modified Six Elements, and Zoo Map subtests of the Behavioral Assessment of the Dysexecutive Syndrome to be useful in discriminating Brazilian patients with AD from controls. It would be interesting to see whether these subtests can be modified so they can be applied with patients who have little to no formal education.

The results in the cognitive domain of language showed that (adapted) versions of the Boston Naming Test were most often studied. This is remarkable, as it is known that even healthy people who are illiterate are at a disadvantage when naming black-and-white line drawings, such as those in the Boston Naming Test, compared to people who are literate [29].This disadvantage disappears when a test uses colored images or, better yet, real-life objects [28,29].Considering low-educated patients, Kim et al. [100] describe an interesting finding: although participants with a low education level scored lower on the naming test, remarkable differential item functioning was discovered; the items "acorn" and "pomegranate" were easier to name for low-educated people than higher-educated people, and the effect was reversed for "compass" and "mermaid". The authors suggest that this may be due to these groups growing up in rural versus urban areas, thereby acquiring knowledge specific to these environments. New naming tests might therefore benefit from differential item functioning analyses with regard to education, but also other demographic variables. It was surprising that none of the studies examined a cross-culturally and cross-linguistically applicable test, even though such a test has been developed, that is, the Cross-Linguistic Naming Test [17]. The Cross-Linguistic Naming Test has been studied in healthy non-Western populations from Morocco, Colombia, and Lebanon [157,158], as well as in Spanish patients with dementia [158]. These studies preliminarily support its cross-cultural applicability, although more research is needed in diverse populations with dementia.

Memory was the cognitive domain that was most extensively studied, in different formats and with stimuli presented to different sensory modalities: visual, auditory, and tactile. Both adaptations of existing tests and assembled tests were studied. The memory tests in this review generally had the best discriminative abilities of all cognitive domains that were studied. Although this is a positive finding, given that memory tests play a pivotal role in assessing patients with AD, memory tests alone are insufficient to diagnose, or discriminate between, other types of dementia, such as VaD, DLB, FTD, or PDD.

For the majority of the test batteries that were described, information about the validity of the subtests was not provided. An exception is the study of the CNTB [144]. Largely in line with the other findings in this review, the memory tests of the CNTB performed best, whereas the tests of naming and graphomotor construction performed worst. Attention tests, such as the Color Trails Test and Five Digit Test, performed relatively well. In sum, the CNTB encompasses a variety of potentially useful subtests. Similar to the CNTB, the LICA also includes less traditional tests, such as Stick Construction and Digit Stroop, but the lack of information about the discriminative abilities of the subtests makes it hard to judge the relative value of these tests for the cross-cultural assessment of dementia.

In this review, special attention was paid to the influence of education on the performance on neuropsychological tests. Interestingly, the discriminative abilities of the tests were consistently lower for low-educated participants than high-educated patients [110]. It has been suggested that tests with high ecological validity may be more suitable for low-educated populations than the (Western) tests that are currently used. Perhaps inspiration can be drawn from the International Shopping List Test [159] for memory, the Multiple Errands Test for executive functioning [160], or even its Virtual Reality (VR) version [161],

or other VR tests, such as the Non-immersive Virtual Coffee Task [162] or the Multitasking in the City Test [163].

Some limitations must be acknowledged with respect to this systematic review. It can be argued that this review should not have been limited to dementia or MCI, and should have also included studies of healthy people—for example, normative data studies—or studies of patients with other medical conditions. The inclusion criterion of patients with dementia or MCI was chosen as it is important to know if and how the presence of dementia influences test performance, before a test can be used in clinical practice. That is: is the test sufficiently sensitive and specific to the presence of disease and to disease progression? If this is not the case, using the test might lead to an underestimation of the presence of dementia, or problems differentiating dementia from other conditions.

Furthermore, with regard to the definition of the target population of this review, questions may be raised whether African American people from the USA should have been included. Although differences in test performance have indeed been found between African Americans and (non-Hispanic) Whites, these differences mostly appear to be driven by differences in quality of education, as opposed to differences in culture [39,164,165]. Although a very interesting topic for further research, the absence of cultural or linguistic barriers in this population has led to the exclusion of this population in this review.

Lastly, a remarkable finding was the relative paucity of studies from regions such as Africa and the Middle East. It is important to note that, although the search was thorough and studies in other languages were not excluded from this review, some studies without titles/abstracts in English, or studies that were published in local databases, may not have been found. For example, a review by Fasfous et al. [166] describes how Arabic-speaking countries have their own data bases (e.g. Arabpsynet) and how an adequate word for "neuropsychology" is lacking in Arabic. Similar databases are known to exist in other regions as well, such as LILACS in Latin America [82].

A strength of this review is that it provides clinicians and researchers working with non-Western populations with a clear overview of the tests and comprehensive test batteries that may have cross-cultural potential, and could be further studied. For example, researchers might use tests from the CNTB as the basis of the neuropsychological assessment, and supplement it with other tests. If preferred, memory tests can also be chosen from the wide variety of memory tests with good AUCs in this review, such as the Fuld Object Memory Evaluation. Researchers are advised against using measures of attention and construction that are paper-and-pencil based, and instead to use tests such as the Five Digit Test for attention, or the Stick Design Test for construction. With regard to executive functioning, it is recommended to look for new, ecologically valid tests to supplement existing tests such as the category verbal fluency test and the Five Digit Test. Furthermore, it is recommended to use language tests that are not based on black-and-white line drawings, but instead use colored pictures, photographs, or real-life objects. The Cross-Linguistic Naming Test might have potential for such purposes.

Other recommendations for future research are to study patients with a variety of diagnoses, including—but not limited to—FTD, DLB, VaD, and primary progressive aphasias.

However, as this review has pointed out, this will remain difficult as long as adequate tests to assess these dementias are lacking. It is therefore recommended that future studies support the diagnosis used as the reference standard by additional biomarkers of disease, such as magnetic resonance imaging scans or lumbar punctures. Another suggestion is to carry out validation studies in patients with dementia for instruments that have only been used in healthy controls or for normative data studies. Lastly, it is recommended that test developers use the most up-to-date guidelines on the adaptation of cross-cultural tests, such as those by the International Test Commission [167] and others [168,169], and report in their study how they met the various criteria described in these guidelines.

In conclusion, the neuropsychological assessment of dementia in non-Western, low-educated patients is complicated by a lack of research examining cognitive domains such as executive functioning, non-graphomotor construction, and (the cross-cultural assessment of) language, as well as a lack of studies investigating other types of dementia than AD. However, promising instruments are available in a number of cognitive domains that can be used for future research and clinical practice.

Acknowledgments and funding
This study was supported by grant 733050834 from the Netherlands Organization of Scientific Research (ZonMw Memorabel). The authors would like to thank Wichor Bramer from the Erasmus MC University Medical Center Rotterdam for his help in developing the search strategy.

Conflict of interest
The authors have nothing to disclose.

Chapter 2.3 Supplementary material

Supplementary Table 1. Quality criteria.

Study, year	Population (country)	Tests	Illiteracy	Language	Adaptations	Feasibility	Psychometrics	Circularity	Sources	Criteria	Sum
[111] 2010	Brazil	CDT, CVF	1	1	1	0	1	1	1	2	8
[129] 2012	Korea	SVLT, KSRT	0	1	1	1	1	0	2	2	8
[112] 2005	Nigeria	SDT, CP	1	0	0	1	1	0	2 (C)	1	6
[31] 2007	Brazil	CVF	1	0	0	0	1	0	1	2	5
[99] 2002	Chinese (HK)	Olfactory Identification Test	1	1	3	1	1	0	0	1	8
[113] 2005	Chinese (HK)	CDT	1	1	0	1	2	1	1	1	8
[130] 2010	Taiwan	CVVLT	1	1	1	0	1	1	2	2	9
[121] 1997	Chinese (HK)	CVF	1	0	1	1	2	1	0	1	7
[131] 2009	Chinese (HK)	FOME	1	0	0	1	2	1	0	0	5
[114] 2007	India	CP, CVF, Object naming test, Memory test	1	1	1	0	2	0	1	2	8
[104] 2010	Brazil	Digit Span (WAIS-III), Corsi Blocks, CDT, CVF, Token Test	0	0	0	0	1	0	1	2	4
[122] 2012	Brazil	Tower of London	0	0	0	0	0	0	0	2	2
[115] 2013	Brazil	CDT, SDT	1	0	0	1	1	0	1	2	6
[105] 2013	Brazil	Digit Span (WAIS-III), CDT, SDT, CVF, letter fluency, TN-LIN Naming*, Token Test, RAVLT	1	1	0–1	0–1	2	0	2 (C)	2	8–9
[106] 2015	Brazil	Five Digit Test*	0	0	0	0	1	1	2 (C)	2	6
[125] 2013	Argentina	Cordoba Naming Test*	0	1	3	0	1	0	2 (C)	2	9
[132] 2014	Latino (USA)	pFCSRT	0	1	1	1	1	0	1	1	6
[107] 2014	Brazil	Digit Span (unspecified), CDT, CVF, Naming Test (BCSB)*, List learning (BCSB; picture based)*	0	0	0	0	1	0	2 (C)	0	3
[108] 2014	Korea	TMT Black-and-White	1	0	1	1	1	0	0	2	6
[100] 2017	Korea	BNT-Korean (CERAD)	0	1	3	1	1	0	2	1	9
[116] 1998	Chinese (HK)	CDT	0	1	0	1	2	1	0	1	6
[101] 2002	Korea	CERAD battery	1	1	3	1	2	0	0	2	10
[102] 1993	Cuban American (USA)	Digit Span (WAIS-R), Block Design, Object Assembly (WAIS-R), letter fluency, BNT, Comprehension (WAIS-R), WMS-LM, WMS-VR	0	1	3	0	0	0	1	2	7
[133] 1995	Hispanics (USA)	FOME	0	1	0	1	1	0	2	2	7
[135] 2016	Mixed (France)	TNI-93*	1	0	3	0	1	1	0	1	7
[134] 2017	Mixed (France)	TMA-93*	1	1	3	1	1	1	0	1	9

Supplementary Table 1. Continued.

Study, year	Population (country)	Tests	Illiteracy	Language	Adaptations	Feasibility	Psychometrics	Circularity	Sources	Criteria	Sum
[127] 2008	Hispanics (USA)	MBNT-S, TNT, 15-item Spanish Naming Test	0	1	1	0	2	0	0	1	5
[128] 2009	Colombia	BNT (CERAD), BNT-S, TNT	0	1	1	0	2	0	0	1	5
[123] 2004	Chinese (HK)	CVF	1	1	0	1	1	0	0	2	6
[140] 2018	Mixed (Denmark)	CNTB	1	1	0	1	1	0	2 (C)	2	8
[117] 2018	Korea	CC-SIT	0	0	0	0	1	0	0 (C)	2	3
[109] 2016	China	Digit Span (WAIS-R), Block Design (WISC-III), FOME	1	1	0	0	1	0	1	0	4
[124] 2007	Brazil	CVF, BNT (CERAD), Naming Test (BCSB)*	1	1	0	1	1	0	1	2	7
[136] 2012	Latino (USA)	FOME	0	1	1	0	1	0	1 (C)	1	5
[118] 2002	Chinese (SG)	CP, Block Design, Object Assembly (WAIS-R), CVF, BNT, Word list memory	1	1	1	0–1	1	1	2	2	9–10
[137] 2006	Turkey	Enhanced Cued Recall	1	1	1	1	1	1	2	2	10
[110] 1995	China	Digit Span (WAIS-R), TMT, Digit Symbol Substitution (WAIS-R), CDT, Block Design (WAIS-R), CVF, BNT, Vocabulary (WAIS-R), FOME	1	1	0–1	0–1	0–1	0	0	0	3–5
[103] 2015	Korea	LICA	1	1	3	0	2	0	2	2	11
[119] 2002	Mixed NESB (Australia)	CDT	0	1	0	0	2	1	0	1	5
[138] 2006	Brazil	List learning (CERAD), List learning (BCSB; picture based)*	1	1	0	0	1	0	1 (C)	2	6
[141] 1999	Jamaica	CERAD battery	0	1	1	1	1	1	0 (C)	1	6
[139] 2012	India	PMIS*	1	1	3	1	1	1	1 (C)	1	10
[142] 2017	China	NLCA	0	1	1	1	2	0	0	2	7
[120] 2007	Chinese (SG)	CDT	1	1	0	1	1	1	2 (C)	2	9

Abbreviations: CDT = Clock Drawing Test; CVF = category verbal fluency; HK = Hong Kong; SVLT = Seoul Verbal Learning Test; KSRT = Korean Story Recall Test; SDT = Stick Design Test; CP= Constructional Praxis; CVVLT = Chinese Version Verbal Learning Test; FOME = Fuld Object Memory Evaluation; TN-LIN = The Neuropsychological Investigations Laboratory Naming Test; RAVLT = Rey Auditory Verbal Learning Test; pFCSRT = Free and Cued Selective Reminding Test (picture version); BCSB = Brief Cognitive Screening Battery; TMT = Trail Making Test; BNT = Boston Naming Test; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; SG = Singapore; WAIS-R = Wechsler Adult Intelligence Scale-Revised; WMS-LM = Wechsler Memory Scale Logical Memory (original); WMS-VR = Wechsler Memory Scale Visual Reproduction (original); TNI-93 = Test des Neuf Images du 93; TMA-93 = Memory Associative Test of the district of Seine-Saint-Denis; MBNT-S = Modified Boston Naming Test Spanish; TNT = Texas Spanish Naming Test; CNTB = European Cross-Cultural Neuropsychological Test Battery; CC-SIT = Cross-Cultural Smell Identification Test; WISC-III = Wechsler Intelligence Scale for Children-III; LICA = Literacy Independent Cognitive Assessment; NESB = non-English speaking background; PMIS = Picture based Memory Impairment Screen; NLCA = Non-Language Based Cognitive Assessment
* Assembled test

CHAPTER 2.3

Cross-cultural neuropsychological assessment in the European Union: a Delphi expert study

Sanne Franzen
Janne M. Papma
Esther van den Berg
T. Rune Nielsen

Archives of Clinical Neuropsychology, 2021, 36(5), 815–839

Abstract

Introduction:
The increasing ethnic diversity in the European Union (EU) calls for adaptations to neuropsychological assessment practices. The aims of this study were to examine the current state of cross-cultural neuropsychological assessment in EU-15 countries and to provide recommendations for researchers and policy makers.

Method:
Twelve experts from nine EU-15 countries participated in a Delphi consensus study involving two sequential rounds of web-based questionnaires and an in-person consensus meeting. The experts individually rated Delphi topics on the basis of importance (scale 1–10). The degree of consensus was determined by assessing first and third quartiles (Q1 and Q3) and medians.

Results:
Consensus outcomes showed the following priorities: a) the development of tests (median importance rating 10, Q1-Q3: 9–10), b) the collection of normative data (median importance rating 9, Q1-Q3: 8–10), and c) more training, awareness, and knowledge regarding cross-cultural assessment among neuropsychologists in the EU (median importance rating 9, Q1-Q3: 8–10). Whereas memory tests were often available, tests measuring social cognition (median 9, Q1-Q3: 8–10) and language (median 9, Q1-Q3: 7–10) are particularly lacking. Recommendations were made regarding essential skills and knowledge necessary for cross-cultural neuropsychological assessment.

Conclusions:
This study in a small group of experts suggests that the development and availability of cross-cultural tests and normative data should be prioritized, as well as the development and implementation of training initiatives. Furthermore, EU guidelines could be established for working with interpreters during neuropsychological assessment. Before implementing these recommendations, follow-up studies are recommended that include more minority neuropsychologists and community stakeholders.

1 Introduction

Although a certain degree of diversity has always been present in European Union (EU) countries, diversity levels have increased greatly over the last seven decades, starting with the immigration of labor workers from countries outside the EU from 1950–1974 and the immigration of people from once colonized countries, followed by the influx of asylum seekers and refugees in more recent years [170]. Therefore, the 15 original EU-countries, or EU-15, have had to adjust rapidly to the increasing diversity in their societies. Several minority ethnic groups in EU-15 countries are at an increased risk of medical conditions that are associated with cognitive impairment, such as stroke [47], diabetes mellitus [47], and dementia [46,171]. Furthermore, other conditions that can influence cognition may occur in some minority ethnic groups, such as tropical diseases like malaria [172] and schistosomiasis [173] in people who recently emigrated from endemic areas, malnutrition in refugees [174], and exposure to occupational hazards, such as pesticides, in labor workers [175]. As a result, neuropsychologists in EU-15 countries will increasingly encounter patients from minority ethnic groups in their daily practice.

Several characteristics of minority ethnic groups may pose unique challenges to neuropsychologists. First, limited proficiency in the host country language is widespread among older people in some minority ethnic groups in EU-15 countries, including Moroccans and Turks in the Netherlands [51], South Asians in the UK [176], Turks in Germany [177], and Turks and Vietnamese in Belgium [177]. The language in which neuropsychological tests are administered, as well as the level of formality used, can significantly impact communication, rapport, and subsequent test scores [86,87,178]. Interpretation through (formal or informal) interpreters is often needed to assess these patients in their native language. Second, low education levels or illiteracy are common among (older) people in various minority ethnic groups in EU-15 countries [51,176,179]. For example, more than 80% of Moroccan first-generation immigrants in the Netherlands did not complete any form of formal education [51]. Illiteracy, a limited number of years of education, as well as a low quality of education significantly impact neuropsychological test scores across several cognitive domains [28,29,32,39,77,91,180,181]. Patients who are illiterate may also experience more discomfort in testing situations due to unfamiliarity with the setting, the content of the tests, or due to differences in what is considered a good response [13]. Third, neuropsychologists in EU-15 countries may encounter substantial cultural barriers in their clinical practice. In particular, the "guest workers", who came to EU-15 countries as labor migrants in the post-World War II period, may have limited levels of acculturation to the dominant culture as they were initially expected to return to their countries of origin after a number of years—often resulting in a delay of decades in the development of policies promoting social integration and acculturation [170,182,183]. Cultural differences may impact the neuropsychological assessment in several ways. The patient may have different expectations of (the purpose of) the assessment, of what is relevant information, and of what information may be shared with a stranger [14]. Additionally, culture influences communication styles, idioms of distress, and the way symptoms may manifest themselves [13]. In addition, Al-Jawahiri & Nielsen [21] showed that lower levels of acculturation are associated with poorer performance on tests of mental speed and executive functioning—even when tests were administered in the person's native language and scores were corrected for other demographics. Furthermore, culture and acculturation may influence

test scores when Western items are used in tests (e.g. in naming tests) or when the tests involve culture-specific testing elements and strategies [12,17,178,180].

In sum, language, (quality of) education, literacy, and culture substantially influence neuropsychological assessment. Thorough adaptations or newly developed neuropsychological tests are needed, but such tests are often lacking [184]. Although neuropsychologists in several of the countries of origin of minority ethnic patients are working on the validation of cognitive tests, these initiatives mostly seem to focus on tests originally designed for (educated) populations in North America and Europe, such as the Trail Making Test [185] or the Montreal Cognitive Assessment [186] in Morocco, and tests from the BİLNOT battery in Turkey [187]. Furthermore, people who are low educated or illiterate were not included in these validation studies or in the normative data samples.

Taking these barriers into consideration, administering a cross-cultural neuropsychological assessment requires neuropsychologists to acquire culture-competent skills and knowledge. Some general directions for training of psychologists are presented in the "Guidelines on Multicultural Education, Training, Research, Practice, and Organizational Change for Psychologists" by the American Psychological Association [188]. However, these guidelines are not specific to neuropsychologists. Additionally, EU- and USA-based neuropsychologists state that the ability to handle cultural diversity is a "vital functional competency" for clinical neuropsychologists worldwide [189] and, more specifically, "one of the foundational entry-level competencies for neuropsychologists" [190]. However, no details are provided on the specific knowledge or skills that EU-based neuropsychologists should acquire to attain sufficient competence to handle the substantial barriers in culture, language, and education.

All these factors pose challenges to the neuropsychological assessment of patients from minority ethnic groups. The question thus arises how neuropsychologists in EU-15 countries have adapted their clinical practice to the growing population of patients from minority ethnic groups. The first aim of this study was therefore to investigate the current state of the field of adult cross-cultural neuropsychological assessment in EU-15 countries. A second aim was to generate recommendations for researchers and policy makers on the main issues that should be addressed and the potential ways to approach these issues.

2 Methods

To systematically gather expert opinion data and reach a consensus among these experts, a Delphi study method was used, focusing on the former EU-15 countries—Belgium, Denmark, Germany, Finland, France, Greece, Ireland, Italy, Luxembourg, the Netherlands, Austria, Portugal, Spain, the UK, and Sweden. These EU-15 countries share a history of similar immigration patterns with prominent (de)colonialization [170,191] and post-World War II labor immigration [170,192]. The following definition of minority ethnic people was used in this study [193]: "people who are first-generation immigrants or refugees from countries outside the extended EU, Canada, USA, Australia or New Zealand". The study was split into three rounds [194]. The first two rounds consisted of web-based surveys in which the panelists were blinded to other the panelists' responses. The last round was a

face-to-face meeting. As not all panelists could participate in the face-to-face meeting due to time and distance constraints, an additional video conference was organized.

2.1 Delphi expert panel selection

Potential panelists from EU-15 countries were identified based on an extensive search of the international peer-reviewed literature about neuropsychological assessment in patients from minority ethnic groups over the past 5 years. Search terms in several languages were used to make sure all experts were identified. Panelists who published within the last 5 years were selected in order to ensure that the experts were still actively involved in their field. Included panelists were asked to identify any other relevant experts, a technique known as snowballing [195]. Three panelists were included in the final panel based on this technique, two of which were additional experts from the same country who were asked to complement the expertise of the original panelist.

2.2 First Delphi round: Determining current status and collecting qualitative data

We drafted a survey containing seven sections aimed at exploring the current status of the field (see supplementary material). The first section inquired about general panelist data and data on the clinic in which they worked. The second section gathered information about the minority ethnic groups visiting the panelists' clinics. The neuropsychological assessment of these patients was detailed in the third section. The use of interpreter services was the topic of the fourth section. The fifth section discussed training for neuropsychologists, specifically concerning cross-cultural aspects. The sixth section examined the assessment of the following nine cognitive domains in patients from minority ethnic groups: language, memory, working memory, visuospatial functioning, orientation (time/place), attention, mental speed, executive functioning, and social cognition—largely following Lezak et al.'s [98] classifications of cognitive domains. The last section requested the panelists to provide their recommendations for researchers and policy makers. A pilot version of the survey was emailed to two neuropsychologists with ample experience in cross-cultural neuropsychology, after which minor adjustments were made. The final survey was distributed in May 2019.

2.3 Second Delphi round: Rating and ranking priorities

After the data collection for the first survey was complete, the results were integrated into a presentation format and sent by email to all panelists. A second survey was then drafted based on the results of the first survey (see supplementary material). The main aim of the second survey round was to rate and rank various priorities identified in the first survey. The second survey contained four sections. The first section asked panelists to rate the importance of each of the clustered general recommendations generated in the first survey. The second section asked panelists to rate the importance of having an interpreter present for patients with little, some, and a good understanding of the test administrator's language. Furthermore, panelists rated the importance of having a trained interpreter and the importance of having a formal interpreter—as opposed to an informal interpreter. In addition, this section contained an open-ended question on how to improve the use of interpreter services during a neuropsychological assessment. The third section required panelists to rate the importance of training programs to become a neuropsychologist and the importance of training in cross-cultural neuropsychological assessment specifically. It also contained an open-ended question about the skills and knowledge required to carry

out cross-cultural neuropsychological assessments. In the fourth section, panelists rated the importance of more research into the nine different cognitive domains and were asked to provide more detailed recommendations for research on this topic. The survey was distributed in August 2019.

2.4 Third Delphi round: Confirming consensus and generating further recommendations

The third Delphi round was aimed at confirming consensus and discussing final recommendations. This round was split into two parts. One in-person meeting was held with four panelists at the conference of the Federation of European Societies of Neuropsychology in Milan, Italy, on September 7, 2019. An additional video conference was held on October 4, 2019 with six panelists, one of whom also attended the in-person meeting. The results of the first two surveys were sent to all participants before the meeting so they could first independently consider their opinions and relevant comments. Both meetings started with a summary of the main results of the survey rounds. During the meeting, all panelists were given turns to voice their opinions and/or comments. Group discussion of divergent views was encouraged. Panelists were all asked specifically about their opinion on the items that showed consensus in the second survey. Subsequently, all ratings for which wide quartile ranges remained after the second survey round were discussed in depth. All panelists consented to the recording of the meetings.

3 Results

3.1 Delphi expert panel

The number of panelists that participated in each round is displayed in Fig. 1. We could not identify experts from Luxembourg, Sweden, Finland, Greece, Portugal, or Ireland. A total of 16 potential panelists from the remaining EU-15 countries were approached for participation in the survey; three declined participation and two did not respond. One of the invited panelists joined only in the last phase of the study. In the second survey stage, one additional panelist was included. Overall, 12 experts from nine countries contributed to the Delphi study: one from Denmark, Germany, Belgium, England, Italy, and Austria and two from the Netherlands, France, and Spain.

3.2 Round one: Determining current status and collecting qualitative data

For half of the panelists, 5–15% of the patients in their clinic were estimated to be from a minority ethnic group, followed by 15–25% (three out of ten panelists). Some minority ethnic groups were seen in clinics in multiple countries, such as patients from Turkey (five clinics in four countries) and North African minorities (five clinics in four countries), whereas other populations were only seen by panelists from one country, such as South Asian and Afro-Caribbean minorities in the UK, Surinamese and Dutch Antillean minorities in the Netherlands, and Latin American people in Spain. The education level of the patients from minority ethnic groups was generally experienced to be lower compared to that of patients from the majority culture (mostly primary school education or lower in eight out of ten clinics).

Two panelists reported that they did not use any cross-cultural neuropsychological tests or test batteries; the other eight panelists all made use of one or more cross-cultural tests. The

European Cross-Cultural Neuropsychological Test Battery [140] or its subtests were used in four out of nine countries. The Rowland Universal Dementia Assessment Scale [60,196,197] was used in four out of nine countries. The Cross-Cultural Dementia screening [57] was used in two countries. Additionally, the use of some tests seemed to be country-specific. For instance, modified versions of the Mini-Mental State Examination [198-200] were used in the UK; the modified Visual Association Test [201], a literacy screener (unpublished), and the Stick Design Test [112] in the Netherlands; the computerized EMBRACED battery (unpublished) in Spain; the Cross-Linguistic Naming Test [17], the WHO/UCLA adaptation of the Rey Auditory Verbal Learning Test [42], and the Multicultural Cognitive Examination [202] in Denmark; and the TNI-93 [135], TMA-93 [134], TFA-93 [203], and Montreal Cognitive Assessment [204] in France.

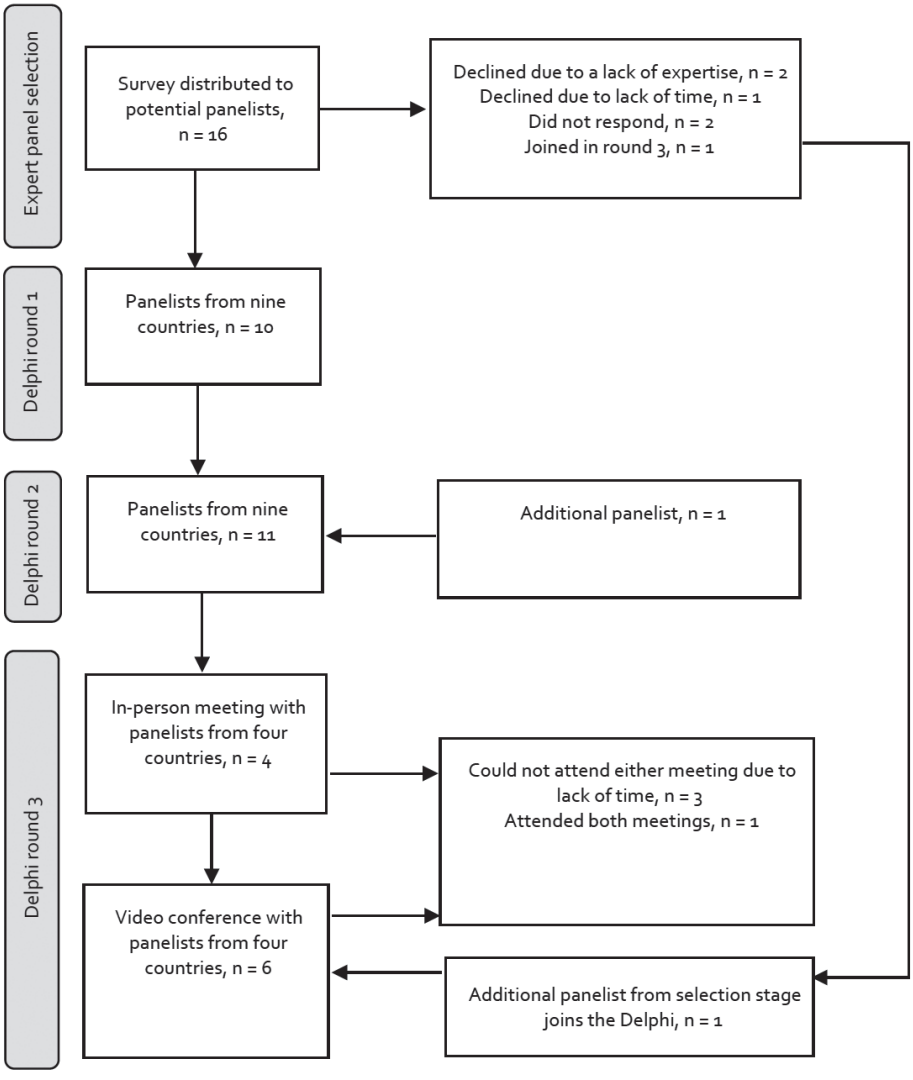


Figure 1. Selection of panelists and participation in each Delphi round.

Three out of ten experts stated that no normative data was available for the tests they used with patients from minority ethnic groups. Five panelists had either normative data for some, but not all tests, or normative data for some, but not all minority ethnic populations. Two panelists noted having normative data for all tests they used. Regarding the person administering the assessment, nine of ten experts reported that, generally, this person was not of the same ethnic background as the patient. To communicate with patients, six panelists used professional interpreters, whereas four panelists did not use interpreters. In Austria and France, interpreters were provided through governmental funding; in Denmark and Belgium, there were different rules depending on the specific case and context; in the other six countries, there were no government-funded interpreter services.

The panelists also rated the degree to which they could assess nine cognitive domains in patients from minority ethnic groups. A 10-point Likert scale was used, ranging from "I cannot assess this cognitive domain at all" (one) to "I can validly and reliably assess this domain and have a sufficient number of tests" (ten). Medians, first quartiles (Q1) and third quartiles (Q3) were extracted. As displayed in Fig. 2A, the domains of social cognition and language were ranked as the most challenging to assess. Various panelists indicated that no tools were available to assess social cognition in patients from minority ethnic groups. Similarly, language was described as hard or even impossible to assess. In contrast, memory was ranked the easiest cognitive domain to assess in these patients.

Clinical training to specialize as a neuropsychologist was available in six out of nine countries. Four panelists described that training in cross-cultural neuropsychological assessment is a part of clinical training, but that it is limited, e.g. voluntary. The panelists from Spain and Denmark described how they, or their clinic, provided their own training on the topic. Only the panelists from France described that cross-cultural assessment was a mandatory part of training in university.

3.3 Round two: Rating and ranking priorities

In round one, a list of general recommendations was generated (see Table 1). These recommendations were then grouped into five broad categories. The first category, "Clinic and staff", contained recommendations such as the employment of ethnically diverse neuropsychologists. The second category, "Training, knowledge, and awareness among neuropsychologists" stressed the importance of cross-cultural knowledge and skills training for neuropsychologists. "Tests", the third category, recommended more research into educational and cultural effects on test performance and for the development of new tests. "Norms", the fourth category, recommended development of normative data that takes into account education, culture, and country of origin, and for tests for which normative data are not required. The last category, "Other", contained a variety of recommendations to improve assessment of patients from minority ethnic groups, such as additional resources in terms of assessment time and interpreter services, and research into specific populations, such as patients with mild cognitive impairment. In the second survey, all panelists were asked to rate the importance of these categories. The category "Tests" was ranked as the most important priority (median: 10, Q1-Q3: 9–10), closely followed by "Norms" (median: 9, Q1-Q3 8–10), and "Training, awareness, and knowledge among neuropsychologists" (median: 9, Q1-Q3 8–10). The recommendations from the "Other" category were ranked as less important (median 7, Q1-Q3: 7–8). Ratings of the

"Clinic and staff" category showed heterogeneous responses, indicating limited consensus (median: 6, Q1-Q3: 3–8).

Figure 2A.

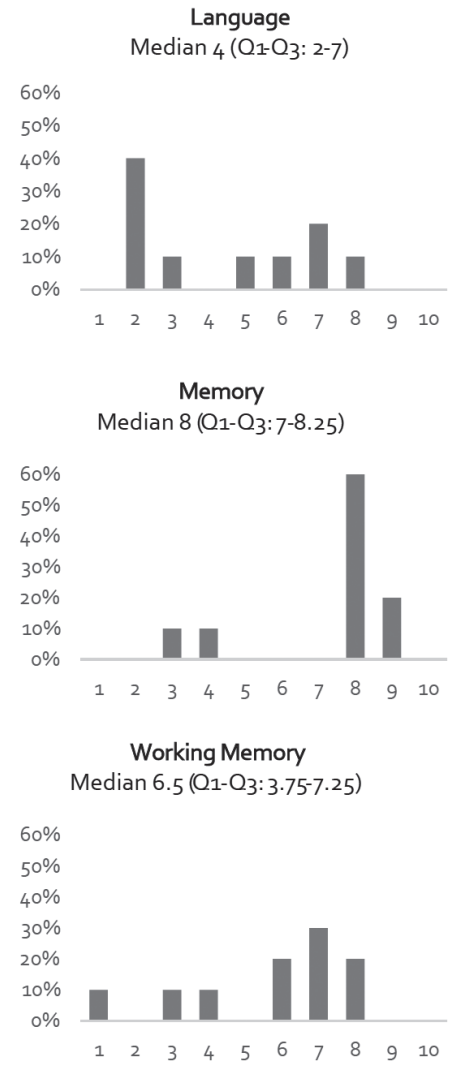


Figure 2B.

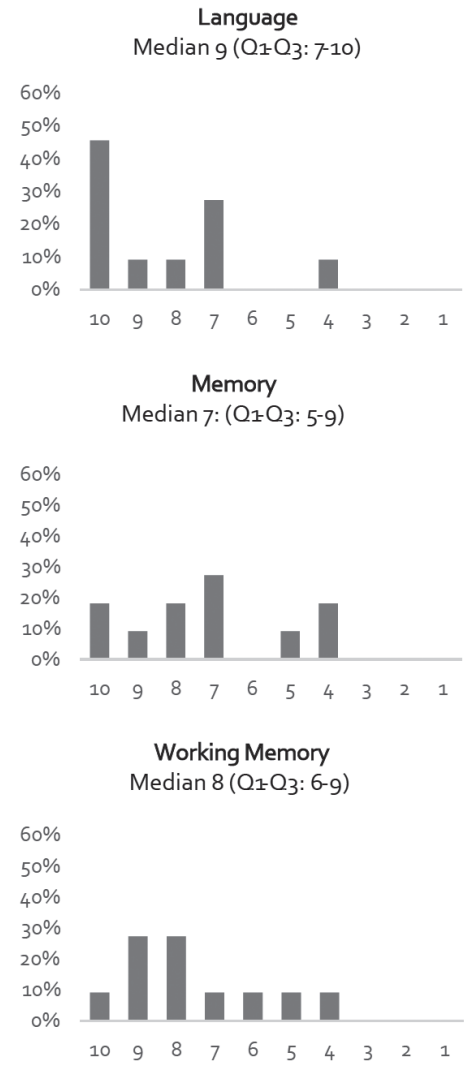
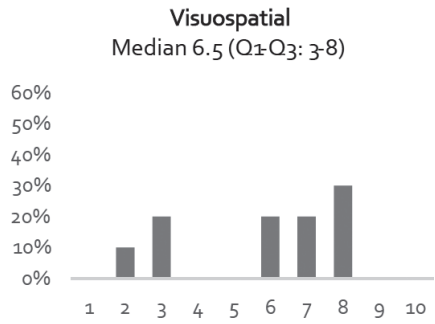
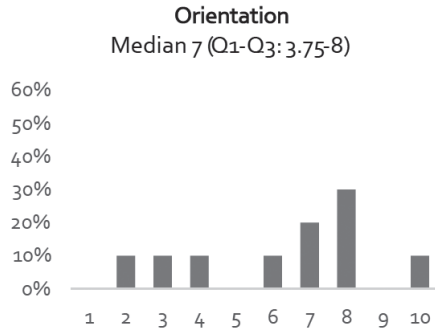


Figure 2A Ease of administration ratings from the first round (% of panelists; medians, first quartile [Q1] and third quartile [Q3]). 2B Importance ratings from the second round (% of panelists, first quartile [Q1] and third quartile [Q3]).Plotted on a reverse axis for ease of comparison with the results from the first round.

Figure 2A. (cont.)





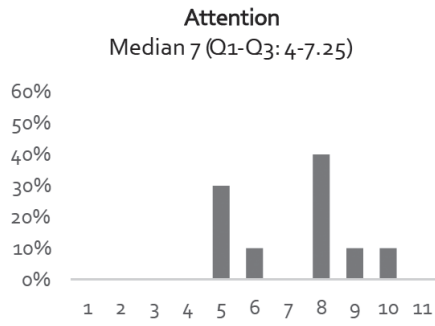
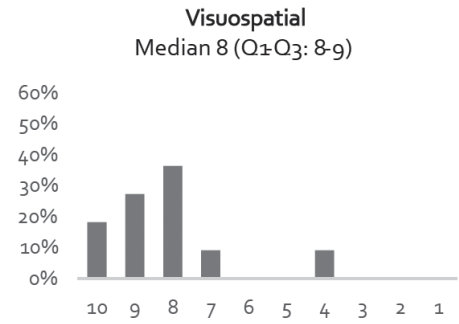
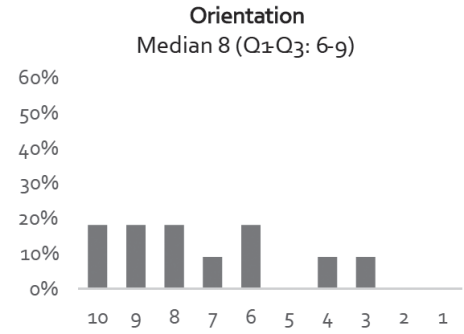
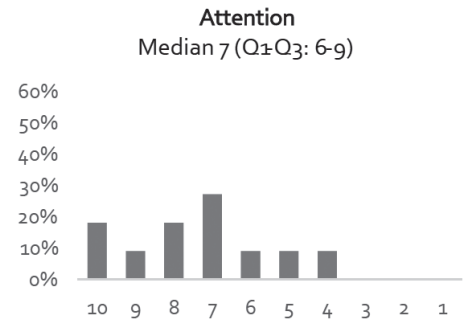


Figure 2B. (cont.)*

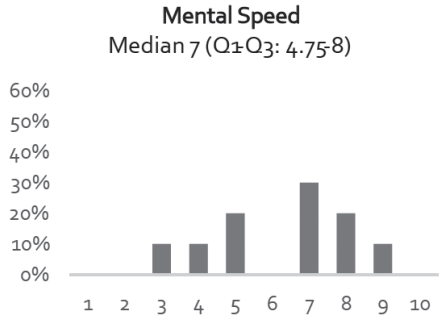


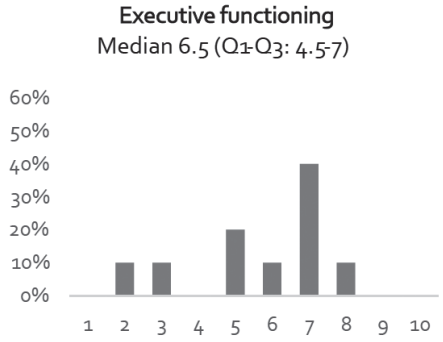




*Plotted on a reverse axis for ease of comparison with the results from the first round.

Figure 2A. (cont.)





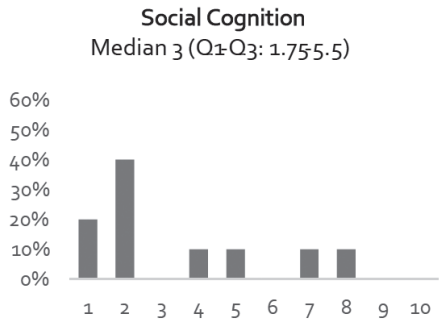
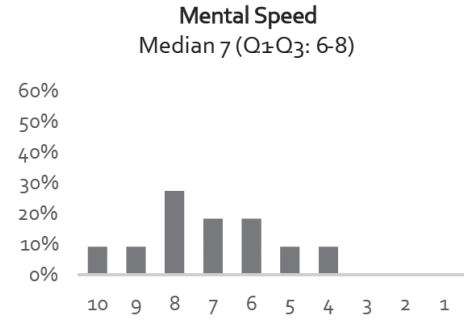
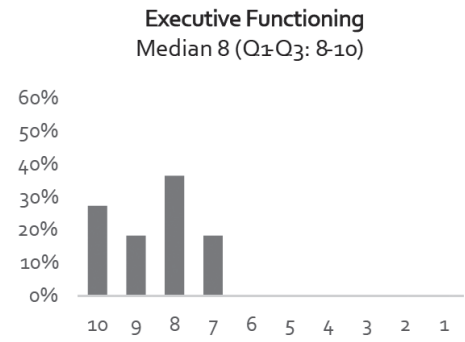
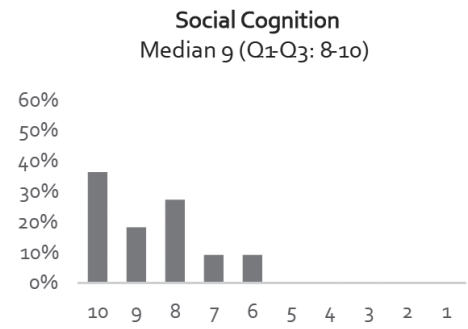


Figure 2B. (cont.)*







*Plotted on a reverse axis for ease of comparison with the results from the first round.

Table 1. General recommendations

Clinic and staff	Training, knowledge, & awareness among neuropsychologists	Tests	Norms	Other
Employment of neuropsychologists with a diverse ethnic background	Professional training in cross-cultural neuropsychology for neuropsychologists, perhaps as a fixed part of specialist training for clinical neuropsychologists, including training on how to work with interpreters	More research on the way education and culture influence neuropsychological test performance for different cognitive domains, so we are better able to adapt neuropsychological tests and/or develop new tests	Development of normative data, specifically those that take into account education, culture, and/or country of origin	More resources needed, such as: interpreters that are trained to work with neuropsychologists, funded by the government; more time to carry out the assessment
Employment of native speakers of minority ethnic languages or neuropsychologists with proficiency in different languages	Increased awareness among neuropsychologists of the influence of factors such as diversity, education, and culture on clinical practice	Develop and/or validate (fine-grained) neuropsychological tests, especially for cognitive domains for which there are few tests.	Development of tests for which culture- or education-specific normative data is not required	More research with regard to specific populations, such as rare diagnoses, mild cognitive impairment, oldest old patients, semi-literate/illiterate patients
Providing services in a specialized, expertise setting, e.g. memory clinic with focus on minorities	Increase in knowledge of cultural aspects			Examining ways to instruct/inform patients from minority ethnic groups about the neuropsychological assessment (test-wiseness) Using additional biomarkers in diagnostic assessment

Table 2. Recommendations from all Delphi rounds for research in each cognitive domain

	Social cognition	Language	Executive functioning	Mental speed	Attention	Visuospatial	Memory	Working Memory	Orientation	Other
Suggestions regarding tests	Emotion recognition	Good naming test suitable for various minorities, with adequate normative data, brief, efficient, sufficiently sensitive and specific, with little need of explanations in administration	Tests of planning, mental flexibility, inhibition and other (sub) functions of executive functioning	Tests less influenced by education and literacy	Tests less influenced by education and literacy	Perhaps use a "format that everyone comes across in life, e.g. a broken cup?"	Tests for episodic memory			Intelligence tests measuring aspects similar to reasoning or matrices scales
	Test capturing social norms and moral/ social dilemmas	A test for the evaluation of speech, such as a complex picture description	Tests less influenced by education and literacy			Tests less influenced by education and literacy	Verbal and nonverbal			Test for suboptimal performance (malingering)
	Theory of mind		More tests for very low educated populations			More alternatives	Normative data for other minorities, e.g. Eastern Europeans, and for higher educated patients			Test for motor coordination

Table 2. Continued.

	Social cognition	Language	Executive functioning	Mental speed	Attention	Visuospatial	Memory	Working Memory	Orientation	Other
Other comments	Needs more research in general	"Will be hard to assess"	Variation between cultures and between individuals	Cultural norms and expectations influence test results, i.e. speed is a cultural norm		"Lack of alternatives to current tests […] Not sure which ones are best for different ethnic groups".		"There are various options for working memory tests, but they have not been validated yet"	Move beyond MMSE orientation questions	
	"Seriously lacking"; "huge need to develop them"	"Some language barriers are unbridgeable for an assessor that does not speak the participant's language"	Need to understand how education modulates executive functioning in healthy population before examining culture			"I think we need to develop tests other than drawings or figure copy that required past school learning of geometry"		"We need to come up with a new testing paradigm with better ecological validity that does not rely on numbers"		
	Should (also) be studied with more qualitative methods	"Most apparent barrier"								
	Highly dependent on culture									

Table 3. Recommendations for knowledge and skills required by neuropsychologists

Knowledge about patients from minority ethnic groups	Culture, education, language and the assessment	Availability of cross-cultural tests	During/after the assessment	Recognize limitations	Reflect on own culture	Training staff
Knowledge about the cultural/ linguistic background of main minority groups; background knowledge about the country of origin, migration history, customs, expectations, social roles and rules, religion, and traditions; knowledge about differences between cultures.	Knowledge about how language, culture, and education influence test performance; knowing how to interpret test performance, e.g. what is normal performance for someone without education.	Knowledge of which cross-cultural tests and batteries are available, perhaps with the help of test publishers?	Get used to working with formal and informal interpreters; know how to instruct them.	Recognize the limits of instruments and normative data.	Reflect on own culture; be aware of own prejudices and cultural beliefs; be aware of how others may perceive you	Be trained as a neuropsychologist before practicing neuropsychology
Knowledge about (cultural) views on cognition and cognitive impairment.	Be aware of the cultural values underlying cognitive assessment		Be flexible, patient, listen well and be creative to work around barriers	Recognize the limits of own expertise and knowledge	Have an open mind	Training of test administrators by neuropsychologists
	Be aware of the necessity of cross-cultural tests and normative data		Know how to handle cultural barriers			
	Knowledge of potential biases of instruments and assessment procedures; knowing which tests are culture-specific.		Being able to do a broad (history taking) interview and use questions to assess cognitive functioning in an ecologically valid way			
	Knowledge about the influence of culture on the brain		Being able to qualitatively interpret results			

Table 4. Recommendations for working with interpreters

1. All neuropsychologists should be trained in working with interpreters to improve the outcome of interpreted assessments
2. Neuropsychologists need to know about existing guidelines for working with interpreters—perhaps these should be included in standard (clinical) training
3. Interpreters should be trained
4. Standard practice before the neuropsychological assessment should entail a briefing with the interpreter about:
   - the aims of the assessment/what the neuropsychologist wants to accomplish
   - the case
   - the instruments
   - the procedures (of the assessment), in particular:
     - failing of the patient is a vital part of the examination
     - interpreters should translate as literally as possible
     - only to intervene/correct if the neuropsychologists says so, i.e. not give any hints, additional information etc.
   - the expected responses
5. Hire interpreters with experience/training in clinical settings, perhaps even create a whitelist (or blacklist) of interpreters
6. Adapted tests are often translated/back-translated multiple times; have interpreters use the official translated instructions instead of interpreting freely based on what the neuropsychologist says
7. Awareness of regional or country-specific variations in language, e.g. it is not ideal to use a Spanish interpreter from Spain for the assessment of Latin American patients
8. Improve availability (funding)

After having been presented with the results of the first survey, the panelists were asked to rate the need of further development of tests within the nine cognitive domains. Figure 2B displays the results, plotted on a reverse axis to facilitate the comparison with Fig. 2A as the answers were formulated in reverse directions for survey 1 and 2—i.e. the degree to which domains could be assessed versus the need for more research. Except for memory, the quartile ranges for all cognitive domains were smaller in the second than the first round, indicating a shift towards consensus. After the second round, the priorities of the cognitive domains could be grouped in three levels. Social cognition and language were ranked as most important for further development. The second most important set of cognitive domains was executive functioning, visuospatial functioning, working memory, and orientation. The domains of attention, mental speed, and memory were indicated to need the least amount of research. Specific recommendations from round one and two for the development of cognitive tests are provided in Table 2. Furthermore, the panelists made recommendations regarding the specific knowledge and skills that neuropsychologist needs to perform cross-cultural assessment (Table 3).

Regarding interpreters, panelists indicated it was critical to have an interpreter present for the assessment of patients with little understanding of the test administrator's language (median: 9, Q1-Q3: 7–10). There was little consensus about the use of interpreters for patients with some understanding of the language (median: 7, Q1-Q3: 5–9) and for patients with a good understanding of the language (median: 5,Q1-Q3: 2–7). Having a formal interpreter present, as opposed to an informal interpreter such as a relative, was rated as important (median: 8, Q1-Q3: 6–9). Furthermore, having an interpreter present who is trained at interpreting during a neuropsychological assessment was rated as important (median: 8, Q1-Q3: 7–9). In the open-ended questions of the second survey, a list of recommendations to improve assessment with interpreters was generated (Table 4).

3.4 Round three: Confirming consensus and additional recommendations

During the face-to-face meeting, the panelists first reached a consensus on the importance of improving or developing cross-cultural tests and normative data and on increasing cross-cultural knowledge and training among neuropsychologists. Subsequently, the lack of consensus about the category "Clinic and staff" was discussed. This lack of consensus was suggested to be due to the limited feasibility of providing same-ethnicity staff in clinics with patients from a wide variety of ethnic backgrounds.

Second, the panelists agreed on the order of importance of development of cross-cultural cognitive tests for the nine cognitive domains. According to the panelists, heterogeneity in ratings on memory tests was caused by differences in the availability of cross-cultural memory tests across EU-15 countries. The results from the second survey indicated that each cognitive domain was assigned either a high, medium or low(er) priority. According to one of the panelists, this three-level hierarchy may reflect differences in conceptual complexity of these cognitive domains: memory and processing speed may be less challenging to capture in a cross-cultural cognitive test than concepts like executive functioning or social cognition. However, the question was raised whether it is possible to develop tests of social cognition and language that are cross-culturally and cross-linguistically applicable, or whether tests should be developed for each individual minority ethnic group. The practicality of this last approach was judged to be limited. Although some large minority groups may be present across various EU-15 countries (e.g. the Turkish minority), the patient population in most clinics is too diverse to have any use for tests that are specific to any individual minority culture or language.

Third, as indicated by the surveys, the panelists agreed on the need for an interpreter for patients with little understanding of the host country language, but no consensus was reached about the need for an interpreter for patients with some or a good understanding of the host country language. One panelist suggested that this probably depends on the level of precision needed for an assessment—an interpreter will be necessary to identify mild cognitive deficits, but may not be necessary to identify more severe cognitive impairment. Although the panelists recognized that there are inherent challenges associated with doing neuropsychological assessments with an interpreter, such as the risk of biased test results, assessment with an interpreter was often necessary as the availability of (same-ethnicity) neuropsychologists fluent in the patient's language was very limited in the panelists' countries. Although one panelist mentioned that patients may feel more at ease with a relative doing the interpretation, the panelists agreed that, generally, the use of formal interpreters was preferable to the use of informal interpreters. The panelists voiced their concerns regarding the ethical aspects of the quality of the interpretation and the potential bias introduced when using informal interpreters. According to the panelists, when assessing a patient with a formal interpreter, neuropsychologists need to be aware of various potential barriers. First, patients may feel ashamed about the fact that they are low educated and speak local, rural dialects, rather than speaking the more formal language of the interpreter. A second potential barrier mentioned by the panelists is a mismatch between the gender of the interpreter and the patient. A third issue is the variable quality of formal interpreter services in some countries, where interpreters do not always have "proven efficacy" and may only work as an interpreter for a short time. This is in contrast with some other EU-15 countries, where formal interpreters have to meet various criteria

and follow a clear code of conduct. Finally, the use of telephone interpreters was identified as a challenge. The panelists agreed that evaluation with a telephone interpreter should be avoided, mostly because the interpreter cannot see the test materials that are used in the assessment.

Fourth, the panelists reached a consensus that (cross-cultural) training was important. It was recommended that this training should include both theoretical and practical training in cross-cultural assessment and working with diverse patients. No consensus was reached about how this training was best implemented as there were significant cross-country variations in neuropsychological training, certification, and licensing in general. Furthermore, the way expertise in cross-cultural neuropsychology was organized varied by country: some countries had expert centers, e.g. a specific multicultural memory clinic, whereas others had more "local expertise". Only in France was a more extensive cross-cultural training provided to neuropsychologists.

In addition to reaching a consensus on these topics, a number of other relevant aspects were mentioned regarding the assessment of minority ethnic populations. One panelist commented that neuropsychologists should be aware of the effect of examiner–examinee ethnic discordance, as well as mentioning the possible effects of stereotype threat on cognitive test performance. Another panelist mentioned the inter- and intra-individual variability in the proficiency in, and use of the majority and minority languages, and the prestige that can sometimes be attached to proficiency in certain languages. Furthermore, two panelists mentioned that people from a different culture will not be familiar with undergoing a neuropsychological assessment, possibly inducing shame or (di)stress in patients, or making the patient feel treated like a child. During the meeting in Milan with panelists from Denmark, Italy, the UK, and the Netherlands, the costs of and access to neuropsychological services were also discussed. The panelists indicated that the assessments were either free or were covered by (mandatory) health insurance. In some countries, the availability of specialized services for minority patients depends heavily on whether patients live in the catchment area. Last, two of the panelists mentioned that, in patients from minority populations, it is important to take a wide range of variables into account: culture, age, gender, education, and lifetime (socio)demographic characteristics.

4 Discussion

The aims of this Delphi study were to examine the current state of the field of adult cross-cultural neuropsychological assessment in EU-15 countries and to generate recommendations for researchers and policy makers. The results showed that a number of instruments and batteries are available in EU-15 countries—in particular cross-cultural memory tests—several of which are currently used in more than one country. A consensus was reached that training of neuropsychologists and the development of cross-cultural tests and normative data are the most pressing matters. A consensus was reached on social cognition and language tests as the first priorities, followed by tests of executive functioning, visuospatial functioning, working memory, and orientation. The panelists agreed that tests that can be used across a variety of minority ethnic groups are preferable over tests specific to one culture or language. The panelists recognized that the use of formal interpreters is important, although neuropsychological assessment with interpreters may never be free

of bias. Various recommendations were provided for working with interpreters and for training in cross-cultural neuropsychological assessment.

This study showed that considerable work has already been carried out in the development and validation of cross-cultural neuropsychological tests in Europe. In particular, the European CNTB and the RUDAS are well validated; these instruments have been studied in people from numerous minority groups, with a wide variety of education levels, in studies from across multiple European countries (CNTB [21,140,205]; RUDAS [196,197,206]). Together, these instruments measure a variety of cognitive functions: general cognitive functioning (RUDAS), memory (Enhanced Cued Recall and Recall of Pictures Test), language (Picture naming and semantic verbal fluency), executive functions (Color Trails Test and Five Digit Test), and visuospatial functions (Clock Reading Test, Clock Drawing Test and copying of simple and complex figures). For some of the other instruments identified in this study, few (if any) validation studies in the target population have been published. Most experts reported using one or more cross-cultural (adapted) cognitive test, but few panelists were familiar with all the tests that were used by the other experts. This highlights that existing tests validated in one country should be better publicized, reviewed, and implemented in other EU countries. This will, at a minimum, require carrying out local validation studies following international standards, such as those of the International Test Commission [167], as well as negotiating with publishers. The development of new cross-cultural tests and normative data was rated as highly important. Merely stratifying normative data by age and education may be insufficient for low educated patients from minority ethnic groups, who often show floor effects on neuropsychological tests requiring any form of school-based procedures [205]. The development of new neuropsychological tests is therefore warranted. To suit a diverse patient population, the international literature recommends designing tests without black-and-white line drawings [28,29,201], culture-specific stimuli [12,17,178], or test elements that require skills learned in school [32]. Additionally, tests that are more ecologically valid may be more suitable for this population [178,184]. As developing tests and collecting normative data can be a costly and time-consuming process, researchers will have to prioritize which cognitive domains to tackle first. The experts in this study particularly agreed on a general lack of tests measuring social cognition and language, as opposed to some of the other cognitive domains for which more tests are available. Memory was the domain that panelists considered to be lacking the least in terms of test development. This finding reflects the better availability of cross-culturally validated memory tests in EU-15 countries, which is probably due to the relative ease with which memory tests can be developed or adapted to suit minority ethnic groups, e.g. by using items that are common and familiar to the minority ethnic group and by presenting them in a suitable format [207]. Adequately validated tests and normative data for the cognitive domain of social cognition generally seem to be lacking in the EU-15, even for native-born adults [208,209]. Aside from social cognition and language, one panelist also suggested to validate or develop performance validity tests. A similar call to action was made at the Sixth European Conference on Symptom Validity Assessment in 2019, stressing that the cross-cultural validity of current performance validity tests is probably limited [210]. Panelists from two countries in this study were working on, or had previously worked on, cross-cultural validation studies of performance validity tests [211]. However, no true experts seem to exist in the EU that specialize specifically in the topic

of performance validity testing in minority ethnic populations in Europe—research on this topic currently seems to be dominated by work from other regions [212-214].

Aside from looking into cognitive tests themselves, it is important to take the cultural context of neuropsychological assessment into consideration. These contextual factors are elegantly summarized by the acronym of the ECLECTIC framework [13]: Education and literacy, Culture and acculturation, Language, Economics (e.g. socioeconomic status), Communication, Testing situation, comfort and motivation, Intelligence conceptualization, and Context of immigration. Although the (design of the) current study mainly highlighted the importance of cognitive tests and norms, several key contextual factors were mentioned by the panelists. For example, an unpublished literacy screening test was used in one country to determine the quality of the patients' education (E). Neuropsychologists from a number of European countries make use of short acculturation scales (C) in their clinics—such as a modified version of the Short Acculturation Scale for Hispanics (SASH [215]). Additionally, the effects of language abilities in both native and host country languages (L) was mentioned, as well as the effects of stereotype threat [216], of being unfamiliar with cognitive testing, and of examinee-examiner ethnic discordance (T) on the assessment. The panelists also mentioned it is important to take into account lifetime (socio)demographic factors and access to and availability of health services (E). Some aspects from the ECLECTIC framework, in particular communication styles and intelligence conceptualization, received less explicit attention in this study. This may in part be due to the way the surveys were designed (i.e. with a relatively heavy emphasis in the forced-choice questions on cognitive tests). Other specific examples of relevant issues to take into consideration in working with minority ethnic groups are traumatic experiences and migration-related distress or grief [217], differences in explanatory models of illness [54,218], exposure to discrimination [219], and differences in symptom manifestation and idioms of distress, such as mixed affective and somatic presentations of depression in Moroccan and Turkish patients [220].

Cross-cultural neuropsychological assessment could benefit from matching patients from minority ethnic groups with same-ethnicity neuropsychologists. The experts in this study agreed that providing same-ethnicity neuropsychologists to all patients from minority ethnic groups in the EU-15 countries is currently not feasible considering the number of different minority ethnic groups and the limited ethnic diversity among neuropsychologist in the EU, which is in line with the reality in the USA [221]. Instead, the panelists identified more cross-cultural training, awareness, and knowledge among neuropsychologists as an important need for cross-cultural assessment. Cross-cultural training of neuropsychologists was also identified as a priority in the USA, where "clinicians often lack in-depth training in assessment of ethnic minorities" [222]. In the present study, a list of important knowledge and skills was generated for training in cross-cultural assessment. These recommendations can be supplemented with existing guidelines, such as those captured in the ECLECTIC framework [13]. With regard to the implementation of these recommendations, a recent study indicated that training to become a neuropsychologist is organized differently across the EU, and the duration of training varies substantially between 12 and 60 months [223]. The way cross-cultural skills are incorporated in neuropsychology training may thus have to be decided separately for each country. Alternatively, cross-cultural training could be realized by organizing a European summer school in cross-cultural neuropsychology, e.g.

in collaboration with the Federation of European Societies of Neuropsychology (FESN) or the European Federation of Psychologists' Associations (EFPA).

Concerning the use of interpreters, the panelists generally agreed that having a formal interpreter present—as opposed to an informal interpreter, such as a family member—was important. However, the panelists also agreed that various challenges will remain, even with a formal interpreter present. Previous studies have similarly indicated that working with interpreters carries risks. The use of relatives as interpreters has been related to the exclusion of the patient from the conversation [35], problems with the adequate translation of medical terminology [36], obscuring of the patient's explanatory models, and difficulties in assessing the level of insight [37]. The use of formal interpreters may be challenging as well, especially for tests with high demands on the abilities of the interpreter or when interpreters have received little formal training [38]. The use of telephone interpreter was discouraged by the panelists as the interpreters would be unable to see the test materials. Additionally, assessment with a telephone interpreter can be hindered by factors such as disturbances in communication due to background noise [224]. We believe that EU guidelines for working with interpreters in the neuropsychological assessment of patients from minority ethnic groups are needed and that these could be extensions of existing guidelines, such as those of the British Psychological Society [225].

Some limitations to this study should be acknowledged. First, the total number of experts that could be identified (12) was relatively small—a typical Delphi study will have between 10 and 50 panelists [226]—and a total of six EU-15 countries were not represented in the panel—Finland, Greece, Ireland, Luxembourg, Portugal, and Sweden. This finding seems to indicate that the field of cross-cultural neuropsychological assessment is largely still a developing field in the EU-15, and formal expertise is localized, rather than widespread. Additionally, the panelists identified using the criteria in the search strategy were nearly invariably of a majority background. Additional research is needed to determine whether the main findings of this study are endorsed by neuropsychologists with a minority ethnic background. This might be accomplished by broadening the inclusion criteria in a follow-up study, such as by replacing the publication criterion with other indicators of expertise, e.g. by peer nomination, self-report, or based on having assessed a specific number of minority ethnic patients. Second, it would have been preferable if all panelists could have participated in one final face-to-face round, which was not possible due to time and distance constraints. By splitting the third round in two meetings, a risk of bias may have been introduced, as smaller groups tend to be more vulnerable to individual panelists holding strong opinions. However, we estimate that these effects were probably minimal, given that a) all panelists received the survey results before the meeting, so they could independently form their opinions, b) all panelists were given turns to speak, c) group discussion of divergent views was encouraged during the meetings, and d) panelists did not reach a consensus on all topics, indicating group pressure to conform was probably negligible. Another limitation of the study was that the majority of the experts worked in a memory clinic setting—although several of them also had experience with assessment of either healthy people from minority ethnic groups or patients from minority ethnic groups in other settings than memory clinics. The overrepresentation of memory clinic experts may partly be influenced by the snowballing technique used in the study, but could also reflect the predominant focus on dementia research in the EU, possibly due to dementia's

large economic and societal costs [227]. A last limitation of this study is that no specific metric intervals were determined to define consensus at the outset of the study.

Given the aforementioned limitations, the results from this study should be seen as a first step towards the development of new policies. More research is needed to ensure that minority groups are represented and their opinions heard. We suggest broadening the scope of this study to represent more neuropsychologists with a minority background, as well as non-expert neuropsychologists, cultural psychologists, and community stakeholders to bring to light all relevant needs and perspectives. Furthermore, the population of interest should be expanded to include immigrants in the wider EU, transnational European minorities, such as Roma people across Europe, and second and third-generation descendants of immigrants. The second generation is more often bilingual and higher educated than the first generation of immigrants, although notable heterogeneity within this group exists—for example, second-generation Turks more often lag behind on Dutch language fluency and are often lower educated than their Moroccans peers in the Netherlands [228]. It will be a challenge to determine which tests and normative data will be most appropriate for this heterogeneous population.

In conclusion, this study indicates that significant work has been carried out in the development and validation of cross-cultural neuropsychological tests in Europe. However, despite recent advances in cross-cultural neuropsychological testing and training in some EU-15 countries, this Delphi expert study highlights the continuing need for development of cross-cultural tests and normative data as well as culture-sensitive training, awareness and knowledge among European neuropsychologists. To improve the field of cross-cultural neuropsychology across the EU-15, countries should increase collaboration—both within the EU and with neuropsychologists from the countries of origin of minority ethnic patients—to a) exchange ideas and methods for cross-cultural neuropsychological assessment, b) validate tests and collect normative data, and c) collaborate in training approaches and the development of guidelines for working with interpreters.

Acknowledgments and funding
The authors thank the following panelists for their invaluable contributions to this study: Ulrike Beinhoff, Leentje Flour, Miriam Goudsmit, Inmaculada Ibanez-Casas, Didier Maillet, Naaheed Mukadam, Pauline Narme, Miguel Perez-Garcia, Simone Pomati, and Stefan Strotzka. This work was supported by The Netherlands Organisation for Health Research and Development (ZonMw Memorabel) [grant number 733050834 to SF, JMP, EvdB].

Conflict of Interest
The authors report no conflict of interest.

Chapter 2.3 Supplementary material
Delphi Survey 1 – Cross-Cultural Neuropsychological Assessment in Europe

This survey is an initiative by Sanne Franzen and Janne Papma from the Alzheimer Center Erasmus MC, Rotterdam, the Netherlands, and Rune Nielsen from the Danish Dementia Research Centre, Copenhagen, Denmark.

As ethnic minority populations across Europe are aging rapidly, and age-related cognitive diseases become more prevalent, the availability of cross-cultural neuropsychological instruments, use of interpreters, adequate normative data, and professional training of neuropsychologists in cross-cultural professional skills, have become pressing issues in neuropsychology. The first objective of this survey is therefore to determine the current status of the field of cross-cultural neuropsychological assessment in Europe. The second objective is to generate recommendations for researchers and policy makers on the issues that should be addressed first and to provide ideas on ways to resolve these issues. To reach these goals, we will consult with European experts in the field of cross-cultural neuropsychology in a Delphi study. We would like to invite you as one of these experts and kindly ask you to fill out this survey. If you have any questions or know other researchers in your country who are experts in this field, please write a comment in the last question, or contact us by email.

General information
1. Name:
2. Job title:
3. Institution, department:
4. If you work in a specialized clinic or (research) center, please specify the name here:
5. City, country:
6. E-mail address:

Ethnic minority groups
In this study, ethnic minority patients are defined as persons who are first-generation immigrants or refugees from countries outside the extended EU, Canada, USA, Australia and New Zealand.

7. What is the percentage of ethnic minority patients in your clinic (if unavailable, please provide your best estimate)?
   - ☐ <5%
   - ☐ 5%–15%
   - ☐ 15–25%
   - ☐ 25–35%
   - ☐ 35–50%
   - ☐ >50%

8. What are the largest ethnic minority groups in your clinic? Please note from largest to smallest (with percentages, if available):

9. What is the most common education level of the ethnic minority patients in your clinic:
   - ☐ 1. No education or illiterate
   - ☐ 2. Less than primary school education
   - ☐ 3. Primary school education
   - ☐ 4. Lower secondary education (e.g. lower/junior secondary school, middle school; often compulsory)
   - ☐ 5. Higher secondary education (e.g. higher secondary school, high school)
   - ☐ 6. Tertiary education (e.g. bachelor's/master's or higher)

10. If available, provide the percentage of patients with each respective education level, i.e. 1: __%, 2: __%, 3: __%, 4:__%, 5:__%, 6:__%:

Neuropsychological assessment

11. Do you make use of cross-cultural (adapted) neuropsychological tests or test batteries for ethnic minorities in your country?
   - ☐ No
   - ☐ Yes (please specify below)

12. If yes, the neuropsychological tests/test batteries that are used are:

13. Are there norms available that are specific to the minority groups you work with?
   - ☐ No
   - ☐ Yes, for all tests
   - ☐ Yes, but only for some tests (please specify below)

14. If such norms are only available for some tests, please specify below for which ones:

15. In general, is the person administering the neuropsychological assessment of the same ethnic background as the minority patient?
   - ☐ No
   - ☐ Yes

Interpreter services

16. Do you make use of interpreter services for cross-cultural neuropsychological assessments?
   - ☐ No
   - ☐ Yes, live professional interpreters
   - ☐ Yes, via phone/video
   - ☐ Other, …

17. Does your government provide reimbursement for the use of interpreters (if only in some cases, please use 'Other' and specify)?
   - ☐ No
   - ☐ Yes
   - ☐ Other, …

Training

18. As far as you are aware, does a professional training program exist in your country to qualify as a neuropsychologist (e.g. at undergraduate/postgraduate level, a BSc/MSc/MMed/post-MSc or post-MMed)?
    - ☐ No
    - ☐ Yes (please specify below)

19. If yes, please specify your answer:

20. As far as you are aware, is cross-cultural neuropsychological assessment part of the professional training of the test administrator (e.g. at undergraduate/postgraduate level, BSc/MSc/MMed/post-MSc or post-MMed)?
    - ☐ No
    - ☐ Yes (please specify below)

21. If yes, please specify your answer:

Future directions in cross-cultural neuropsychology

22. In your professional opinion, what is required to improve assessment of cognition in ethnic minority groups? E.g. concerning neuropsychological assessment methods, use of interpreters, professional training, etc.

Below is a list of cognitive domains. Please rate how well you can assess this cognitive domain in the ethnic minority population in your clinic on a scale of 1 to 10, with 1 = "I cannot assess this cognitive domain at all", and 10 = "I can assess this cognitive domain in a valid and reliable way and have a sufficient number of tests at my disposal".

23. Language

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

24. Memory

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

25. Working memory

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

26. Visuospatial

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

27. Orientation (time/place)

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

28. Attention

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

29. Mental speed

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

30. Executive functioning

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

31. Social cognition

	1	2	3	4	5	6	7	8	9	10	
I cannot assess this cognitive domain at all	o	o	o	o	o	o	o	o	o	o	I can validly and reliably assess this domain and have a sufficient number of tests

32. For all the cognitive domains you scored lower than 7/10, how could the assessment of this domain be improved through changes in clinical practice, research or policy, if there were no constraints in terms of budget, time etc.? Please write the name of the domain followed by your suggestions for each domain.

End of the survey

33. If you have any other relevant comments, please specify them here:

Delphi Survey 2 – Follow-up Survey Cross-Cultural Neuropsychological Assessment in Europe

This is a follow-up survey to the first survey about cross-cultural neuropsychological assessment in Europe. It contains mostly multiple choice questions and should not take up a lot of your time. Please make sure you have looked at the results of Survey 1 (pdf document sent to you by email) before continuing the survey. If you have any questions, please write a comment in the last question, or contact us by email.

1. Name of the participant:

General recommendations

The following questions concern the general recommendations following from Survey 1 (page 11 of the pdf document). Please indicate below how important these general recommendations are for improving cross-cultural neuropsychological assessment.

2. How important are 'Changes in the clinic/staff'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

3. How important is 'More training, awareness and knowledge among neuropsychologists'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

4. How important is 'Development/validation of neuropsychological tests'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

5. How important is 'Development of (extensive norms) for existing tests'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

6. How important are the recommendations from the 'Other' category (more resources, trained interpreters, research in specific subpopulations, better instructions/ information about NPA for ethnic minority patients, more biomarkers)?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

Interpreters

7. How important is it to have an interpreter present in the assessment of ethnic minority patients who have little understanding of the language of the test administrator?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

8. How important is it to have an interpreter present in the assessment of ethnic minority patients who have some understanding of the language of the test administrator?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

9. How important is it to have an interpreter present in the assessment of ethnic minority patients who have a good understanding of the language of the test administrator?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

10. How important is it to have a formal interpreter present in the assessment of ethnic minority patients (as opposed to an informal interpreter, such as a relative)?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

11. How important is it to have an interpreter present who is trained in interpreting during neuropsychological assessments?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

12. Do you have any specific suggestions how the use of interpreters during neuropsychological assessments can be improved?

Training of neuropsychologists

13. How important is it that a professional training program exists for psychologists to qualify as a neuropsychologist?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

14. How important is it to train test administrators in cross-cultural assessment as part of their general training?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

15. Based on your expertise, what do neuropsychologists need to know and what skills do they need to learn before they are able to do a cross-cultural neuropsychological assessment? Please be specific and/or provide examples.

Assessment of cognition
The next questions concern cross-cultural cognitive assessment (page 12 of the pdf document)

16. How important is more research in the cognitive domain of 'Social Cognition'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

17. How important is more research in the cognitive domain 'Language'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

18. How important is more research in the cognitive domain 'Executive Functioning'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

19. How important is more research in the cognitive domain 'Working Memory'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

20. How important is more research in the cognitive domain 'Visuospatial Functioning'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

21. How important is more research in the cognitive domain of 'Attention'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

22. How important is more research in the cognitive domain of "Mental Speed'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

23. How important is more research in the cognitive domain of 'Orientation'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

24. How important is more research in the cognitive domain of 'Memory'?

	1	2	3	4	5	6	7	8	9	10	
Not important	o	o	o	o	o	o	o	o	o	o	Important

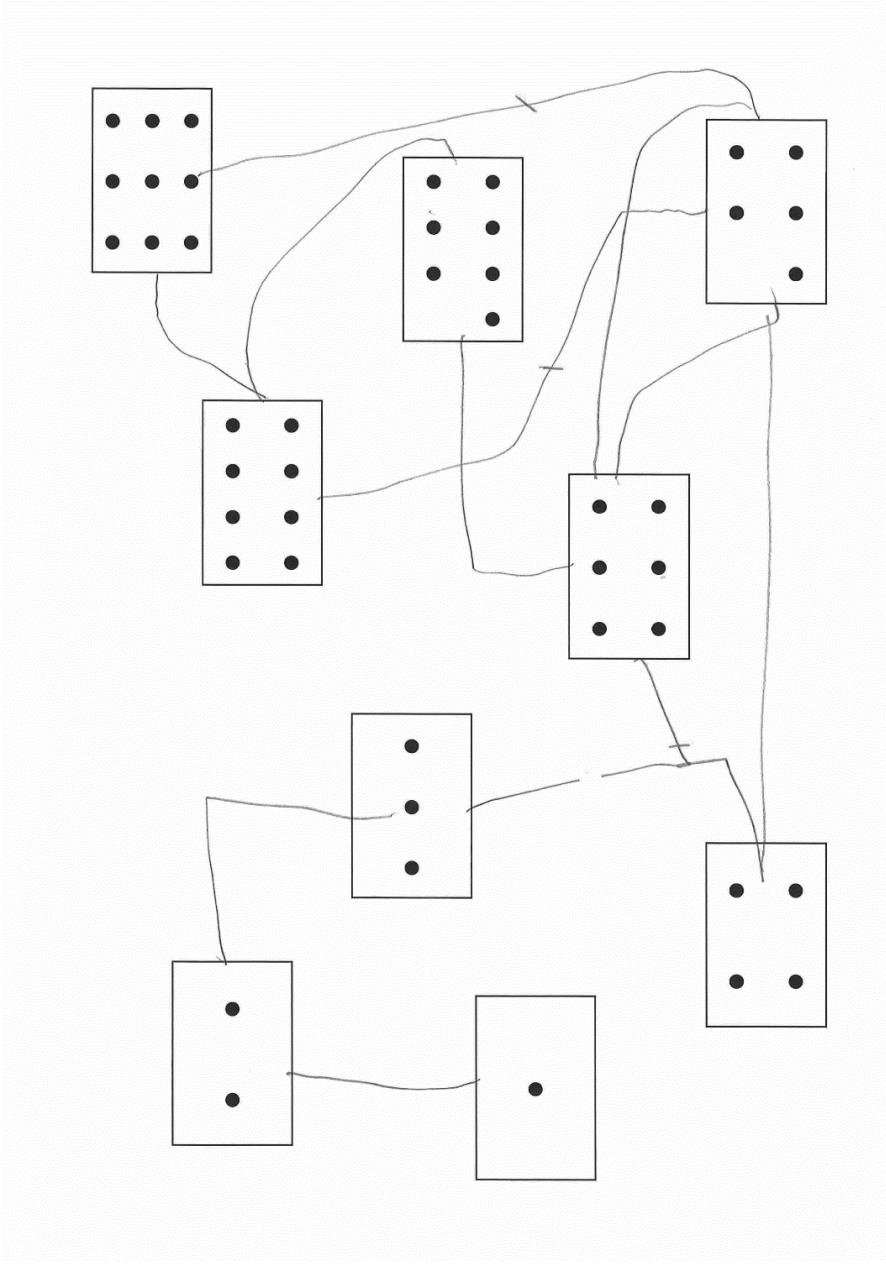
25. Do you have any specific ideas what researchers should study with regard to the cognitive domain(s) you scored as most important? E.g. which test(s) should they study, what aspects of this cognitive domain should be studied etc.?

End of the survey

26. If you have any other relevant comments, please specify them here; don't forget to press submit when you are finished:

CHAPTER 3

IMPROVEMENTS TO THE FIELD OF CROSS-CULTURAL NEUROPSYCHOLOGICAL ASSESSMENT



CHAPTER 3.1

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

The Clinical Neuropsychologist (revisions under review)

Abstract

Introduction:
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.

Methods:
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, $\rho = .33$, 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%).

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.

1 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 [47], stroke [47], and dementia [44]. 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 [13]. Assessment with an interpreter is often necessary, but formal interpreters are inconsistently used across Europe, particularly due to a lack of funding [229]. 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 [35-37,230]. 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 [19,178]. 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 [28,29,32,33,91].

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" [231]. Patients with a diverse background may not be familiar with 'best performance' or speed tests [178]. They may experience distress when the examiner points out errors or stops the test after the pre-set time limit has been passed [231]. 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 [231,232]. Additional practice items may need to be provided [233].

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 [184,193,229,234]. 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 [140]) and the Cross-Cultural Dementia Screening (CCD [57]), as well as European validation studies of the Rowland Universal Dementia Assessment Scale (RUDAS [60]; see also [235]). 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 [205]. 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. The first aim of this study was therefore to develop a suitable neuropsychological test protocol, including existing tests that show promise in cross-cultural neuropsychological assessment and newly developed tests where needed. 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 [236]. 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.

2 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. Last, we present the findings from our feasibility study.

2.1 Development of the TULIPA test battery

To determine which tests should be included in the neuropsychological test battery, we consulted the relevant international literature through a systematic review [184]. 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 (for more detail on the methods, see [229]). In short, 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 [184,229]). 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 three subsequent focus groups with 12 neuropsychologists experienced in assessing diverse patient populations (neuropsychologists present per focus group: 6–9), relevant barriers and facilitators were identified and appropriate tests selected. 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. All session were videorecorded with consent of the participants and were transcribed verbatim. 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.

In these focus groups, several barriers to cognitive testing were identified through thematic analysis of the focus group transcripts. 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 [237].

2.2 The TULIPA test battery

The neuropsychological tests 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. [196], the CCD [57], the modified Visual Association Test [201], 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 [67]. The supplementary battery contained two tests of visuospatial functioning: the Clock Reading Test from the CNTB [140] and the Stick Design Test [112]; 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 [150] and a Turkish version of the Stroop test [238], to be administered only to Turkish-speaking patients who are literate. The Corsi Block Tapping Test [239] was added as a measure of (visual) working memory as the more commonly used digit span is heavily influenced by language of administration [184]. The supplementary battery contained one additional memory test, the Recall of Pictures Test of the CNTB [140]. The Coin-in-the-Hand Test [240] was used to detect suboptimal performance. The Naming Assessment in Multicultural Europe (NAME [241]) 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. The 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 [242,243]) and adapted versions of the Geriatric Depression Scale (GDS [244,245]). In addition, acculturation was measured with a shortened, adapted Short Acculturation Scale for Hispanics (SASH [215]).

2.3 Implementation of the test battery

2.3.1. Consultation with individuals from culturally and linguistically diverse backgrounds

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 [246], 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. 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 asked through open-ended questions about their first impressions of the tests and what the tests might measure. Afterwards, the purpose

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

Core battery	Cognitive domain	Description	Approx. 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

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.

In line with the findings by Aghvinian et al. [231], the goal of each of the individual tests and the relationship with everyday cognitive functioning was often unclear to the participants. 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.

2.3.2 Optimization of test procedures

Subsequently, 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. The 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 [229]. Furthermore, two follow-up meetings were organized after data collection had started to share experiences and ensure test administration was comparable across centers. Last, we attempted to standardize interpreter-mediated assessment with the help of a team of bilingual, bicultural interpreters with a background in medicine, (neuro)psychology, or paramedical disciplines. 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.

Table 2. Demographic characteristics of the full sample[1]

	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(%):				
Zero years of education	42 (24%)	0 (0%)	16 (33%)	29 (30%)
>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[2]	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[3]	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[3]	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%)

Abbreviations: RUDAS = Rowland Universal Dementia Assessment Scale
Values are displayed as mean (standard deviation) unless otherwise specified.
[1] A number of cases are missing for education and years in the Netherlands because patients were asked but were unable to report it
[2] 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.
[3] Median (first quartile-third quartile)

2.4 Feasibility study in the memory clinic

2.4.1. 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 [247] 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.

2.4.2 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 [248], mild cognitive impairment [249], and dementia subtypes (e.g. [96,250]), and the DSM-V for primary psychiatric disorders [251].

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' [236]) 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 on test performance. We collected this information retrospectively from the observations recorded in the neuropsychological reports (see Supplementary Table 2 for example codes). We included both complicating factors that were self-reported as well as those observed by the neuropsychologist. 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.

2.5 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.

3 Results

3.1 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 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 ($\rho = .33$, adjusted $p < .001$). We did not find a significant 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.

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[1]	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%)

Abbreviations: RUDAS = Rowland Universal Dementia Assessment Scale, CCD = Cross Cultural Dementia Screening
[1] 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.

3.2 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 ($\rho = -.23$, adjusted $p = .001$) and female ($\rho = .21$, 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.

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%)

4 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 [193]. The TULIPA battery incorporates some of the psychometrically sound elements of the CNTB [140]—the only battery available for European diverse populations thus far [229]—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. [112,115]), diagnostic accuracy was poor in a later study [252]. 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 [229] 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. [253]), 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 [254]; 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 [212,214]. For example, one quarter of the healthy adults tested with the Test of Memory Malingering in Paraguay were misclassified as displaying insufficient effort [212]. Although the Amsterdam Short-Term Memory test showed more promising sensitivity and

specificity [214], 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. [255]), 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 [256].

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. [257,258]), 2) how often participants required breaks [258], 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.

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 [231-233]. Although this need is in no way unique to diverse populations (see e.g. [259]), 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. [260] 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 [261].

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.

Acknowledgments and funding
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. This work was supported by the Netherlands Organisation for Health Research and Development [grant number: 733050834].

Conflict of interest
The authors report no conflicts of interest.

Chapter 3.1 Supplementary material

Supplementary Table 1. Examples of quotes from participants and how the input was subsequently used.

Topic	Quote (translated to English)	How input was used
General comments	"These tests look childish"	During the consultation, participants were asked how the tests could be made to feel less childish. They responded that the neuropsychologist should stress how important the doctor thinks it is to test their physical condition and to highlight the connection with the brain and memory. Another participant responded that it is important for the neuropsychologist to behave in a serious manner and not talk to the patient like a child.
	"A patient may feel ashamed, anxious or insecure, and they will hide themselves or won't dare [to answer] because they are mistrusting"	Neuropsychologists should provide information to the patient beforehand that some tests may be more challenging than others. Monitor how the patient is feeling and try to comfort patients.
Time pressure	"Give people who are illiterate time to think, do not give them X number of seconds"	These quotes highlight that participants felt nervous/afraid of having to do tests as fast as possible—even when there were no explicit time limits, such as in the Clock Reading Test. In the instructions, neuropsychologists should therefore make sure to explain whether the patient does or does not have to work as fast as possible. In future studies, it should be examined whether the tests can be reduced in length by half without compromising on sensitivity/specificity.
	"It looks easy, but you have to run through it like a high speed train"	
	"If you have sufficient time, you can really look closely"	
	"These fast tests make you dizzy"	
	When asked what the Clock Reading Test is supposed to measure: "reading times on the clock", "reaction speed, how fast you can say what time it is"	
Clock Reading Test	"Not all persons who are illiterate know how to tell the time"	These three quotes highlight that neuropsychologists should not assume that all patients were (premorbidly) able to tell the time on an analog clock. Instead, verify during history taking while the caregiver is present.
	"Everyone knows how to tell the time: 12, 3, 6, 9"	
	"Some older men with a migration background may wear a watch without being able to tell the time. If someone asks them for the time, they will show them their watch"	
	"In Turkey, we say: it is half twelve, but in Dutch we say: it is half one"	Make sure to ask the patient to answer consistently in one language, preferably the language they are most proficient in. If patients report in two languages, take this into consideration in the interpretation.

Supplementary Table 1. Continued.

Topic	Quote (translated to English)	How input was used
Five Digit Test	"This looks just like an exam"	If necessary, acknowledge to the patient that the test may seem overwhelming. Make sure to write down time when patient passes the halfway mark in case the patient becomes overwhelmed.
	"I someone keeps pushing you, you will start to make errors"	Mention beforehand that the test may feel somewhat overwhelming and that this is a normal experience.
	When asked what this test could be meant to measure: "Memory loss? Long term memory?" "Motor skills? Memory? Orientation?" "How fast you can switch?" "How fast you can see whether it is a three, four, or five. Concentration?" "How fast you can recognize [items]?"	These responses highlight that it is not evident from the stimuli what the test measures, and that it may be necessary to explain to the patient either before, during, or after the assessment how the tests that were administered relate to everyday cognitive complaints
Stick Design Test	[The arrow item is a] "clue" "pointing [the participant] in a certain direction"	These quotes show that it is necessary to ensure the patient does not think you are trying to trick them or that the test has a hidden meaning. If necessary, provide an example of how brain damage may impair performance (e.g. neglect).
	"No, this is one of those visual tricks" [where the lines seem of equal length but are not]	
	"It can be confusing if someone has to use items in a different way"	

Supplementary Table 2. Examples of coding for analyses of complicating factors.

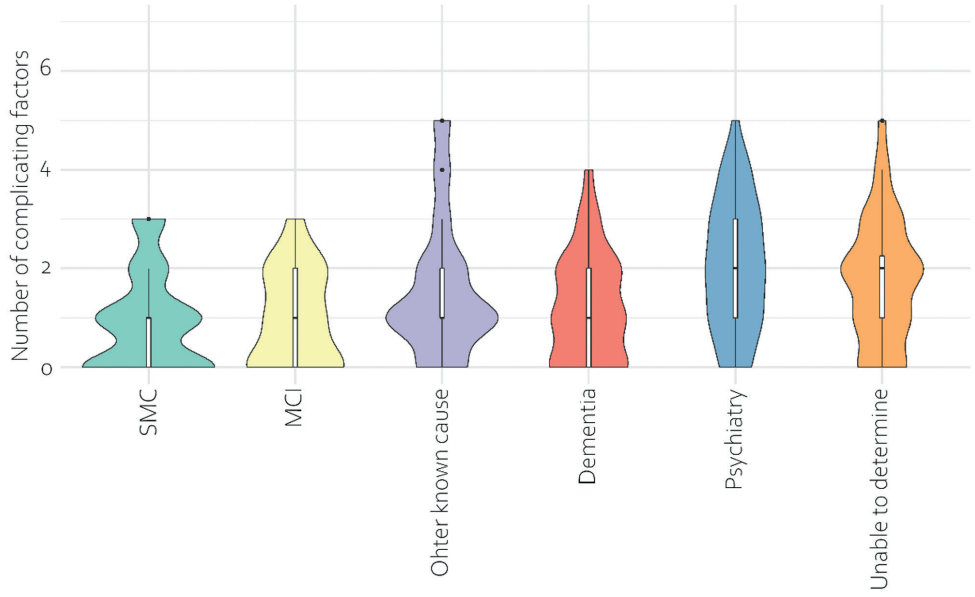
	Examples of included codes (translated to English)	Examples of included codes (Dutch)	Codes not included
Suboptimal effort/ malingering	Limited motivation; uncooperative; patient indicates [multiple times] that they do not want to continue; patient needs to be motivated to continue until optimal performance is reached	'Matig gemotiveerd'; 'niet coöperatief'; 'geeft [meermaals] aan niet verder te willen gaan'; 'moet aangespoord worden om tot een optimale prestatie te komen'	The motivation decreases slightly as testing continues; patient asks if it is the final test; the motivation is slightly decreased; patient has to be encouraged to start with testing
Depression	Sad; depressed; dejected; patient becomes emotional	Stemming is 'gedrukt'; 'somber'; 'gelaten'; patiënt 'wordt emotioneel'	Feeling slightly/very mildly down (sometimes: and this mood improves during testing)
Anxiety	Nervous; anxious; tense; insecure	'Nerveus'; 'gespannen'; 'onzeker'	
Fatigue	Tired; weary; eyelids keep dropping/patient is dropping off	'Vermoeid'; 'moe'; 'ogen vallen dicht'	
Pain	Headache; backache; toothache	'Hoofdpijn'; 'rugpijn'; 'kiespijn'	
Other physical symptoms	Dizziness; 'full head'; tingling sensations	'Duizelig'; 'vol hoofd'; 'tintelingen'	
Motor impairment	Tremor; speech impairment for which patient uses prosthetic device	'Tremor'; 'spraakproblemen' waarvoor spraakprothese	
Sensory impairment	Deaf; blind; visually impaired; blurry or double vision; patient who wears hearing aids but has forgotten to bring them	'Doof'; 'blind'; 'beperkte visus'; 'wazig' of 'dubbel' zien; patiënt is vergeten gehoorapparaat mee te nemen	

Supplementary Table 3. Administration and completion rates per study site.

	Rotterdam 1 (n = 177)		Rotterdam 2 (n = 22)		Enschede (n = 48)		The Hague (n = 98)	
	Administered n(%)	Not completed n (%)	Administered n(%)	Not completed n (%)	Administered n(%)	Not completed n (%)	Administered n(%)	Not completed n (%)
Core battery								
RUDAS	152 (86%)	4 (3%)	20 (91%)	0 (0%)	41 (85%)	0 (0%)	77 (79%)	2 (3%)
CCD Objects test A	152 (86%)	3 (2%)	19 (86%)	0 (0%)	43 (90%)	1 (2%)	84 (86%)	3 (4%)
CCD Objects test B	146 (83%)	5 (3%)	19 (86%)	0 (0%)	39 (81%)	1 (3%)	80 (82%)	4 (5%)
CCD Sun Moon test A	149 (84%)	2 (1%)	19 (86%)	1 (5%)	40 (83%)	2 (5%)	82 (84%)	3 (4%)
CCD Sun Moon test B	146 (83%)	19 (13%)	18 (82%)	1 (5%)	39 (81%)	6 (15%)	78 (80%)	8 (10%)
CCD Dots test A	140 (79%)	25 (18%)	17 (77%)	1 (6%)	39 (81%)	6 (15%)	79 (90%)	17 (22%)
CCD Dots test B	123 (70%)	54 (44%)	15 (68%)	5 (33%)	32 (67%)	9 (28%)	60 (61%)	29 (48%)
Animal fluency	160 (90%)	2 (1%)	21 (96%)	0 (0%)	38 (79%)	0 (0%)	76 (78%)	0 (0%)
Food fluency; supermarket fluency	113; 2 (64%; 1%)	0; 0 (0%)	21 (96%)	0 (0%)	34 (71%)	0 (0%)	18; 33 (18%; 34%)	0; 0 (0%)
mVAT (short or long)	113 (64%)	1 (1%)	21 (95%)	0 (0%)	41 (85%)	0 (0%)	52 (53%)	2 (4%)
Supplementary tests								
Literacy screener total	68 (39%)	1 (1%)	1 (5%)	0 (0%)	1 (2%)	0 (0%)	1 (1%)	0 (0%)
FDT Reading and Counting	47 (27%)	0 (0%)	2 (9%)	0 (0%)	1 (2%)	0 (0%)	1 (1%)	0 (0%)
FDT Choosing and Shifting	47; 41 (27%; 23%)	1; 4 (2%; 10%)	2; 0 (9%; 0%)	0 (0%); n.a.	1; 0 (2%)	0 (0%); n.a.	1 (1%); 0	0 (0%); n.a.
Turkish Stroop Cards 1; 3; 4 (Attention/speed)	2 (1%)	0 (0%)	5 (23%)	0 (0%)	8 (17%)	0 (0%)	2 (2%)	0 (0%)
Turkish Stroop Cards 2; 5 (Executive)	2 (1%)	0 (0%)	5 (23%)	0 (0%)	8 (17%); 7 (15%)	0 (0%)	2 (2%)	0 (0%)
RPT – naming subtest	55 (31%)	0 (0%)	11 (50%)	0 (0%)	7 (15%)	0 (0%)	17 (17%)	0 (0%)
RPT – memory subtests	51 (29%)	0 (0%)	11 (50%)	0 (0%)	7 (15%)	0 (0%)1	17 (17%)	0 (0%)
Corsi Block Tapping Test	57 (32%)	0 (0%)	5 (23%)	0 (0%)	4 (8%)	0 (0%)	0 (0%)	n.a.
Coin in the Hand Test	49 (28%)	0 (0%)	8 (36%)	0 (0%)	43 (90%)	1 (2%)	12 (12%)	2 (17%)
Stick Design Test	51 (29%)	0 (0%)	7 (32%)	0 (0%)	3 (6%)	0 (0%)	11 (11%)	2 (18%)
Clock Reading Test	58 (33%)	6 (10%)	12 (55%)	0 (0%)	11 (23%)	0 (0%)	4 (4%)	1 (25%)
NAME	86 (49%)	0 (0%)	2 (5%)	1 (50%)	2 (4%)	0 (0%)	5 (5%)	2 (40%)

Abbreviations: RUDAS = Rowland Universal Dementia Assessment Scale, CCD = Cross-Cultural Dementia Screening, mVAT = modified Visual Association Test, FDT = Five Digit Test, RPT = Recall of Pictures Test, Association Test, NAME = Naming Assessment in Multicultural Europe

[1] Two patients were not administered the delayed recall and recognition trials



Supplementary Figure 1. Violin plot with superimposed boxplots displaying the number of complicating factors by diagnostic group. Wider sections in the violin plot represent a larger number of patients with that specific number of complicating factors.

CHAPTER 3.2

Assessment of visual association memory in low-educated, non-Western immigrants with the modified Visual Association Test

Sanne Franzen
Esther van den Berg
Yuled Kalkisim
Lotte van de Wiel
Marleen Harkes
Rozemarijn L. van Bruchem-Visser
Frank Jan de Jong
Lize C. Jiskoot
Janne M. Papma

Dementia and Geriatric Cognitive Disorders, 2019, 47(4–6), 345–354

Abstract

Introduction:
Neuropsychological tests are influenced by culture, language, level of education, and literacy, but there are few cognitive tests of which the applicability in ethnic minority populations has been studied. The aim of this study was to assess the reliability and validity of the Visual Association Test (VAT), a test of visual association memory, in a non-Western, low-educated memory clinic population. Additionally, a modified version of the VAT using colored photographs instead of line drawings was studied (mVAT).

Methods:
Both the original VAT and the mVAT were administered to non-Western immigrants (n = 73) from two multicultural memory clinics in Rotterdam, the Netherlands, and a control sample of non-demented Turkish elderly (n = 14) with low education levels (32 and 29% illiterate, respectively).

Results:
Both the VAT and the mVAT were able to discriminate persons with and without dementia (area under the curve: VAT, 0.77–0.88; mVAT, 0.85–0.95). The mVAT had more homogeneous item difficulty levels than the VAT. Administration of parallel versions of the VAT and the mVAT within the same person revealed higher scores on the mVAT ($Z = -3.35$, $p = .001$).

Conclusions:
The mVAT is a reliable and valid measure of memory in non-Western immigrants. Clinicians and researchers should be aware that the memory performance of immigrants may be systematically underestimated when using tests with black-and-white line drawings, such as the original VAT.

1 Introduction

Neuropsychological examination is fundamental to the assessment of dementia. Neuropsychological test performance is known to be substantially affected by culture, language, (quality of) education, and literacy [12,17,30,32,33,39,77,78]. For example, healthy illiterate persons have lower scores on visual naming tests using black-and-white line drawings than literate persons [29,181,262,263]. This difference disappears when colored photographs are used [29], which is most likely related to the higher level of detail provided by the colored photographs [28]. This example illustrates how tests developed for educated, Western people cannot readily be used in other populations. In recent years, new screening instruments for dementia, such as the Cross-Cultural Dementia Screening or CCD [57] and the Rowland Universal Dementia Assessment Scale or RUDAS [60,196,197,206], have therefore been developed and validated for ethnic minority populations. These instruments are designed to screen for dementia. However, domain-specific neuropsychological tests that can determine the underlying etiology are lacking.

The Visual Association Test (VAT) [67] is a test of visual association memory that is particularly able to discriminate between Alzheimer's disease (AD) and other types of dementia, as performance is associated with atrophy of the medial temporal lobe [264]. In the VAT, patients are required to remember two interacting objects (such as a stroller with a bird in it, or a monkey with an umbrella) presented in the form of black-and-white line drawings. Although this test is frequently used in clinical practice and was previously recommended as best practice in the neuropsychological assessment of non-Western immigrants [52], the reliability and validity of the VAT have not been assessed in an ethnic minority population or in people who are illiterate.

Given the above mentioned difficulties regarding the naming of black-and-white line drawings in healthy illiterate people, the question can be raised of whether the VAT puts illiterate or low-educated patients at a disadvantage, for example due to difficulties recognizing and thus remembering the objects. These difficulties may also pose a threat to the validity of the VAT as a measure of visual association memory in these persons. We hypothesize that people who are low educated could benefit from a test that uses colored photographs instead of black-and-white line drawings. The aims of this study were to examine the reliability and validity of the original version of the VAT and a modified version using colored photographs (mVAT) and to compare performance on both tests in a population of low-educated non-Western immigrants.

2 Methods

2.1 Participants

Seventy-three non-Western immigrant patients who visited the outpatient multicultural memory clinics of the Erasmus MC University Medical Center and the former Havenziekenhuis in Rotterdam, the Netherlands, were included between April 2016 and October 2018. The patients had immigrated from Turkey (n = 34), Morocco (n = 13), Cape Verde (n = 10), Pakistan (n = 4), Iraq (n = 2), Afghanistan (n = 2), the State of Palestine (n = 1), Syria (n = 1), Egypt (n = 1), China (n = 1), Venezuela (n = 1), the Dutch Antilles (n = 1), Suriname (n = 1), and Macedonia (n = 1). The diagnostic workup consisted of a

comprehensive clinical evaluation, with history taking by a geriatrician or neurologist, a neuropsychological assessment, laboratory screening with blood tests, and (in a subset of patients) structural brain imaging (computed tomography, n = 23; magnetic resonance imaging, n = 25). Diagnosis was determined in a multidisciplinary consensus meeting which included a neuropsychologist, a neurologist, a radiologist, and a geriatrician, using the diagnostic research criteria for dementia subtypes [96,249,250].

Fourteen healthy Turkish community-dwelling individuals were included as a control group. The inclusion criteria for the control group were: age >50 years, free of self-reported cognitive complaints, and a RUDAS [196] score ≥23. For nine healthy controls, informants filled out the short IQCODE [242], confirming the patient's self-reported absence of cognitive complaints. In the other cases, no informant was available. All healthy controls provided a written informed consent. They were recruited in an urban area through centers providing activities for Turkish elderly and through the personal networks of included participants.

2.2 Measures

2.2.1 The original VAT

In the original VAT, patients are asked to name six stimuli on consecutive black-and-white line drawings (the cue cards) presented in a paper booklet. Patients are then shown cards on which the previous stimuli are interacting with new stimuli, and they are asked to name both items on each consecutive card. Patients are then again shown the initial cue cards and asked to name the missing item. This procedure is repeated in trial 2 (unless there is a maximum score on the first trial). Naming errors are allowed as long as the names are specific enough to be identified as correct or incorrect in the reproduction trial, i.e. "prickly animal", "prickly thing", or even "brush" would be sufficiently specific to indicate a hedgehog and would thus be considered correct.

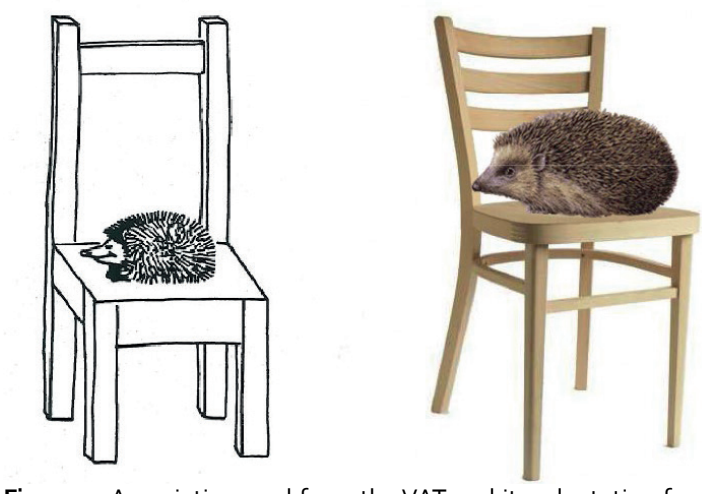


Figure 1. Association card from the VAT and its adaptation for the mVAT (reprinted with permission from the Hogrefe Publishing Group).

2.2.2 Modification of the VAT

To create a photograph version of the VAT, henceforth called the modified VAT (i.e. mVAT), all black-and-white line drawings from the original VAT were replaced with colored photographs of the same objects (Fig. 1). No other changes were made to the test. Stimuli were presented on A5-sized booklets, similar to the original VAT. The original VAT has multiple parallel versions. Therefore, modifications of both the original VAT version A and the parallel version B were made.

2.2.3 Other cognitive tests

The patients were tested with the CCD [57], a comprehensive screening test for dementia, assessing the domains of memory, mental speed, and executive functioning. It was specifically developed and validated for a large sample of immigrants in the Netherlands, and normative data are available. Furthermore, patients were administered the RUDAS [196] and/or the Mini-Mental State Examination (MMSE) [198]. In some patients, other tests, such as the Location Learning Test [265] (n = 15) and category verbal fluency (supermarket) (n = 10), were administered as well.

2.3 Procedure

All participants from the outpatient clinic underwent a neuropsychological examination. Interpreters were present during these examinations, as well as during the intake interview, for all participants who were insufficiently fluent in Dutch (96%, n = 70). The administration procedures for the VAT (and thus the mVAT) according to the test manual [67] were followed, including prorating of the second trial. The administration time of the VAT (and the mVAT) is approximately five minutes. As the stimuli were identical for both the VAT and the mVAT (apart from line drawings vs. photographs), patients could not be administered version A of both the VAT and the mVAT. Therefore, all patients were administered either VAT version A and mVAT version B or VAT version B and mVAT version A. In the majority of cases, the mVAT was administered before the VAT to ensure that a higher score on the mVAT was not caused by the participant being aware at the first trial that the stimuli needed to be remembered. In a subset of patients, this order was reversed.

Control participants were assessed in Turkish, either at home or at the day activity center, by a neuropsychologist who is a native speaker of Turkish (YK). Similar to the patients, the order of administration of the VAT versus the mVAT, as well as the versions that were used, was varied.

2.4 Statistical analysis

Differences in demographic characteristics were analyzed with χ^2 tests for nominal data and t tests for continuous data. Cronbach's α and item-total correlations were used to determine the internal consistency, and Cochran Q tests were used to assess item difficulty levels. For reliability analyses, both parallel versions of the VAT and the mVAT were analyzed separately. For all other analyses, parallel versions A and B of the original and A and B of the mVAT were merged. As the scores of both the VAT and the mVAT were not normally distributed, a Wilcoxon signed-rank test was used to analyze intraindividual differences in the performance on the mVAT versus the VAT. Spearman correlations were used for correlational analyses to determine the convergent validity and the relationship with demographic variables. Receiver operating characteristic curves were used to assess

diagnostic accuracy. In secondary analyses differences were analyzed in participants with a low education level (less than primary school) versus participants who had received more education (primary school and up).

3 Results

As is shown in Table 1, controls were more often female, but the groups did not differ in terms of age, years of residence in the Netherlands, or level of education. For 14% of the patients in the sample, the diagnosis could not be determined definitively with the current diagnostic procedures, or additional procedures to determine the diagnosis, such as a lumbar puncture, failed or were refused by the patient. These patients remained in the analyses. All of the included patients and controls had normal or corrected-to-normal vision. None of the patients or controls reported color blindness.

3.1 Reliability

3.1.1 Internal consistency

The first trial of VAT version A displayed a Cronbach's α of 0.75, while the first trial of VAT version B had a Cronbach's α of 0.41 (Table 2). Cronbach's α for the first trial of mVAT version A was 0.63, and Cronbach's α was 0.81 for the first trial of mVAT version B. Item-total correlations were strong for the mVAT, as well as for VAT version A (Table 2). VAT version B, however, had two items with nonsignificant correlations with the total score for trial 1.

3.1.2 Item difficulty

Table 3 shows the percentage of people who remembered an item correctly at the first trial of the VAT and mVAT versions A and B. Overall, the item difficulty was lower for the mVAT than for the VAT, without reaching ceiling effects. Item difficulty levels were homogeneous for VAT version A and mVAT version A. Heterogeneity in item difficulty levels was observed for VAT version B (Cochran's $Q = 24.8$, d.f. = 5, $p < .001$). Version B of the mVAT did not have homogeneous difficulty levels either (Cochran's $Q = 22.0$, d.f. = 5, $p = .001$), as item 5 (leaf-syringe) was more difficult than the other items, but after deleting this item the test was homogeneous (Cochran's $Q = 0.7$, p = n.s.).

For the distribution of the scores in the first trial for patients and controls, see Figure 2. There were 324 correct responses in the first trial of the mVAT (n = 87); of the subjects who also completed trial 2, there were 304 cases followed by a correct response on the same item in the second trial (93.8%). Similarly, on the VAT, 260 correct responses were recorded for the first trial (n = 81), 250 cases (96.2%) of which were followed by a correct response on the same item in the next trial.

3.2 Validity

3.2.1 Intraindividual performance on the VAT and the mVAT

The intraindividual performance of all of the participants in the first trial of the mVAT was significantly higher than in the first trial of the VAT ($Z = -3.35$, $p = .001$). The results were comparable when the controls were analyzed separately ($Z = -2.31$; $p = .021$) or the patients separately ($Z = -2.83$; $p = .005$).

Table 1. Demographic characteristics, cognitive test scores, and group comparisons for the whole sample.

	Controls (n = 14)	Memory clinic patients (n = 73)	Significance
Age, years	62.21 ±11.49	68.48±11.00	n.s.
Education			
0 years of education/illiterate	4 (29)	23 (32)	n.s.
1 year of education up to primary education	6 (43)	27 (37)	
>primary education	4 (29)	21 (29)	
Male gender, %	21	55	p = .02
Time in the Netherlands, years	36.9 ±16.3	38.5±7.4	n.s.
RUDAS	26.86±1.92	22.1±5.0[a]	
MMSE	–	16.6±5.8[b]	
CCD objects test A[c]	–	113.2±9.6	
CCD objects test B[c]	–	104.7±12.2	
Diagnosis, n(%)			
Subjective memory complaints	–	13 (18)	
Mild cognitive impairment	–	13 (18)	
Dementia (AD, VaD, mixed, and other)	–	22 (30)	
Primary psychiatric disorder (e.g. depression)	–	8 (11)	
Cognitive disorder due to another known medical condition	–	7 (10)	
Could not be determined	–	10 (14)	

Abbreviations: VaD = vascular dementia. Values are displayed as means ± SD or numbers (%) unless otherwise specified. [a] n = 26 patients. [b] n = 54 patients. [c] The maximum score for the objects test A (immediate recognition) and B (delayed recognition) is 122. The general cut-offs for dementia are ≤118 for objects test A and ≤109 for objects test B.

Table 2. Item-total Spearman's correlations and Cronbach's α for the first trial of all versions.

	Trial 1 mVAT A (n = 32)	**Trial 1 mVAT B (n = 55)**	**Trial 1 VAT A (n = 55)**	**Trial 1 VAT B (n = 25)**
Item 1 (ρ)	0.64	0.77	0.61	0.57
Item 2 (ρ)	0.57	0.59	0.59	0.60
Item 3 (ρ)	0.54	0.74	0.72	0.50
Item 4 (ρ)	0.46	0.73	0.68	0.34 (n.s.)
Item 5 (ρ)	0.56	0.73	0.68	0.69
Item 6 (ρ)	0.67	0.68	0.71	0.18 (n.s.)
Cronbach's α	0.63	0.81	0.75	0.41

All correlations are significant at p < 0.05 unless otherwise specified.

Table 3. Percent correct per item of mVAT A and B and VAT A and B.

	Trial 1 mVAT A	**Trial 1 mVAT B**		**Trial 1 VAT A**	**Trial 1 VAT B**
A item 1	69	55	B item 1	62	64
A item 2	75	54	B item 2	68	28
A item 3	72	64	B item 3	66	44
A item 4	60	44	B item 4	64	72
A item 5	75	56	B item 5	36	64
A item 6	53	56	B item 6	64	24

Values are presented as percentages.

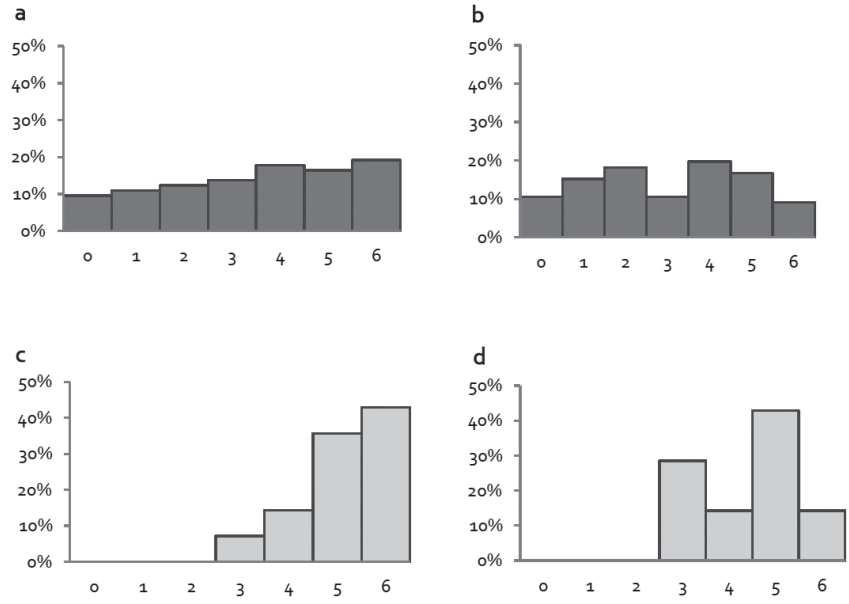


Figure 2. Scores on the first trial of the mVAT (a, c) and VAT (b, d) for patients (dark gray) and controls (light gray).

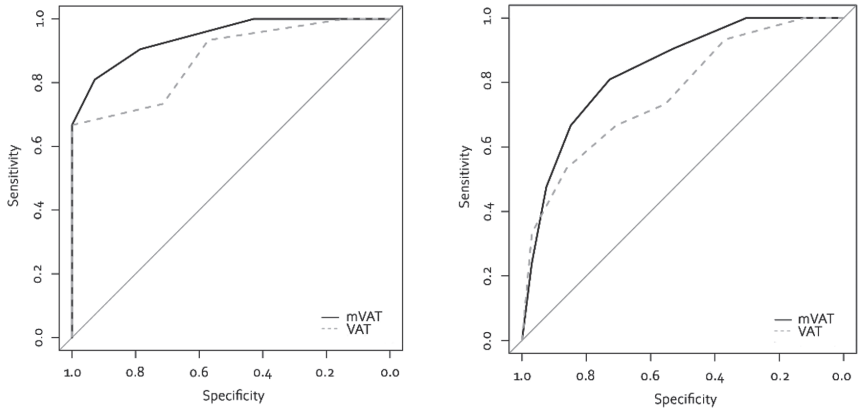


Figure 2. Scores on the first trial of the mVAT (a, c) and VAT (b, d) for patients (dark gray) and controls (light gray).

3.2.2 Discriminative validity

The discriminative abilities of the mVAT and VAT (first trial) for patients with dementia versus healthy controls are displayed in Figure 3. The area under the curve (AUC) was 0.95 for the mVAT (n = 36) and 0.88 for the VAT (n = 30). When discriminating patients with dementia from the entire sample (including healthy controls), the AUC for the first trial of the mVAT was good (i.e. 0.85, n = 87), and the AUC for the first trial of the VAT was fair (i.e. 0.77, n = 80; Figure 3). A separate analysis of only the Turkish participants revealed similar results. The sample sizes for this study were too small to examine specific dementia subtypes.

3.2.3 Convergent validity

The mVAT was moderately correlated with the MMSE (ρ = .51, p < .001, n = 54), the RUDAS (ρ = .61, p < .001, n = 39), and the CCD objects test part A (ρ = .58, p < .001, n = 71) and part B (ρ = .63, p < .001, n = 70). The VAT showed similarly moderate correlations with these cognitive screening measures (MMSE: ρ = .54, p < .001, n = 49; RUDAS: ρ = .57, p < .001, n = 38; CCD objects test part A: ρ =.57, p < .001, n = 65; and CCD objects test part B: ρ = .54, p < .001, n = 64). After splitting the group into a low-educated group and a highly educated group, the correlations of the mVAT with the CCD and MMSE remained similar, but the correlation with RUDAS was no longer significant in the low-educated group (Table 4).

Table 4. Correlation of mVAT trial 1 with demographic variables and neuropsychological tests by education level

	Lower than primary school	**Primary school and higher**
Education	.586**	–.026
Age	–.310*	–.409**
Gender	–.019	.200
Time in the Netherlands (years)	–.425*	–.276
RUDAS	.244	.640**
MMSE	.532*	.678**
CCD objects test A	.576**	.630**
CCD objects test B	.608**	.686**
VAT trial 1	.672**	.712**

* Correlation is significant at the .05 level. ** Correlation is significant at the 0.01 level.

3.2.4 Relationship with demographic variables

The mVAT was weakly correlated with education level (ρ = .25, p = .02) and uncorrelated with gender (ρ = .06, p = n.s.), as was the VAT (education: ρ = .29, p = .01; gender: ρ = .01, p = n.s.). The VAT was moderately correlated with age (ρ = –.43, p < .001) and weakly correlated with years of residence in the Netherlands (ρ = –.38, p < .001), and the mVAT was weakly correlated with age (ρ = –.37, p < .001) and years of residence in the Netherlands (ρ = –.32, p < .01). After splitting the group into a low-educated and higher educated group, the correlation of the mVAT with education level and with years of residence in the Netherlands only remained significant for the low-educated group (Table 4). The correlations with age and gender remained similar.

4 Discussion

This study showed that both the VAT and the mVAT are useful measures of visual association memory in non-Western immigrants. The mVAT, however, had a better discriminative ability and less heterogeneous item difficulty levels than the VAT.

Administration of parallel versions of the VAT and the mVAT within the same person revealed that, both for controls and for patients, higher scores were obtained on the mVAT. This finding indicates that the memory performance of non-Western immigrants may be systematically underestimated when using the original VAT with its black-and-white line drawings. This is an important finding, as memory tests for ethnic minority populations often contain visual stimuli to circumvent language barriers, and black-and-white line drawings, such as in the picture version of the Free and Cued Selective Recall Test [266] and the Location Learning

Test [265,267], are still widely used. Even some newly developed tests for illiterate and low-educated subjects contain black-and-white line drawings [134,135].

The better performance on the mVAT is best explained by the added information about the color provided by the colored photographs (as opposed to surface texture information) [28,268]. The added color may be particularly important for low-educated people, as the ability to decode black-and-white line drawings (such as those in the VAT) is thought to be acquired through education and literacy training [28,91].

The reliability analyses indicated that only one item of the mVAT would have to be replaced before the mVAT could be applied in ethnic minority populations. This may seem counterintuitive, as some of the items of the mVAT do not necessarily appear to be universal. This finding may best be explained by the fact that misnamed or misidentified objects are not scored as incorrect as long as they are specific enough to be scored in the recall trial. Therefore, a hedgehog that is not recognized as such but is instead called a "brush" may have the same item difficulty as an item that seems more cross-culturally recognizable.

A remarkable finding was that performance on the mVAT was associated with education in a group of illiterate to minimally educated persons but not in a group with primary education levels and higher. This is in line with findings that even one year of education may substantially alter the performance on cognitive tests [33] and supports the notion of a nonlinear effect of years of education on cognitive performance [269].

A limitation of this study is that a final diagnosis could not be determined in a subset of patients. For example, in some patients it remained unclear whether the symptoms were caused by dementia, a primary psychiatric disorder, other medical conditions (e.g. diabetes and cardiovascular disease), or a combination of these factors. This seems inevitable, as diagnosing dementia in non-Western immigrants is challenging [193,270] and both under- and overdiagnosis of dementia are common [75]. Strengths of this study were the relatively large sample size and the administration of both the VAT and mVAT within the same person, enabling a direct comparison between the performance on the two tests.

In conclusion, the mVAT is a valid, brief, and easy-to-administer test to measure visual association memory in low-educated immigrant populations. Clinicians should be aware that memory performance may be systematically underestimated when using memory tests with line drawings. Future studies with the mVAT should be aimed at including more dementia patients with a wide variety of dementia diagnoses, so the discriminative abilities can be examined for various subtypes of dementia and dementia stages.

Acknowledgments and funding
This study was supported by grant 733050834 from The Netherlands Organization of Scientific Research (ZonMw Memorabel).

Conflict of interest
The authors have no conflict of interests to declare.

Statement of Ethics

The study protocol has been approved by the research institute's committee on human research.

CHAPTER 3.3

The Naming Assessment in Multicultural Europe (NAME): development and validation in a multicultural memory clinic

Sanne Franzen
Esther van den Berg
Yavuz Ayhan
Djaina D. Satoer
Özge Türkoğlu
Elif Genç Akpulat
Evy G. Visch-Brink
Esther A. Scheffers
Judi Kranenburg
Lize C. Jiskoot
Judy van Hemmen
Janne M. Papma

Journal of the International Neuropsychological Society: JINS, 2022.
Advance online publication

Abstract

Introduction:
Traditional naming tests are unsuitable to assess naming impairment in diverse populations, given the influence of culture, language, and education on naming performance. Our goal was therefore to develop and validate a new test to assess naming impairment in diverse populations: the Naming Assessment in Multicultural Europe (NAME).

Methods:
We carried out a multistage pilot study. First, we generated a list of 149 potentially suitable items—e.g. from published cross-linguistic word lists and other naming tests—and selected those with a homogeneous age of acquisition and word frequency across languages. We selected three to four colored photographs for each of the 73 remaining items; 194 controls selected the most suitable photographs. Thirteen items were removed after a pilot study in 15 diverse healthy controls. The final 60-item test was validated in 39 controls and 137 diverse memory clinic patients with subjective cognitive impairment, neurological/neurodegenerative disease or psychiatric disorders in the Netherlands and Turkey (mean age: 67, SD: 11). Patients were from 15 different countries; the majority completed primary education or less (53%).

Results:
The NAME showed excellent reliability (Spearman-Brown coefficient: 0.95; Kuder-Richardson coefficient: 0.94) and robust correlations with other language tests (ρ = .35–.73). Patients with AD/mixed dementia obtained lower scores on most (48/60) NAME items, with an area under the curve of 0.88. NAME scores were correlated with age and education, but not with acculturation or sex.

Conclusions:
The NAME is a promising tool to assess naming impairment in culturally, educationally, and linguistically diverse individuals.

1 Introduction

Naming impairment is frequently reported across a variety of neurological diseases, such as in temporal lobe epilepsy [271], post-stroke [272], in brain tumors [273], and in various neurodegenerative diseases, such as Alzheimer's disease dementia (AD) and frontotemporal dementia (FTD [274]). An assessment of naming impairment is therefore an important part of neuropsychological assessment. It is traditionally measured by presenting a series of items (often images) to the patient. The process of naming such visually presented items requires intact visual perception, accurate semantic processing of the stimulus, accurate selection of the lexical item, and correct (motor) execution of the stimulus' name [275]. The difficulty level of an individual item depends on a number of factors, such as the age of acquisition of the lexical item, the word frequency and familiarity, phonemic complexity, morphological length, and several other factors [276].

The Boston Naming Test (BNT [65]) is the most widely available and used test to assess naming impairment in the USA, Canada, and Europe [277,278]. It contains 60 black-and-white line drawings and has been demonstrated to be effective in detecting naming impairment across a variety of neurological diseases. Three decades of research, however, have indicated that tests such as the BNT cannot readily be applied to culturally, linguistically, and educationally diverse populations. Studies in the USA revealed large differences in BNT performance between white, African-American, Latino/a, and Asian participants [86,279], even after controlling for age, general cognitive impairment, formal education, and reading level [279]. Research suggests that the test stimuli themselves may be systematically biased against certain groups [86], and studies from Australia [280], New Zealand [18], French-speaking Canada [281], and Korea [126], identified several items that are not culturally appropriate in those settings, such as the pretzel, beaver, and asparagus. Furthermore, some items may be less suitable depending on whether participants come from a rural versus an urban environment within the same country [100]. As item difficulty levels depend on the cultural and language background of the person being assessed, the optimal order of administration of the items will also vary [282]. Controversial items such as the noose—an item that is considered particularly harmful because of its connection with historical racism—provide further reasons to use tests other than the BNT in diverse populations [283]. Although some of these issues may be addressed by using normative data specific to these diverse populations, this approach has been criticized for potentially increasing false negative rates in some cases [232,284].

In addition to the effects of language and culture on naming test performance, another major factor to impact performance on traditional naming tests is education. A higher level of education may directly influence test score through increased vocabulary and exposure to certain items not otherwise encountered in daily life, but can also (indirectly) impact the test score through differences in the processing of the stimuli. Reis et al. [29] have shown that people who are illiterate are significantly better at naming colored photographs of everyday objects than black-and-white line drawings of the same objects. On further evaluation [28], it was found that this was most likely related to the added detail that the color provided.

Few tests are currently available that address these issues in culturally, linguistically, and educationally diverse patients [184,276]. The Multilingual Naming Test (MINT)—which was originally developed to assess Spanish, English, Mandarin, and Hebrew bilinguals—was described by its authors as "relatively culture-neutral" [285]; however, culturally, educationally, and linguistically diverse individuals in Europe may never have encountered some of the MINT's stimuli in their daily lives—such as the porthole, gauge, and witch on a broomstick—and the black-and-white line drawings also make this test less suitable for educationally diverse populations in Europe.

Another test that was developed was for cross-linguistic purposes was the Cross-Linguistic Naming Test (CLNT [17]). The CLNT consists of a set of 40 items that have corresponding words in many languages according to the Swadesh list [286], and that are presented in the form of colored photographs. Studies with this instrument show preliminary support of its cross-cultural properties and its usefulness in assessing dementia-related naming impairments in dementia patients from Spain [158]. Ardila warned, however, that his test may have low sensitivity due to ceiling effects, which were observed in control participants across several countries [157,158]. Although the CLNT is a promising test, items with a higher difficulty level are likely needed to increase sensitivity.

Because of this issue with sensitivity, some recent efforts have focused mainly on developing naming tests using colored items that can be used in specific, local populations, such as the Argentinean Psycholinguistic Picture Naming Test [287], and the Test de Dénomination de Québec-60 images [288]. However, such an approach has limited feasibility in memory clinics characterized by marked diversity. For example, an estimated fifth of the patients visiting memory clinics in large European cities have a 'minority ethnic' background—many of them being first-generation immigrants from North Africa, the Middle East, and South America—and a substantial share of these patients have received only limited education [229]. Language-specific or local naming tests have limited use in these settings, and a widely applicable naming test was therefore identified as one of the major priorities for cross-cultural neuropsychological assessment in Europe [229].

Consequently, building on the work by Ardila with the CLNT, the first goal of this study was to develop a cross-cultural naming test that can be used to assess naming impairment in culturally, linguistically, and educationally diverse individuals. Second, we aimed to carry out a preliminary validity study of this newly developed test in a diverse European memory clinic setting. To this end, we examined 1) the convergent and divergent validity of the NAME, 2) its relationship with demographic variables, and 3) its diagnostic accuracy in discriminating patients with AD or mixed dementia (Alzheimer's with comorbid vascular cognitive impairment) from other patients visiting the memory clinic and healthy controls. Given the frequent occurrence of naming impairment in persons with AD, we hypothesized that patients with AD/mixed dementia would obtain lower scores on the NAME than patients with other diagnoses visiting the memory clinic and neurologically healthy controls.

2 Method

2.1 Development and pilot studies of the Naming Assessment in Multicultural Europe

2.1.1 Item selection

The first step in developing the Naming Assessment in Multicultural Europe (NAME) consisted of generating a comprehensive list of potential items. The initial set of stimuli included the Swadesh list, as suggested by Ardila [17], as well as items from various other sources, such as the dataset by Snodgrass and Vanderwart [289]. Regarding selection criteria, we 1) only included words that would likely be familiar to individuals from a wide range of backgrounds and 2) excluded items that would be hard to capture in a photograph, i.e. personal and demonstrative pronouns, prepositions, conjunctions, cardinal numbers and quantifiers, and adjectives. This resulted in a list of 149 potential items (nouns and verbs).

In language test design, Ivanova & Hallowell [276] recommend taking into account a large number of potentially relevant factors. We focused on age of acquisition and word frequency, as data on many of the other potentially relevant factors are not available for the languages of interest. We examined several Indo-European languages, two Semitic languages, and Turkish. Age of acquisition and word frequency data were available for English ([290,291], project Gutenberg), Dutch [291-293], Spanish ([291,294], opensubtitles.org), Polish ([291], opensubtitles.org), and Turkish ([291], opensubtitles.org). Age of acquisition data only was available for Portuguese [295,296], French [297], Italian [291], German [291,298], Swedish [291], Russian [291] and Hebrew [291]. Frequency data only was available for Arabic ([299], opensubtitles.org). As different methods were used across the age of acquisition and word frequency studies, comparing absolute values between languages was not possible. For each language, we therefore divided the set of items in half; the items that had the highest frequency and lowest age of acquisition were labeled 'easy', and the items that had the lowest frequencies and latest age of acquisition were labeled 'hard'. The words that were consistently labeled 'easy' or 'hard' across languages were subsequently selected for the following stage. This resulted in a set of 73 potential items—11 verbs and 62 nouns. The nouns could broadly be categorized into the following categories: nature, animals, colors, the body and its parts, objects, and occupations.

2.1.2 Selection of images

Subsequently, a survey was performed with the aim of selecting the photographs that best represented the target word, to ensure they were suitable for a diverse population. For all potential items (except for the colors black and white), three to four photographs were selected from open source databases and stock photography websites. The aim was to have as much variation as possible in terms of background details (i.e. isolated vs. rich context), perspective (e.g. frontal vs. profile), depiction in part vs. whole, ethnic/cultural diversity, and type of actor (e.g. animals vs. humans). The survey was distributed online through 1) the networks of the authors, 2) a professional network for culture-sensitive dementia care, and 3) a team of bicultural, bilingual interpreters. The survey was filled out by 194 respondents (mean age: 40.6, SD: 15.2). Twenty-one participants self-identified as bilingual/multilingual with a Dutch background, 21 were bilingual/multilingual participants with a diverse background (defined as being born, or having one or more parent born outside Europe), and 6 were monolingual diverse participants. These diverse participants consisted of first or second generation immigrants from North and sub-Saharan Africa, former

Dutch colonies (Indonesia, Suriname), South America (Brazil), Oceania (new Zealand), Asia (Turkey, Afghanistan, Papua New Guinea) and several countries in Europe. All other participants (n = 148) identified as monolingual individuals with a Dutch background. For each item, participants were displayed the three or four photographs simultaneously on the screen. After clicking on the image they felt best matched the target word, the survey displayed the photographs for the next item (and so on). One example item was provided to explain the goal and answer format of the survey. For the majority of the items, the same photograph was preferred by both diverse and non-diverse participants. In the seven cases of disagreement (defined by an [uncorrected] p-value on a chi-square test of < .05), we generally selected the item that was preferred by participants with a diverse background, which in six cases was the second most preferred item of the other participants.

2.1.3 Pilot study

We pilot-tested the subsequent 73-item instrument in 15 Turkish-speaking healthy controls, the majority of whom had a primary school education level or lower (73%), which, in the case of Turkey, constituted ≤five years of education. These controls were recruited in community centers and the personal network of a bicultural, bilingual neuropsychologist in training. Thirteen items were removed after this pilot stage. For eight nouns, the photographs elicited substantial response heterogeneity—e.g. 'bedroom' instead of bed; for two other nouns, the item itself often was not recognized—'anchor' and 'horn'. In addition, three verbs were removed, either because of substantial response heterogeneity—e.g. 'digging' was named 'scraping', 'working the earth' etc.—or because the actor instead of the action was named. For the verbs used in the study, ten out of 15 participants reported the verb in gerund (e.g. 'walking'), while five participants reported the verb in the third person present singular (e.g. 'walks'). Consequently, the gerund, the third person present singular, the infinitive form, and durative/continuative verb constructions (common in Dutch) were considered correct in the final test.

Figure 1. Example items of the 60-item NAME (laugh, nose, policeman, butcher).

2.1.4 Final test

The final version of the test consists of 60 items, 52 nouns and eight verbs; 31 items had easy difficulty levels based on frequency and age of acquisition data and 29 were labeled as medium or hard items (see Table 1). Some example items are provided in Figure 1. Contrary to Ardila [17] we did not present items from semantically related categories in sequence, as this may inadvertently lead to perseverative error in patients with a dysexecutive syndrome. The item order was therefore randomized. After this randomization, any successive

Table 1. Percent correct per NAME item by group.

Item		Controls (n = 39)	Other patients (n = 106)	AD/ mixed (n = 30)	p-value AD/ mixed vs. rest†	Item		Controls (n = 39)	Other patients (n = 106)	AD/ mixed (n = 30)	p-value AD/ mixed vs. rest†
Nature						Objects*					
Tree	(E)	100%	91.3%	76.7%	< .05	Boat	(E)	100%	93.4%	66.7%	< .001
Sun	(E)	92.3%	86.6%	70.0%	< .05	Book	(E)	100%	90.6%	83.3%	< .05
Moon	(E)	87.2%	75.7%	46.7%	< .01	Table	(E)	100%	97.1%	86.7%	< .05
Sea	(E)	100%	90.6%	73.3%	< .01	Chair	(E)	100%	99.0%	96.7%	n.s.
Fire	(E)	100%	84.0%	33.3%	< .001	Pants	(E)	100%	95.3%	90.0%	n.s.
						Bread	(E)	100%	95.1%	76.7%	< .01
Animals						Apple	(E)	100%	98.1%	96.7%	n.s.
Dog	(E)	100%	100%	100%	n.s.	Rope	(H)	100%	96.1%	86.7%	< .05
Fish	(E)	100%	99.0%	90.0%	< .05	Bucket	(H)	97.4%	94.2%	73.3%	< .01
Bird	(E)	100%	95.3%	90.0%	n.s.	Candle	(H)	100%	97.2%	76.7%	< .001
Ant*	(M)	94.9%	66.0%	36.7%	< .001	Football	(H)	100%	98.1%	90.0%	= .05
Snake	(H)	100%	95.2%	80.0%	< .01	Key	(H)	100%	99.1%	93.3%	n.s.
Worm	(H)	84.6%	57.3%	16.7%	< .001	Axe	(H)	94.9%	84.5%	80.0%	n.s.
						Cigarette	(H)	100%	96.2%	86.7%	< .05
Colors						Ring	(H)	82.1%	62.3%	36.7%	< .01
Red	(E)	100%	97.2%	76.7%	n.s.	Envelope	(H)	97.4%	89.6%	80.0%	n.s.
Green	(E)	100%	94.3%	73.3%	< .01	Scissors	(H)	100%	99.1%	90.0%	< .05
Black	(E)	100%	98.1%	86.7%	< .05	Match	(H)	94.9%	82.1%	53.3%	= .001
White	(M)	100%	96.2%	83.3%	< .05	Glasses	(H)	100%	100%	86.7%	< .01
Verbs						Occupations*					
Eat	(E)	97.4%	91.3%	56.7%	< .001	Doctor	(M)	100%	94.3%	53.3%	< .001
Drink	(E)	100%	99.0%	86.7%	< .01	Teacher	(M)	100%	89.6%	70.0%	< .01
Sit	(E)	94.9%	89.6%	80.0%	n.s.	Policeman	(M)	100%	83.7%	60.0%	< .01
Walk	(E)	94.9%	98.1%	80.0%	< .01	Baker	(H)	94.9%	72.6%	46.7%	< .01
Sleep	(E)	100%	98.1%	90.0%	= .05	Butcher	(H)	97.4%	78.3%	40.0%	< .001
Laugh*	(E)	100%	96.2%	86.7%	< .05	Dentist	(H)	100%	85.6%	50.0%	< .001
Swim	(H)	100%	90.6%	63.3%	< .001	Firefighter	(H)	97.4%	75.5%	16.7%	< .001
Drive*	(H)	100%	98.1%	86.7%	< .01	Chef	(H)	84.6%	61.2%	30.0%	< .001
Body and body parts											
Hair*	(E)	94.9%	92.2%	83.3%	n.s.						
Ear	(E)	100%	95.2%	83.3%	< .05						
Eye	(E)	100%	97.1%	86.7%	< .05						
Nose	(E)	100%	96.2%	76.7%	< .01						
Tongue	(E)	92.3%	87.7%	70.0%	< .05						
Foot	(E)	100%	96.1%	86.7%	< .05						
Hand*	(E)	100%	97.2%	96.7%	n.s.						
Bone	(H)	97.4%	85.8%	63.3%	< .01						
Wing*	(H)	92.3%	75.5%	46.7%	< .01						
Feather*	(H)	87.2%	76.0%	43.3%	< .001						

Abbreviations: AD = Alzheimer's Disease Dementia; NAME = Naming Assessment in Multicultural Europe

* Words and categories marked with an asterisk were newly added to items from the CLNT by Ardila [17]. Words marked with (E), (M), or (H) signify easy, medium, or hard items based on the frequency/age of acquisition database; † p-value corrected for FDR

Table 2. Demographic characteristics, cognitive test scores, and group comparisons for the whole sample.

	Controls (n = 39)	Rotterdam cohort (n = 75)	Ankara Hacettepe cohort (n = 61)	
Age	61.8 (7.2)	64.8 (12.7)	74.0 (8.5)	p < .001
Education n(%):				p = .43
Zero years of education	2 (5.1%)	9 (12%)	15 (24.6%)	
>0 but <completed primary education[1]	9 (23.1%)	12 (16.2%)	3 (4.9%)	
Completed primary education	12 (30.8%)	12 (16.2%)	19 (31.1%)	
Higher than primary education	23 (41%)	42 (55.4%)	24 (39.3%)	
Sex n (%male)	12 (30.8%)	36 (48.0%)	23 (37.7%)	p = .19
Years in the Netherlands	39 (11, n = 24)	38 (13)	-	-
RUDAS[2]	27.7 (1.8)	22.2 (5.1, n = 62)		-
3MS[2]	-	-	60.9 (22.9)	-
Diagnosis n(%)				
Subjective cognitive impairment	-	10 (13.3%)	12 (19.4%)	-
Mild cognitive impairment	-	12 (16.0%)	19 (30.6%)	
Dementia	-	20 (26.7%)	27 (43.5%)	
AD	-	8 (40.0%)	17 (63.0%)	
Mixed AD/VaD	-	(15.0%)	(7.4%)	
Other or unable to discriminate	-	9 (45.0%)	8 (29.6%)	
Psychiatric disorder	-	18 (24.0%)	2 (3.2%)	
Cognitive disorder due to other known medical condition	-	6 (8.0%)	1 (1.6%)	
Could not be determined	-	9 (12.0%)	1 (1.6%)	

Abbreviations: RUDAS = Rowland Universal Dementia Assessment Scale; AD = Alzheimer's Disease Dementia; VaD = Vascular Dementia.
Values are displayed as mean (standard deviation) unless otherwise specified.
[1] Primary education duration in the country of origin is defined according to UNESCO [85] –often five or six years
[2] The maximum score for the RUDAS is 30, with a cut-off score of <22 in diverse populations in the Netherlands. The 3MS has a maximum score of 100, and relies on normative data that is stratified by age and education level instead of a single cut-off score.

items from the same category that remained—e.g. occupations presented two times in a row—were manually rearranged. All participants were administered the test items in the same, fixed order. The items were not ordered based on the (presumed) difficulty level. In the current study, no time limits were imposed and no semantic or phonological cues were provided. Administration time varied from a few minutes (controls) up to ~20 minutes for some patients. No discontinuation rules were provided. All answers provided by the participant were recorded verbatim and items were scored correct (1) or incorrect (0). For participants with any proficiency in both Dutch and their first language, responses in either language were considered correct.

2.2 Validation study

2.2.1. Participants

One control sample and two patient samples were collected for the validation study (see Table 2 for demographic characteristics). The control sample consisted of 39 first-generation immigrants residing in the Netherlands (n = 3 from Morocco, n = 36 from Turkey). All controls were >50 years of age, free of self-reported cognitive complaints, and had a Rowland Universal Dementia Assessment Scale (RUDAS [60]) score ≥22. The first patient sample, hereafter called the 'Rotterdam cohort', was enrolled in the Netherlands at the multicultural memory clinics of the Erasmus University Medical Center in Rotterdam and the Haaglanden Medical Center in The Hague. It consisted of 75 first-generation immigrant patients, who mainly originated from Turkey (n = 29), Morocco (n = 14), Cape Verde (n = 8), Suriname (n = 7), and Iran (n = 5), in addition to ten other countries (n = 12). The

second patient sample (n = 62), or 'Ankara Hacettepe cohort', consisted of native Turkish patients and was enrolled at the Hacettepe University Medical Center in Ankara, Turkey.

2.2.2 Other measures

The neuropsychological assessment in patients of the Rotterdam cohort consisted of several tests suitable for diverse populations in Europe, such as the Cross-Cultural Dementia Screening (CCD [57]), modified Visual Association Test (mVAT [201]) and RUDAS [196]. In this test battery, language functioning was assessed with one minute semantic verbal fluency (animals and foods) and the 10-item picture naming subtest of the Recall of Pictures Test which uses colored line drawings [144]. Demographic data were collected at the neuropsychological assessment, with level of education scored according to the system of Verhage [300], with the addition of one extra level ('Verhage level 0') for patients with no education. An adapted version of the 'Language use' subscale of the Short Acculturation Scale for Hispanics (SASH [215]) was used to measure acculturation. The Ankara Hacettepe cohort was administered a different neuropsychological test battery, specific to the Turkish population in Turkey. For example, patients were administered either the 3MS version for minimally educated persons or educated persons instead of the RUDAS, as this screening test is better validated in Turkey [301].

2.2.3 Procedure

All patients in the Rotterdam cohort were referred to the memory clinic for cognitive assessment, consisting of an examination by a geriatrician or neurologist, as well as the comprehensive, culture-sensitive neuropsychological assessment (described in *Other measures).* In the majority of cases, formal interpreters (76%) or an informal interpreter (e.g. a relative, 8%) were present during the neuropsychological assessment. The NAME was administered as part of this culture-sensitive test battery used as standard clinical practice. The aim was to administer the NAME to all consecutive patients, but exceptions were made if feasibility was limited due to e.g. severe fatigue or visual impairments. Score sheets with the correct answers printed on them were available for Turkish, Moroccan-Arabic, and Dutch. For all other languages, the patients' answers were written down by the interpreter during testing and scored by consensus with the interpreter after the patient had left. All data from controls and patients were checked after data collection had finished to ensure consistent scoring across groups. Results from the neuropsychological assessment, laboratory screening with blood tests, and structural brain imaging (in a subset of patients), were discussed in a multidisciplinary consensus meeting, using the diagnostic research criteria for subjective cognitive impairment [248], mild cognitive impairment [249], and dementia subtypes (e.g. [96,250]), and the DSM-V for primary psychiatric disorders [251]. Although neuropsychologists were not blinded to patients' performance on the NAME, the diagnosis was based on the other available sources of information.

The procedure for the Ankara Hacettepe cohort was broadly similar—although no interpreters were needed for the assessment of this cohort. Diagnoses were determined in a multidisciplinary consensus meeting based on an extensive clinical evaluation including a neuropsychological assessment with tests validated in Turkey (see *Other measures)*, MRI-scans, and FDG-PET (on indication).

The control sample was assessed by a Turkish-Dutch bilingual neuropsychologist in training (with a trained interpreter present for Moroccan controls), either at their home or in a quiet room at a community center. The neuropsychologist in training was trained in test administration by a neuropsychologist with ample experience in assessing diverse populations (SF). All procedures used in this study adhere to the tenets of the Declaration of Helsinki. This study was approved by the IRB of the Erasmus Medical Center [MEC-2019-0036].

2.3 Statistical analyses

Differences in demographics between controls and the two patient cohorts were analyzed with Fisher exact tests (for the variable sex) and Kruskal-Wallis tests for age and education level, as the data was not normally distributed. We used Kuder-Richardson reliability (an equivalent of Cronbach's alpha for binary data) and Spearman-Brown split-half reliability analyses to determine the internal consistency of the NAME. NAME total scores were not normally distributed, and the analyses of convergent and divergent validity, relationship with demographic variables, and group comparisons involving the NAME total score were therefore conducted with non-parametric statistical tests. Fisher exact tests were used to test whether patients with AD/mixed dementia differed from the rest of the sample (controls and patients with other syndromes) for each of the individual 60 items of the NAME, correcting for the False Discovery Rate (FDR) using Benjamini-Hochberg adjusted p-values. As the assumptions of normality was violated for a paired samples t-test and the distribution of difference scores was asymmetrical, we used a related-samples sign test to compare the percent correct for the easy versus the medium to hard items. Spearman correlations were used to determine convergent validity with other tests measuring language (semantic verbal fluency) and with general cognitive functioning (RUDAS, 3MS), as well as to analyze divergent validity with tests measuring memory, mental speed, and executive functioning (mVAT trial 1, CCD subtests Objects A, Sun-Moon A, Sun-Moon B). To examine the relationship of the total score with demographic variables, we ran a generalized additive model using the variables sex, smooth functions of age and education, and AD/mixed dementia status across the full sample. Given the limited number of ordinal categories of the Verhage scale [300] measuring education, we used k = 6 basic functions for education; automatic smoothing parameter selection was used for age. We ran a separate model which also included smooth functions of the SASH acculturation-scores for the subset of the sample for which SASH data were available (n = 70). The ability of the NAME to discriminate between patients with AD/mixed dementia and the rest of the sample (all other patients and controls) was analyzed using (forced entry) binary logistic regression taking into account age, education, and sex. As the assumption of linearity of the logit showed a minor violation, we also ran a generalized additive model in R including smooth functions of the NAME score, age (both with automatic smoothing parameters selection), and education (k = 6), with sex as a categorical variable. Last, we ran a binary logistic regression in which we predicted AD status in AD patients versus controls only (including sex, education, and age in the model), to investigate diagnostic NAME accuracy in a more homogeneous sample.

3 Results

One patient with AD from the Ankara Hacettepe cohort was removed from the analyses as an outlier because she obtained extremely low scores on all cognitive tests, including the NAME and 3MS. The control sample and two patient samples differed significantly in age ($H = 42.2$, $p < .001$; see Table 2); controls were slightly younger than patients from the Rotterdam cohort ($U = 1073.0$, $p = .02$), who were in turn younger than the patients from the Ankara Hacettepe cohort ($U = 1302.5$, $p < .001$). There was no difference between the samples in the patients' sex ($Z = 3.42$, $p = .19$) or education level ($H = 1.7$, $p = .43$).

Across the full sample, the NAME showed excellent split-half reliability (Spearman-Brown Coefficient: 0.95); the Kuder-Richardson coefficient was similarly high (0.94). Figure 2 shows the distribution of the NAME scores across different diagnostic groups. The median total score was 59 (interquartile range [IQR]: 2) for controls, 58 (IQR: 3) for patients with subjective cognitive impairment (SCI), 55.5 (IQR: 6) for mild cognitive impairment (MCI), 55 (IQR: 4) for patients with primary psychiatric disorders such as major depression, 47 (IQR: 17) for AD/mixed dementia, and 53 (IQR: 17) for patients with other dementia subtypes. The percent correct was higher for the easy items (median percent correct: 97%) than the medium to hard items (median percent correct: 90%; $Z = -9.3$, $p < .001$). Table 1 shows the percentage of participants that correctly named each item by group. Patients with AD/mixed dementia had lower scores on 48 out of 60 items compared to the rest of the sample (controls and patients with other diagnoses combined). In AD patients, the items elicited numerous sorts of errors; patients frequently used descriptions—e.g. "small things we used to burn" for matches—and semantic paraphasias were common, e.g. "millipede" or "grasshopper" for ant. There were occasional errors in gnosis, e.g. "table" for boat.

3.1 Association with demographic variables

Higher scores on the NAME across the full sample (correcting for AD/mixed dementia status) were non-linearly associated with age (approximate $F = 4.71$, $p = .001$) and education (approximate $F = 4.82$, $p = .001$). Specifically, there was no clear relationship between age and NAME score until approximately age 70, after which more advanced age became associated with lower NAME scores; for education, higher levels of education were associated with higher NAME scores mainly for participants with a primary school education level or lower—i.e. educational attainment beyond primary school level did not seem to contribute to higher NAME scores (see Supplementary Figure 1 for smooth plots). Acculturation (measured with SASH) was not a significant predictor in the model (approximate $F = 0.92$, $p = .34$), nor was sex ($t = 1.72$, $p = .09$; see Supplementary Figure 2 for smooth plots).

3.2 Convergent and divergent validity

The NAME was significantly correlated with other measures of language as measured by semantic verbal fluency and the naming subtest of the Recall of Pictures Test (see Table 3). In addition, there was a significant correlation with the score on the RUDAS (Rotterdam cohort and controls) and the 3MS (Ankara Hacettepe cohort). Regarding divergent validity, lower NAME scores were significantly associated with worse memory performance (mVAT and CCD objects test A) and reduced mental speed (CCD Sun-Moon test A), but there was no significant association with executive functioning (CCD Sun-Moon test B).

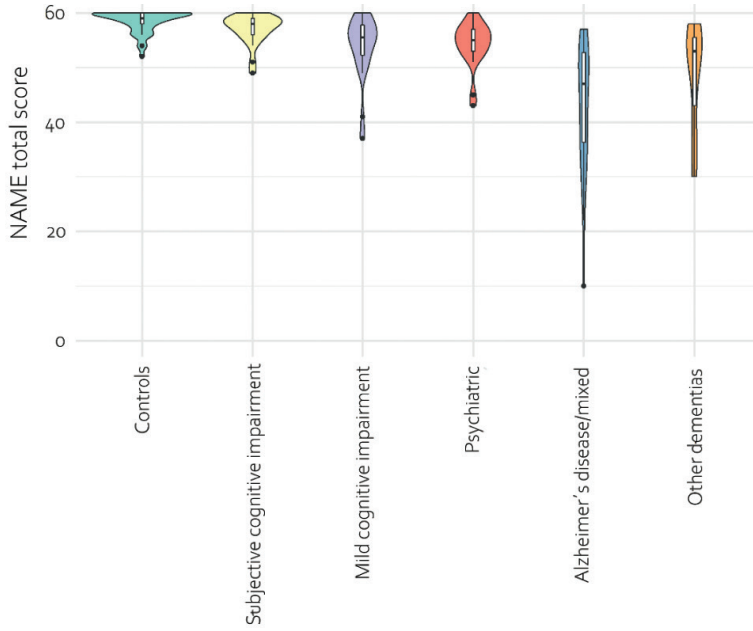


Figure 2. Violin plot of the NAME scores by diagnosis type.

Table 3. Correlations between NAME total score and tests measuring similar (convergent validity) and dissimilar (divergent validity) cognitive domains

	N	ρ	p-value
Convergent validity			
Animal fluency	154	.73	< .001
Foods fluency	93	.58	< .001
Naming subtest of Recall of Pictures Test	69	.35	.004
RUDAS	99	.68	< .001
3MS	60	.82	< .001
Divergent validity			
Modified Visual Association Test	61	.54	< .001
CCD Objects test A	60	.61	< .001
CCD Sun-Moon test A	62	−.35	.005
CCD Sun-Moon test B	59	−.25	.06

Abbreviations: 3MS = Modified Mini-Mental State Examination; CCD = Cross-Cultural Dementia Screening; RUDAS = Rowland Universal Dementia Assessment Scale

3.3 Preliminary validity analyses of the NAME

A first analysis of the predictive validity of the NAME score was carried out using binary logistic regression, correcting for the demographics age, education level, and sex. The model as a whole predicted 45% (Nagelkerke R^2) of the group status (AD/mixed dementia vs. all other patients and controls) and correctly classified 86% of all cases. NAME score (B = −.106, p < .001, OR = 0.90 [95% CI = 0.85-0.95]) and age (B = .100, p < .01, OR = 1.11 [95% CI = 1.04–1.18]) were significant predictors of group status. Education level and sex did not significantly predict group status in the model. The model had an acceptable fit (Hosmer-Lemeshow χ^2 = 5.05, p = .75, see Supplementary Figure 3A–3C for probability plots). The AUC of the full model was 0.89. Running these analyses as a generalized additive model did not notably change the results (see Supplementary Figure 4 for smooth plots). In a

standalone model without accounting for demographic characteristics, the AUC for the NAME total score was 0.88, with a sensitivity of 70% and specificity of 91% at the optimal cut-off score of <50. In a subsample of AD patients and controls only, NAME scores showed near-perfect classification rates (classification accuracy: 95%; Nagelkerke R^2: 90%).

4 Discussion

The aim of this study was to develop a naming test that is suitable to detect naming impairment in culturally, educationally, and linguistically diverse individuals. In addition, we provided preliminary data on its reliability and validity in a diverse European memory clinic setting. We carried out a multistage pilot study, in which 73 items that showed a homogeneous age of acquisition and word frequency across multiple languages were selected from an initial pool of items. We piloted several photographs per item to select the most suitable image, and pilot-tested this 73-item version in a sample of healthy diverse controls. The final 60-item version of the NAME was used in a (preliminary) validity and reliability study. The NAME showed promising reliability, convergent validity, and diagnostic accuracy in detecting naming impairment in diverse memory clinic patients. With regard to divergent validity, NAME scores were correlated with performance in memory and mental speed, but not with executive functioning; either naming impairment also affected memory performance and (naming) speed on the Sun-Moon test, or impairments in these cognitive domains co-occurred in this patient population.

Few naming tests are currently available that use culture-sensitive, colored items to assess patients from a wide range of backgrounds, and this (preliminary) diagnostic accuracy study showed that the NAME has the potential to detect naming impairment in such diverse settings. Previous studies in diverse populations using the CLNT [17] and the Recall of Pictures Test of the European Cross-Cultural Neuropsychological Test Battery [144] highlighted issues with sensitivity/ceiling effects and limited diagnostic accuracy—the CLNT had a specificity of 94.6%, but sensitivity of only 58.3% [158] and the naming subtest of the Recall of Pictures Test displayed a very modest AUC of .65 (controls vs. dementia). The NAME may have benefited from the addition of a number of relatively difficult items (such as occupations) as compared to the CLNT and a more substantial length in comparison to the rather brief naming subtest of the Recall of Pictures Test. In its current form, the NAME is relatively long in comparison to other instruments (RPT: 10 items, CLNT: 40 items, MINT: 32 items). For future research and clinical purposes, the NAME might be shortened by removing items that lack sensitivity/specificity in discriminating between controls and specific patient populations (e.g. patients with AD, temporal lobe epilepsy, or stroke). In addition, the items may now be arranged in order of increasing difficulty based on the data collected in this study, including a discontinuation rule for the assessment of patients with AD. Last, future studies should consider adding a time limit for each item (e.g. 20 seconds) to examine whether this may further improve sensitivity.

Patients from the memory clinic cohorts with AD/mixed dementia scored significantly lower on the majority of the individual items than controls and other patients, and the NAME total scores likewise were lowest for those with AD/mixed dementia. Patients with AD/mixed dementia made different kinds of errors, such as semantic paraphasias, descriptions, and—occasionally—errors in gnosis. Patients with other diagnoses had

more variable scores, intermediate between patients with AD/mixed dementia and controls. This is likely due to the inclusion of patients with AD-(co)pathology in this group, such as a number of patients with Lewy body dementia—in whom AD-copathology has been associated with lower naming test scores [302]. In addition, this sample contained patients whose dementia subtype could not be determined, e.g. because the severity of the dementia made it impossible to determine a cognitive profile, who may have had AD/mixed dementia. These difficulties in determining the dementia subtype are common in diverse individuals in Europe, in which dementia diagnosis can be challenging [75,270].

Performance on the NAME was non-linearly associated with age and education, but was not associated with sex or level of acculturation. Such non-linear effects of age on naming abilities are well-established, with little longitudinal change in individuals in their 50s and 60s, but a more notable decline in the seventh and eighth decades of life [303]. Similarly non-linear effects were found for education; that is, receiving one or more additional years of education has more impact on the test performance of individuals without any formal education than those who are already highly educated. Although education was associated with NAME scores, it was not a significant predictor of AD status above and beyond NAME scores. Combined with the lack of association with acculturation, this indicates that the NAME may be an especially promising instrument in a culturally and educationally diverse memory clinic setting—although it would be worthwhile to collect additional data to confirm there is no difference in performance by nationality/ethnicity.

This study has several strengths. First, the items that were selected were specifically chosen to reflect diversity at an international level, with a similar relative age of acquisition and word frequency across a number of languages. This was followed by an extensive pilot testing phase and analysis in a substantial number of diverse patients. Another strength was that the neuropsychological assessments of patients took place in memory clinics with ample experience in working with diverse populations using culturally appropriate cognitive tests. In addition, most of the assessments in the Dutch multicultural memory clinics were carried out in the presence of interpreters who received specific training in interpreting during neuropsychological assessments.

Some limitations should be acknowledged. The interpreters assisted in determining whether a non-standard answer was a correct synonym or an incorrect answer, particularly for local language dialects that are not formally written and for which no formally published lexicon is available, such as regional dialects within the Tamazight language (Moroccan-Berber). However, as interpreters were used for all patients and not just the AD/mixed AD patients, it seems unlikely that this would have significantly influenced our results. Ideally, all patients would be assessed by a neuropsychologist with a similar cultural and linguistic background, but unfortunately, the current situation in Europe is far removed from this ideal due to a lack of diversity in the workforce of neuropsychologists [229]. Second, the pilot study and control sample consisted predominantly of Turkish persons residing in the Netherlands, and more normative data across age and education will have to be collected before this test can be implemented in clinical practice. This may subsequently result in a (more) comprehensive list of acceptable synonyms mentioned by controls to guide decisions on whether items should be considered correct or incorrect in clinical practice. Third, as mentioned above, a subset of the patients could not be diagnosed; the percentage

of these patients without a conclusive diagnosis was similar to the percentage reported in another study in a similar population [201].

In addition to the collection of more comprehensive normative data, future studies should be conducted in other European countries to confirm its applicability in these contexts, such as through the European Consortium on Cross-Cultural Neuropsychology [232]. Furthermore, future studies may aim to extend our findings to multicultural populations with anomia due to other medical conditions, such as acquired brain injury. Furthermore, it would be interesting to study the effects of fluency in, and attrition of, the first and second language on NAME performance. In the current study, participants were allowed to answer in both their first or second language; future research should examine how naming in the first or second language may affect the diagnostic accuracy of the NAME, as first and second languages may differentially deteriorate over time in neurodegenerative diseases [304]. Additionally, it would be interesting to study differences in performance on the NAME noun items versus NAME verb items across different diseases, as noun and verb naming may be differentially impaired in some diseases (e.g. [305,306]). A number of additional verb naming items may be helpful to provide a more in-depth analysis of verb naming in patients who specifically show impaired verb naming on the NAME. Last, follow-up studies may examine the types of errors made in more detail, as well as relevant qualitative aspects of language production, such as naming speed, that are increasingly studied in cross-cultural language paradigms such as word fluency tasks (e.g. [307]).

In conclusion, the NAME is a promising new instrument to assess naming impairment in culturally, educationally, and linguistically diverse individuals, such as diverse patients visiting European memory clinics. Next steps are the collection of normative data and a more extensive study of the instrument's validity to ultimately implement this instrument in clinical practice.

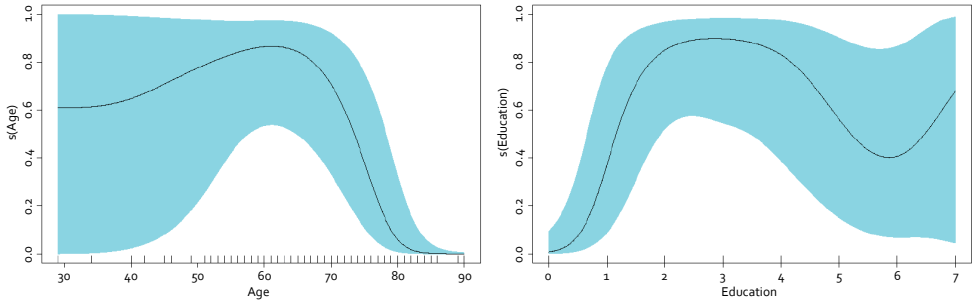
Acknowledgments & funding
The authors would like to thank Amy van Hattem for drafting the age-of-onset and frequency database, Dalila Oulel for her contributions to the Moroccan-Arabic score sheets, and Reyhan Özcan, Margot van der Zee, and Fatma Karagöz for their contribution to the data collection in healthy controls. This work was supported by the Netherlands Organisation for Health Research and Development [grant number: 733050834]. The authors report no conflicts of interest.

Chapter 3.3 Supplementary material

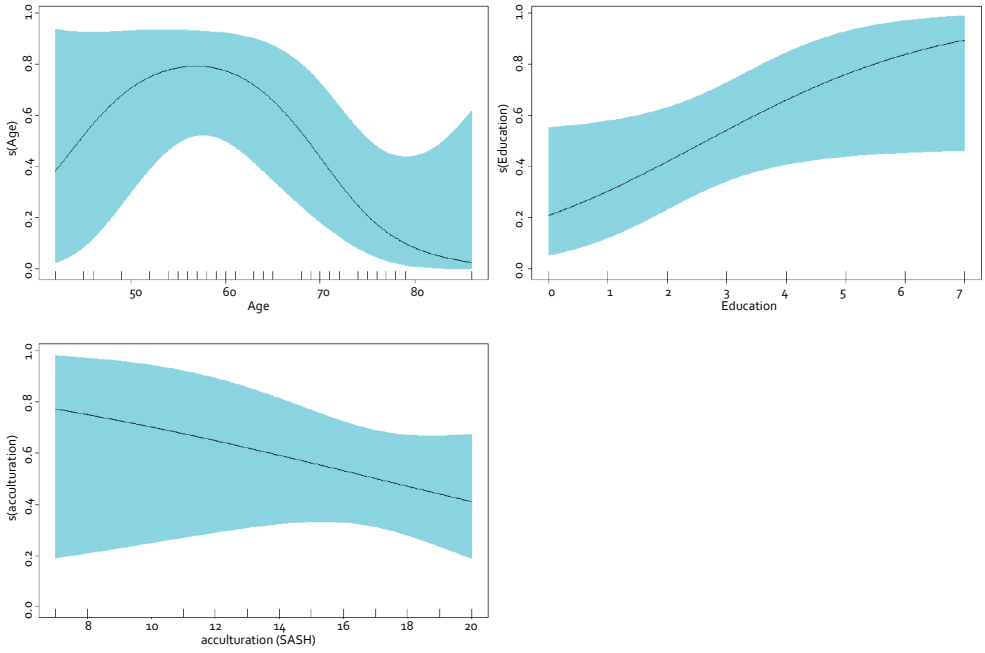
Supplementary Table 1. Mean word frequency for the NAME items

Item	Difficulty	Mean AoA	Mean word frequency	Item	Difficulty	Mean AoA	Mean word frequency
Nature				Objects*			
Tree	(E)	2.5	22,500	Boat	(E)	3.2	25,980
Sun	(E)	2.5	37,937	Book	(E)	3.4	43,527
Moon	(E)	3.2	21,217	Table	(E)	3.2	36,037
Sea	(E)	3.6	40,513	Chair	(E)	2.7	20,190
Fire	(E)	3.6	58,309	Pants	(E)	3.1	6,988
Animals				Bread	(E)	2.9	15,583
Dog	(E)	2.3	27,739	Apple	(E)	2.6	4,236
Fish	(E)	2.7	14,783	Rope	(H)	4.9	6,960
Bird	(E)	2.8	11,285	Bucket	(H)	4.3	1,775
Ant*	(M)	3.4	1,078	Candle	(H)	3.9	4,678
Snake	(H)	3.6	4,374	Football	(H)	4.6	5,054
Worm	(H)	4.6	2,658	Key	(H)	3.8	11,456
Colors				Axe	(H)	5.2	2,693
Red	(E)	3.4	29,334	Cigarette	(H)	6.7	5,594
Green	(E)	3.5	26,003	Ring	(H)	3.7	13,934
Black	(E)	3.5	42,557	Envelope	(H)	4.9	70,244
White	(M)	4.0	61,483	Scissors	(H)	3.5	1,351
Verbs				Match	(H)	4.6	11,906
Eat	(E)	2.2	36,423	Glasses	(H)	3.5	4,842
Drink	(E)	2.5	20,395	Occupations*			
Sit	(E)	2.8	19,012	Doctor	(M)	4.3	24,187
Walk	(E)	2.5	34,574	Teacher	(M)	4.1	11,834
Sleep	(E)	2.3	35,044	Policeman	(M)	4.9	35,112
Laugh*	(E)	3.0	18,786	Baker	(H)	4.5	2,153
Swim	(H)	3.7	3,869	Butcher	(H)	7.1	1,626
Drive*	(H)	4.7	14,914	Dentist	(H)	5.0	1,388
Body and body parts				Firefighter	(H)	4.4	842
Hair*	(E)	3.3	30,504	Chef	(H)	6.3	3,616
Ear	(E)	2.4	16,918				
Eye	(E)	2.5	39,177				
Nose	(E)	2.4	12,391				
Tongue	(E)	3.8	14,474				
Foot	(E)	3.0	32,003				
Hand*	(E)	2.7	130,127				
Bone	(H)	4.2	5,612				
Wing*	(H)	5.5	5,180				
Feather*	(H)	4.4	2,773				

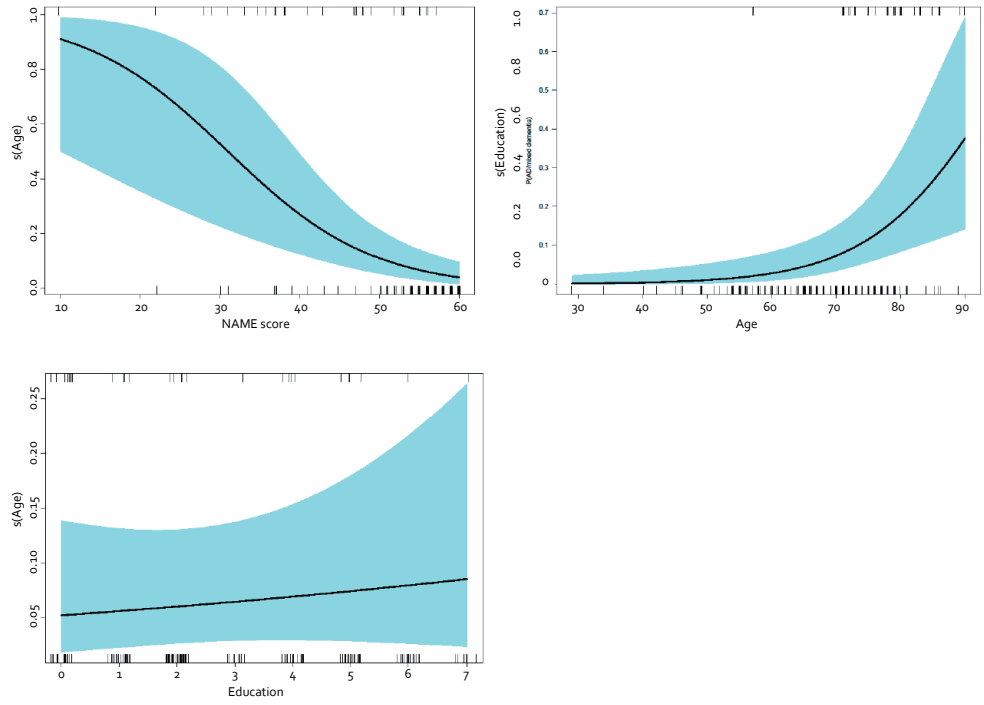
Abbreviations: AoA = Age of acquisition
* Words and categories marked with an asterisk were newly added to items from the CLNT by Ardila [17]. Words marked with (E), (M), or (H) signify easy, medium, or hard items based on the frequency/age of acquisition database.



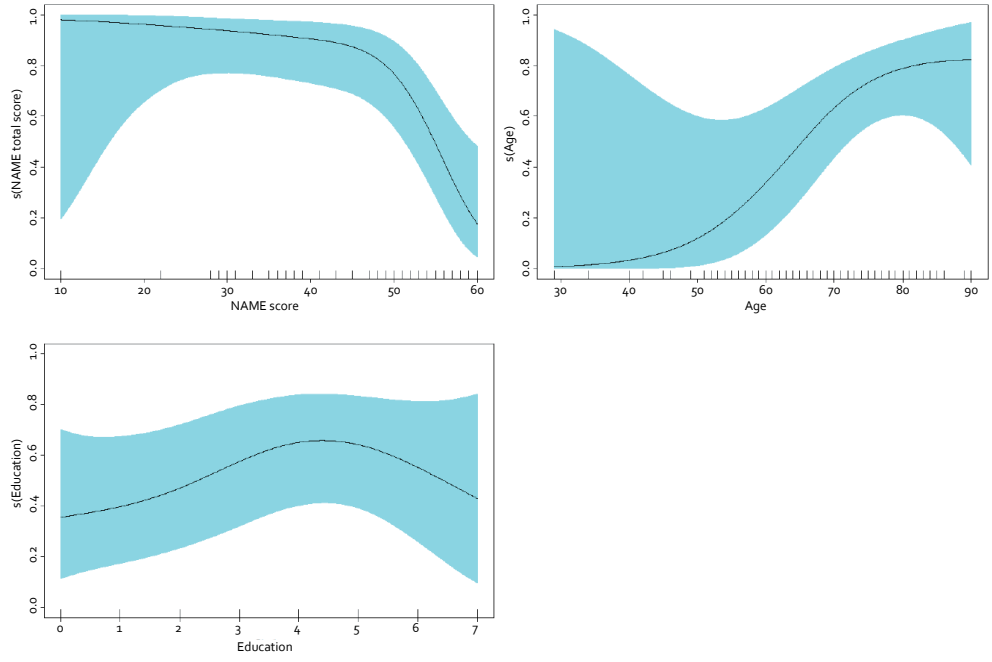
Supplementary Figure 1. Smooth plots for GAM-model predicting NAME-score using age (left), education (right), and sex.



Supplementary Figure 2. Smooth plots for GAM-model predicting NAME-score using age (upper left), education (upper right), sex, and SASH-acculturation score (lower left).



Supplementary Figure 3A-C. Probability plots for NAME score (upper left), age (upper right), and education (lower left) in the binary logistic regression model.



Supplementary Figure 4. Smooth plots for GAM-model predicting AD/mixed dementia status using NAME-score (upper left), age (upper right), education (lower left), and sex.

CHAPTER 3.4

Caregiver burden in a culturally diverse memory clinic population: the Caregiver Strain Index-Expanded

Sanne Franzen
Willem S. Eikelboom
Esther van den Berg
Lize C. Jiskoot
Judy van Hemmen
Janne M. Papma

Dementia and Geriatric Cognitive Disorders, 2021. Advance online publication.

Abstract

Introduction:
Although qualitative studies have highlighted substantial barriers to dementia diagnosis and care in culturally diverse populations in Europe, quantitative studies examining the level of caregiver burden in these populations have been lacking thus far and are urgently needed.

Methods:
We compared the caregiver burden levels on the Caregiver Strain Index (CSI)-Expanded of 63 culturally diverse patient-caregiver dyads from a multicultural memory clinic with 30 native Dutch patient-caregiver dyads and examined the association between caregiver burden and determinants of burden.

Results:
Informal caregivers in the multicultural memory clinic cohort experienced a high level of caregiver burden (mean CSI-score multicultural cohort: 6.1 [SD: 3.3]; mean CSI-score native Dutch cohort: 4.8 [SD: 3.2]). Burden was significantly associated with impairment on proxy-rated and objective measures of cognitive functioning, such as the Informant Questionnaire on Cognitive Decline and the Rowland Universal Dementia Assessment Scale, and with instrumental activities of daily living. Burden was the highest in spousal caregivers. The positive subscale of the CSI-Expanded provided limited additional information.

Conclusions:
Caregivers of culturally diverse patients experience a high level of caregiver burden, in particular at more advanced disease stages. This study highlights the need to screen culturally diverse caregivers in European memory clinics on caregiver burden to identify those in need of caregiver support.

1 Introduction

Over the past century, European countries have become increasingly diverse. In these diverse populations—particularly in migrant populations from Asia and Africa—the prevalence of dementia is higher than in older adults born in Europe [44], likely due to a higher prevalence of risk factors for dementia, such as cardiovascular disease, diabetes, and limited cognitive reserve. Dementia care in these groups is often viewed as a responsibility of the family [55,308], and caregivers may fear losing the respect of the wider family or social network if they do not provide care to the person with dementia [309]. In addition, there are numerous barriers to dementia diagnosis and care in these populations [308,310,311]; therefore, formal dementia care services are often accessed only when the level of caregiver burden becomes exceptionally high [308].

Traditional caregiver burden instruments mainly focus on aspects of care that can increase the level of burden, such as increased emotional strain; however, preliminary studies in culturally diverse caregivers of persons with dementia in the Netherlands suggest that positive aspects of taking care of a family member—such as appreciation expressed by the wider social network—may balance out some of the "negative" effects in these culturally diverse populations [55,309]. An instrument is therefore needed that covers both these positive and "negative" aspects. To that end, Al-Janabi et al. [312] developed an extended version of the Caregiver Strain Index [313], adding 5 items measuring "positive" aspects of care that may decrease caregiver burden. Some factors that may influence burden scores are caregiver characteristics [314,315], patients' neuropsychiatric symptoms [316-318], functional impairment—particularly in instrumental activities of daily living (iADL [316,317,319])—and objective cognitive impairment.

Given the increasing numbers of culturally diverse individuals with dementia in Europe, the goal of this study was to determine the level of caregiver burden in these caregivers and examine the relationship with these potential determinants of burden.

2 Methods

2.1 Participants

We included 63 caregiver-patient dyads from the outpatient multicultural memory clinic of the Erasmus MC University Medical Center in Rotterdam, the Netherlands. The patients were first-generation immigrants from Turkey (n = 27), Morocco (n = 14), Suriname (n = 7), Cape Verde (n = 4), and other countries (n = 11). In addition, we included 30 native Dutch patient-caregiver dyads from the outpatient memory clinic of the Erasmus Medical Center.

2.2 Procedure

All patients were referred to the memory clinic for cognitive evaluation and underwent a comprehensive clinical evaluation, after which they were discussed in a multidisciplinary meeting (see [201]). Patients were diagnosed according to established research criteria for dementia subtypes [96,249,250] or the 5th Edition of the Diagnostic and Statistical Manual of Mental Disorders for primary psychiatric disorders [251]. Imaging biomarkers (CT or MRI) to support the diagnosis were collected in 73% (46/63) of the culturally diverse patients; imaging data were available less often in culturally diverse individuals diagnosed with

primary psychiatric disorders (3/9 patients, 33%) and subjective memory complaints (SMC; 8/13 patients, 61%). Lumbar punctures were only used on indication (5%). Based on the intake interview with the informant, the clinician scored the level of functional impairment in basic activities of daily living (ADL [320]) and iADL [321]). The CSI-Expanded and other informant-based measures were administered to the caregivers in a separate room while the patients underwent neuropsychological testing. Caregivers could choose between the Dutch or an adapted Moroccan-Arabic or Turkish version of the CSI-Expanded. Caregiver-patient dyads were included between January 2019 and January 2021. The majority of the native Dutch caregivers (90%) was recruited as part of a study about neuropsychiatric symptoms in memory clinic patients with specific requirements on the minimum amount of time the caregiver spent with the patient.

2.3 Measures

2.3.1 Caregiver Strain Index-Expanded

The CSI-Expanded [312] is an extended version of the original 13-item Caregiver Strain Index [313]. The original 13-item instrument covers aspects such as emotional strain, physical demands, and time constraints, with a cutoff score of ≥7/13. The CSI-Expanded contains 5 additional items that focus on aspects of caregiving that may decrease burden, such as the patient showing appreciation of the care provided by the informal caregiver. Although the original study totaled the subscale scores (i.e. with a total score between −5 and 13), we followed Kruithof et al. [322] in analyzing both scales separately to determine the added value of the positive subscale. The Dutch CSI-Expanded was previously translated and validated [312]; in the current study, Moroccan-Arabic and Turkish versions were developed following the translation recommendations by the International Test Commission [167], with forward and backward translation and a subsequent evaluation and revision by a team of bicultural, bilingual native speakers to evaluate the cultural and linguistic appropriateness of the translations.

2.3.2 Other informant-based measures

Caregivers of the multicultural memory clinic cohort (as well as a subset of caregivers in the native Dutch cohort) filled out the short version of the Informant Questionnaire on Cognitive Decline (IQCODE [242]); the IQCODE aims to capture cognitive decline and consists of 16 items rated on a 5-point Likert scale. The average score on all items is calculated, resulting in a final score ranging from 1 (marked improvement in cognitive functioning) to 5 (marked decline). For older first-generation immigrants in the Netherlands, a cutoff score of 3.8 was determined to be optimal [243]. In addition, we collected information on the caregivers' sex and the type of relationship to the patient.

2.3.3 Cognitive, functional, and neuropsychiatric measures (patients)

All patients in the multicultural memory clinic underwent a neuropsychological assessment, which consisted predominantly of cognitive and behavioral measures that have been validated in culturally diverse populations in the Netherlands, such as the Cross-Cultural Dementia screening (CCD [57]), modified Visual Association Test (mVAT [201]), and Rowland Universal Dementia Assessment Scale (RUDAS [196]). The CCD covers the domains of memory (Objects tests A and B), mental speed (Dots test A and Sun-Moon test A), and executive functioning (Dots test B and Sun-Moon test B). The modified Visual Association Test is a test of visual association memory consisting of colored photographs. The RUDAS

is a cognitive screening test specifically designed for use in culturally, linguistically, and educationally diverse populations and similar to the MMSE in its scope and administration time, with an optimal cutoff of <22/30 for culturally, linguistically, and educationally diverse individuals in the Netherlands [196]. In addition, patients filled out the Dutch, Turkish, or Moroccan-Arabic 15-item Geriatric Depression Scale (GDS-15 [244,245]). Acculturation was measured with a shortened, adapted Short Acculturation Scale for Hispanics (SASH [215]), consisting only of the four "Language use" items, in which we substituted "Spanish" with the first language of the patient and "English" with "Dutch". Clinicians rated patients on the ADL and iADL scales. Patients in the native Dutch cohort were administered a different neuropsychological test battery which included the MMSE [198].

2.4 Statistical analysis

Differences in demographic characteristics between native Dutch and multicultural memory clinic participants were analyzed in R with χ^2 tests for nominal data and t tests for continuous data. To compare burden levels on the original CSI, we ran a robust linear regression in which we corrected for sample differences in patients' sex and relationship status. We did not correct for differences in the patients' educational attainment as these reflect existing disparities in educational attainment in the general population [51]. As the positive subscale showed substantial skewness and the native Dutch cohort was modest in size, no meaningful group comparison could be carried out on the positive subscale while correcting for sample differences in sex and relationship status. We therefore used a Mann-Whitney U test (uncorrected for sex and relationship status) to analyze group differences on the CSI-Expanded positive subscale. We used Pearson correlations (or nonparametric equivalents) to determine the relationship between caregiver burden and its possible determinants. We corrected for multiple testing using the false discovery rate (FDR) based on Benjamini-Hochberg adjusted p values. ANOVA (or a nonparametric equivalent) was used to compare caregiver burden levels by relationship type and across dementia stages—SMC, mild cognitive impairment, and dementia.

3 Results

Fifty-eight culturally diverse caregivers filled out the Dutch version of the CSI-Expanded, while four preferred the Turkish version and one the Moroccan-Arabic version. Three culturally diverse caregivers were accidentally administered the original CSI—these caregivers remained in the analyses of the original CSI, but were excluded from the analyses of the CSI-Expanded positive subscale. The native Dutch cohort contained relatively more spousal caregivers compared to the multicultural memory clinic cohort (see Table 1). Table 2 shows the characteristics of the patients included in the sample. The patients from the multicultural memory clinic had a lower education level than native Dutch patients. In addition, the native Dutch sample contained more male patients. The patient groups did not differ in age or diagnoses.

3.1 Level of caregiver burden

In the multicultural memory clinic cohort, 29 (46%) caregivers scored above the original CSI cutoff score of ≥7 based on the 13 original items, in comparison with eight (27%) native Dutch caregivers. After correcting for sample differences in relationship type and patients' sex, caregivers in the multicultural cohort experienced significantly higher levels of caregiver

burden (original CSI) than the native Dutch cohort (t = 2.48, p = .01). The native Dutch and multicultural memory clinic cohort did not differ in their CSI-Expanded positive subscale score (U = 795.0, p = .30). A substantial proportion of the caregivers showed a maximum score on this subscale (multicultural memory clinic n = 43 (67%)) and native Dutch cohort n = 16 (53%)). Ceiling effects were particularly present for items 14 and 18 of the positive subscale ("I am happy to care for him/her" and "Taking care of him/her is important to me"). In the multicultural cohort, the positive and negative scales were highly correlated (r = −.58, unadjusted p < .001). There was a medium to large correlation in the Dutch cohort (r = −.39, unadjusted p = .03), which remained significant after adjusting for FDR.

Table 1. Caregiver characteristics and scores on the Caregiver Strain Index-Expanded

	Multicultural memory clinic cohort (n = 63)	**Native Dutch cohort (n = 30)**	**Significance**
CSI-Expanded informant			
Spouse n(%)	8 (13%)	24 (80%)	p < .001
One or more adult child(ren) n(%)	49 (78%)	3 (10%)[1]	
Other n(%)[2]	6 (10%)	3 (10%)	
Sex n males (%)	18 (29%)[3]	6 (20%)	n.s.
CSI-Expanded score			
Score on the negative items (original scale)	6.1 (3.3)	4.8 (3.2)	p = .01*
Score on the positive items[4]	−5.0 (1)	−5.0 (1.25)	n.s.
Distribution of positive subscale scores:	−5: 67% −4: 13% −3: 17% −2: 3% −1: 0%	−5: 53% −4: 23% −3: 20% −2: 3% −1: 0%	

Values are displayed as mean (standard deviation) unless otherwise specified.
* P-value after correcting for sample differences
[1] One adult child verified his answers with the spouse of the patient
[2] For example, second-degree relative, friend, neighbor, parent
[3] Two CSI-Expanded were filled out by two informants of different sexes (e.g. brother and sister)
[4] Median (IQR); A "yes" on an item of the original scale is scored as 1, and a "yes" to an item on the positive subscale is scored as −1; a "no" is scored as 0 on both scales.

Table 2. Patient demographic characteristics, cognitive test scores, and diagnosis of the patients

	Multicultural memory clinic cohort (n = 63)	**Native Dutch cohort (n = 30)**	**Significance**
Age	70.9 (10.5)	73.1 (8.4)	n.s.
Education level n(%):			
0 years of education/illiterate	17 (27%)	0 (0%)	p < .001
1 year of education up to primary education	27 (43%)	0 (0%)	
> primary education	19 (30%)	30 (100%)	
Sex n males (%)	25 (40%)	23 (77%)	p = .001
Number of years in the Netherlands	41.6 (10.6)	-	-
RUDAS	21.2 (5.0; n = 57)	-	-
IQCODE	4.0 (0.6; n = 55)	3.7 (0.5; n = 14)	-
MMSE	19.4 (3.8; n = 17)	23.9 (5.7; n = 21)	-
Diagnosis n(%)			
Subjective memory complaints	13 (21%)	5 (17%)	n.s.
Mild Cognitive Impairment	9 (14%)	8 (27%)	
Dementia	19 (30%)	12 (40%)	
Primary psychiatric disorder (e.g. depression)	9 (14%)	1 (3%)	
Cognitive disorder due to other known medical condition (e.g. epilepsy)	4 (6%)	2 (7%)	
Could not be determined	9 (14%)	2 (7%)	

Abbreviations: RUDAS = Rowland Universal Dementia Assessment Scale; IQCODE = Informant Questionnaire on Cognitive Decline; MMSE = Mini Mental State Examination
Values are displayed as mean (standard deviation) unless otherwise specified.

3.2 Relationship of CSI-Expanded with patient demographics, cognitive and functional impairment, and depression

In the multicultural cohort, there were no correlations between the original CSI and patient demographics (education level, sex, years living in the Netherlands, and SASH acculturation score) or self-reported depressive symptoms (GDS-15). The scores on the original CSI showed moderate positive correlations with the level of impairment in iADL ($r = .38$, $p < .01$), but not with impairment in basic ADL ($r = .22$, $p = .10$). In terms of cognitive impairment, higher scores on the original items of the CSI were strongly associated with more severe cognitive impairment on the short IQCODE ($r = .59$, $p < .001$) and moderately with more impaired general cognitive functioning (RUDAS, $r = -.33$, $p = .01$) and memory performance (mVAT, $r = -.40$, $p = .02$, CCD Objects test B, $r = -.28$, $p = .04$). There were no significant correlations with CCD measures of mental speed or executive functioning. After correcting for FDR, only the associations with the IQCODE and iADL remained statistically significant. The positive subscale did not show any significant correlations after correcting for FDR.

3.3 Caregiver Strain Index in relation to relationship type and patient diagnosis

Spousal caregivers, adult children, and "other" caregivers of culturally diverse patients experienced different levels of caregiver burden on the original CSI (see Fig. 1; $F = 4.4$, $p = .02$). Post hoc analyses (corrected for FDR) revealed a higher level of spousal caregiver burden (mean CSI: 8.6, SD: 1.7) in comparison with both adult children (mean CSI: 5.9, SD: 3.3; $p = .04$) and "other" caregivers (mean CSI: 3.7, SD: 3.2; $p = .02$). The scores on the positive subscale were similar across relationship types ($H = 3.7$, $p = .16$).

There were also significant differences in caregiver burden by dementia stage (Fig. 2; $F = 5.9$, $p = .02$). Post hoc analyses (corrected for FDR) revealed that caregiver burden was higher in caregivers of persons with dementia than persons with SMC (mean difference: −3.07, $p = .04$), while the other comparisons were not significant. The scores on the positive subscale were similar across dementia stages ($H = 0.5$, $p = .8$).

4 Discussion

In this study, we found that informal caregivers of culturally diverse patients experience a high level of caregiver burden as evidenced by the substantial number of individuals scoring above the cutoff on the CSI; these burden levels were associated with dementia severity on proxy-rated and objective cognitive measures, as well as functional measures, and with relationship type. Contrary to our expectation, the positive subscale of the CSI-Expanded provided little additional information.

This study demonstrated that caregiver burden levels in caregivers of culturally diverse patients are high, in line with other studies investigating caregiver burden in neurodegenerative disease (e.g. [323-326]). Several factors may contribute to these high levels of burden. In the early stages of dementia, it is common for one person in culturally diverse families to serve as the primary caregiver [55]. As dementia symptoms progress, this primary caregiver may increasingly dedicate their time to caring for the person with dementia, giving up on their own personal activities and social life, which can subsequently result in isolation of the caregiver [327]. The strong feelings of filial or religious duty

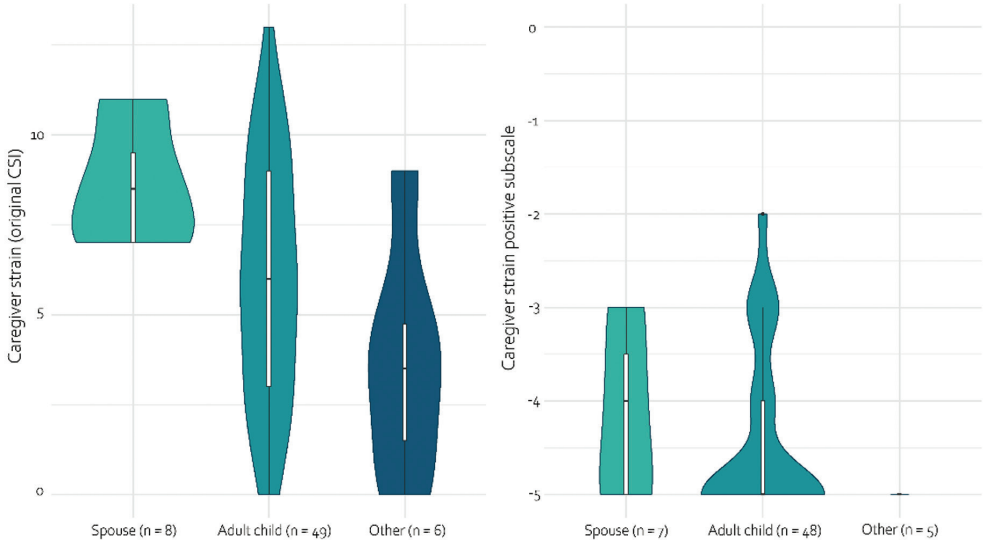


Figure 1. Scores on the original CSI and positive subscale by relationship type in multicultural cohort.

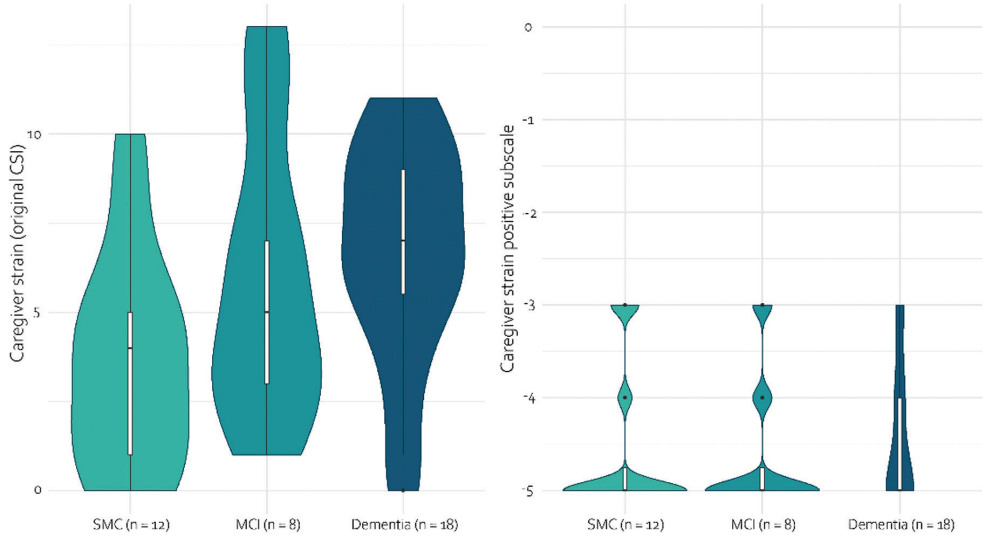


Figure 2. Scores on the original CSI and positive subscale by dementia stage in multicultural cohort.

experienced by these primary caregivers may motivate them to continue to provide informal care despite increasing levels of burden [327]. We therefore recommend general practitioners and memory clinics to routinely monitor caregiver burden and arrange subsequent intercultural caregiver support if necessary.

We found associations between burden levels and (proxy-rated and objective) measures of cognitive functioning, functional impairment, dementia stage, and relationship type. This is in line with previous studies, although some studies in less culturally diverse populations have found weak or no correlations between objective measures of cognition and caregiver

burden [328]—possibly due to the inclusion of more severely cognitively impaired patients in these study samples (e.g. [314,329]). Regarding relationship type, spousal caregivers experienced significantly higher levels of burden than adult children and "other" caregivers. The levels of burden in adult children showed substantial variation. Previous studies in less culturally diverse populations suggest that burden may be influenced by different mechanisms across different caregiver roles; for example, adult children may experience particular uncertainty over the future, such as "increased worry over how long they can maintain their level of caregiving in addition to other responsibilities" [330]. Such differences require further study and should be addressed in caregiver support strategies.

Somewhat contrary to our expectation, there was little variation in the scores on the positive subscale and no clear correlations with possible determinants of caregiver strain. Kruithof et al. [322] similarly found limited added value of the positive subscale in a sample of caregivers of stroke patients and suggested modifications to the items or answer format or the use of a different instrument. It may also be interesting to examine whether the addition of this subscale may improve the overall user experience of caregivers filling out this questionnaire—for example, caregivers may feel more comfortable discussing burdensome aspects of care if such topics are alternated with more positive factors.

This study has several strengths. It was carried out in a specialized multicultural memory clinic, in which the staff has ample experience in assessing patients with culturally diverse backgrounds. In addition, we were able to include individuals from a wide variety of cultural, educational, and linguistic backgrounds. For example, over two-thirds of the patients included in the study received little formal education. We used several instruments and questionnaires that were previously validated in culturally, linguistically, and educationally diverse elderly in the Netherlands, such as the IQCODE, RUDAS, CCD, and mVAT, ensuring a valid assessment of cognitive impairment. Some limitations should be acknowledged. This was a retrospective analysis of data collected in routine clinical care, and the study lacked information on some potential determinants of caregiver burden (e.g. caregivers' education level). Furthermore, it was not possible to examine the association between caregiver burden and neuropsychiatric symptoms other than depression in our multicultural memory clinic cohort, given that no validation studies have been carried out on instruments such as the Neuropsychiatric Inventory [331] in culturally diverse populations in the Netherlands. Last, although both native Dutch and culturally diverse caregivers on average scored close to the cutoff score for dementia on the MMSE and the RUDAS, respectively—indicating that they likely had similar levels of cognitive impairment—we could not formally compare the level of cognitive and functional impairment in these two populations because of the different instruments used across groups. Therefore, we were unable to examine whether or not the differences in caregiver burden between native Dutch and culturally diverse individuals are perhaps in part attributable to differences in the level of cognitive and functional impairment between these groups.

In conclusion, this study highlights that caregiver burden levels in caregivers of culturally diverse patients in the multicultural memory clinic are high, and general practitioners and memory clinics should actively monitor and subsequently arrange support for those caregivers experiencing severe levels of caregiver burden.

Acknowledgments and funding
The authors would like to thank Amy den Teuling and Daphne Pol for their contribution to the data collection in caregivers of native Dutch patients. This work was supported by The Netherlands Organisation for Health Research and Development (ZonMw Memorabel) (Grant No. 733050834 and 733050823). The funder did not play a role in any part (initiation, design, analysis, interpretation, writing of the report, or decision to submit) of this manuscript.

Conflict of interest
The authors have no relevant financial or nonfinancial interests to disclose.

Statement of Ethics
IRB approval was obtained from the IRB of the Erasmus MC University Medical Center in Rotterdam (MEC-2019- 0036 and MEC-2020-0341). The procedures used in this study adhere to the tenets of the Declaration of Helsinki. Dutch native caregivers provided written informed consent, while written informed consent was waived for culturally diverse individuals by the IRB of the Erasmus MC University Medical Center.

CHAPTER 4

IMPLEMENTATION: DIVERSITY IN CLINICAL PRACTICE AND RESEARCH

	Semantisch	
	Dieren	**Eten**
1.	شیر leeuw	نان Brood
2.	گربہ kat	گوشت vlees
3.	ٹائیگر tijger	پیچ pisrache
4.	ڈیوک Duif	کباب kebab
5.	گورخر zebra	چاول rijst
6.	موش muis	گندم boter
7.	کاکروچ kakkenlak	جام jam
8.	گارنیکل garnacl	پنیر Kaas
9.	De sprikhaan	انڈے eieren
10.	ہزار duizendpoot	بادام bannan
11.	ڈائنوسار Dynasores.	کیوی kiwi
12.		ڈیٹ Dadel
13.		یوگرٹ Yogurt
14.		
15.		
16.		
17.		
18.		
19.		
20.		
21.		
22.		
23.		
24.		
25.		
26.		
27.		
28.		
29.		
30.		
31.		
32.		
33.		
34.		
35.		

Totaal = 11 Totaal = 13

CHAPTER 4.1

Cross-cultural neuropsychological assessment of adult and elderly memory clinic patients

Cross-culturele neuropsychologische diagnostiek bij volwassenen en ouderen op de geheugenpoli

Sanne Franzen
Esther van den Berg
Janne M. Papma

Tijdschrift voor Neuropsychologie, 2021, 16(1), 14–28

Abstract

Since the 1950s, the Netherlands has been characterized by increasing diversity, starting with labor migration in the 1960s and 1970s, followed by the independence of Suriname in 1975 and the influx of asylum seekers and refugees in the 1980s. Neuropsychological assessment of persons with a diverse background is challenging due to factors such as language, culture, and education. In this paper, we give an overview of these challenges; in addition, we provide recommendations for clinical practice based on the scientific literature as well as our experiences in the multicultural memory clinic of the Alzheimer Center of the Erasmus MC.

1 Introduction

Over the past century, the Dutch population has become increasingly diverse. This increasing diversity started with the independence of current-day Indonesia in the late 40s of the twentieth century and was followed by a period of labor migration from countries such as Turkey and Morocco—but also from Italy and Spain—between 1950–1974. As it was originally expected that these labor workers would return to their countries of origin, little attention was paid to their integration and to proficiency in the Dutch language in this population [170]. This period of labor migration was followed by the independence of Suriname in 1975, after which a large share of the Surinamese population (consisting of, among others, Hindustani, Creole, Javanese, and Chinese Surinamese individuals) moved to the Netherlands in the following five years. In the 80s and 90s, refugees and asylum seekers came to the Netherlands from several different countries, such as Iran, Iraq, Armenia, Eritrea, and Somalia, but also Aramaic peoples (from across the Middle East) and persons from the former Yugoslav Republic. Recent years have seen an influx of refugees from Syria, as well as seasonal labor workers from within the European Union.

As these individuals age, the likelihood of developing cognitive impairment increases. Some subpopulations are at a higher risk of developing cognitive complaints due to the higher prevalence of conditions such as stroke [47], diabetes [47], and dementia [46]. Neuropsychologists working in memory clinics will therefore increasingly encounter patients with a diverse background in their clinical practice. Neuropsychological assessment of patients with a diverse background requires neuropsychologists to modify their approach, their selection of instruments, and their subsequent reporting. We previously illustrated this using a case study from our own clinical practice [332]. Having suitable instruments in itself is not sufficient for a sensitive neuropsychological assessment of individuals with a culturally, linguistically, and educationally diverse background. For example, neuropsychologists need to take into account different contextual cultural factors that may influence neuropsychological assessment, such as migration history, acculturation, the testing situation, and communication styles (interested readers may read the paper by Fujii [13] for a more detailed elaboration of these factors in the ECLECTIC framework).

In this article, we will elaborate on some of these factors within the Dutch context and will provide practical tools for each subsequent step in the neuropsychological assessment based on the scientific literature and our own experiences in the multicultural memory clinic of the Erasmus MC. This outpatient clinic was founded in 2015 with the explicit goal of improving neuropsychological assessment of culturally, educationally, and linguistically diverse patients. The subsequent TULIPA study (2017–2021) was the starting point for a large-scale national collaboration aiming to compile a sensitive neuropsychological test battery for diverse populations. Even with these suitable tests, however, no one-size-fits-all method exists for cross-cultural neuropsychological assessment. It is important to recognize that the variation within groups can be as large, or much larger, than the variation between groups.

2 Neuropsychological assessment

2.1 Preparing for a culture-sensitive neuropsychological assessment

In preparing for a neuropsychological assessment, it is important to verify whether someone prefers to be assessed in Dutch, and if not, what language/dialect the patient prefers to be assessed in. For example, several dialects of Arabic are spoken across North Africa and the Middle East that are not interchangeable. A Moroccan patient—in addition to speaking Dutch—may speak Darija (Moroccan-Arabic), one of several Berber languages (that until recently were not used in writing), but also French or Spanish. Bilingual or multilingual patients may perceive some languages as more prestigious than others and may indicate that they are equally proficient in these languages even when this is not necessarily the case. Ask patients which language they used to speak at home when they were growing up, as well as the language they currently speak with family, friends, and acquaintances. Even for patients who speak Dutch well, such as many Surinamese patients, it may be worthwhile to ask about their preferred language. A sensitive way to do so can be: "I know that several different languages are spoken within Suriname and that some persons of Surinamese descent (even) speak multiple languages. Can you tell me a little more about what the situation is/was like for you?".

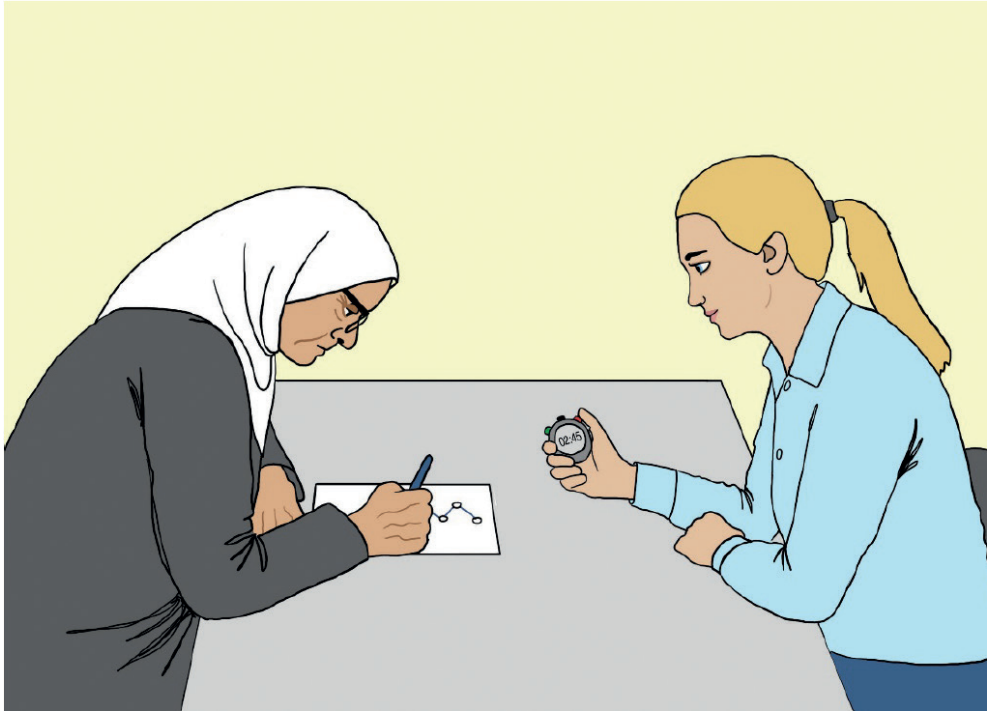
After determining the preferred language, an interpreter can be hired—provided there is funding available. Subsequently, patients may be informed about their appointment. It is important to take into consideration issues of literacy in some patients—for example, 80%–90% of female Turkish and Moroccan first-generation labor immigrants did not complete any formal education [51]. A letter about the appointment may be hard to read for some, even if it is written in the preferred language. It can therefore be worthwhile to send information materials that are written in easy-to-understand language (level B1 or lower), preferably accompanied by images. One such example is the information booklet about the diagnostic trajectory of dementia [333], in which the neuropsychological assessment ('Neuropsychologisch onderzoek') is explained using an image (see Figure 1).

In our experience, it is helpful to call patients (or their caregivers) in advance to ask whether they have any questions about their appointment. First, this will create an opportunity to verify whether the letter describing the appointment was received and understood. In addition, this provides an opportunity to prepare the patient and/or caregiver—who often do not know what to expect from a neuropsychological assessment—for the duration of the assessment and any practical aspects, such as the need to bring reading glasses. Furthermore, it may be useful to inform patients about the (possible) presence of a professional interpreter. We often mention that the interpreter is present as an extra set of ears and eyes for the neuropsychologist, e.g. to monitor any changes in language; we explicitly let the caregiver know that the interpreter is not present because we question their capabilities as an interpreter.

2.2 The interpreter

It is important to instruct the formal interpreter on the procedure and the purpose of the neuropsychological assessment before starting the assessment. In some cases, the interpreter may not have prior experience with interpreting in this specific context. For example, in legal settings that often rely on formal interpreters from interpreter agencies,

it is essential that no room is left for interpretation; frequently repeating questions or information may be justified in this setting, whereas repeating information in memory tests may decrease their validity in neuropsychological assessment. At a minimum, it is advisable to inform the interpreter that you are trying to measure changes in (e.g.) language and memory and that it is therefore vital to interpret as literally as possible—including any mispronunciations or linguistic errors of the patient—and to not repeat any information unless asked to do so during the assessment. It can also be helpful to show the test materials to the interpreter in advance. Asking the patient for permission regarding the presence of the interpreter is important as well—if possible without the interpreter being present. Some communities in the Netherlands are small and close-knit—the Cape Verdean community in Rotterdam for example has settled down in one particular neighborhood of the city—and it may be possible that the interpreter and patient know each other (through acquaintances), which in turn may lead the patient to feel less at ease.



Neuropsychological assessment
Yasmina has brought her glasses and hearing aid. That is important.
Yasmina will be examined to find out why she has problems remembering things.
This is called a neuropsychological assessment.
First, Yasmina has to answer questions.
Then she has to do exercises and puzzles.
For example, Yasmina has to remember pictures.
Some tests are difficult.

Figure 1. Example image and explanatory text (translated from Dutch) included in the information booklet of the Alzheimer Center Erasmus MC.

In case of insufficient funding for a professional interpreter, neuropsychologists may have to resort to using informal interpreters, most often a relative. A professional interpreter, however, is strongly recommended in the international literature [229], as this allows the patient to communicate directly with the neuropsychologist—if necessary without the informal caregiver present, in case of sensitive topics such as (severe) mood symptoms. Furthermore, the informal interpreter may leave out information provided by the patient because he/she may believe the information to be irrelevant, or because of feelings of shame. If the assessment takes place with an informal interpreter, it is even more important to explain the goal and procedure of the neuropsychological assessment, and what is being expected from the informal interpreter.

2.3 History taking

The questions from the 'Cultureel Interview' [334] or Cultural Formulation Interview [335] are a useful tool for the history taking interview, as they may help bridge the gap between individuals from different cultures. In our experience, it can be beneficial to learn more about geographical aspects of the country of origin of the patient and address this topic in the conversation; by doing so, you immediately show your interest in the patient. The cultural interview provides the opportunity to learn about the wishes, needs, and customs of the patient, but also allows for an exploration of the words used by patients and caregivers to describe their complaints. No proper, neutral terminology exists for some concepts, like dementia and depression, in several languages; instead, descriptions are used, or terms are used with negative connotations such as going 'crazy'—e.g. *bunama(k)* instead of the less widely known *demans* or *Alzheimer* in Turkish.

Obtaining information about cognitive functioning as pertaining to daily activities in elderly individuals with a diverse background can sometimes be challenging, because they may always have been dependent on others for certain activities—e.g. for financial administration in persons who are illiterate or insufficiently fluent in Dutch—or because these responsibilities are carried out by someone else due to reasons other than cognitive impairment, such as because of physical conditions. In history taking, it can be useful to ask about specific activities that are relevant to patients with specific characteristics. For example, a patient who practices Islam may be asked about their visits to the mosque, e.g. whether they are able to find their way there (orientation), whether they can keep up with what the Imam is saying (attention/mental speed), or whether they recognize people from the community (gnosis). Some additional knowledge may be required. For practicing Muslims, it may be useful to inquire whether patients are able to correctly perform prayer rituals, and, more specifically, whether they are able to remember how many times they have repeated certain actions—the Islamic prayer ritual is characterized by different numbers of iterations depending on the time of day. In contrast, many patients with early-stage memory impairment due to Alzheimer's disease are often able to reproduce Quranic texts without any difficulty. Regarding orientation to time, it may be useful to ask specific questions about Fridays, as this is an especially important day of the week for Muslims—comparable to Sundays for practicing Christians. Although such knowledge is by no means required, it may be helpful in order to better understand specific cognitive processes.

Asking direct questions about mood symptoms or other (neuro)psychiatric symptoms, and subsequently discussing these symptoms freely, may not be as evident in some cultures

(particularly in the presence of relatives). In some cases, it may be helpful to first ask about more 'acceptable' symptoms, such as fatigue, pain, or a tense feeling in the muscles or body, before exploring whether patients may sometimes experience that their head is 'full of thoughts', or 'being worried' or 'feeling sad'. It is important to take into consideration that subtle nuances are often lost in translation.

2.4 Selecting the test materials

In addition to exploring someone's cultural and linguistic background, it is of great importance to thoroughly examine a person's education and their level of literacy, because it is widely known that (il)literacy strongly impacts performance on neuropsychological tests (see e.g. [91]). It is useful to not only look at the formal level of education, but also take into consideration the quality of the education. For example, in our outpatient clinic, we sometimes assess persons with a primary school education level who are nevertheless not fully able to read and write. The reverse is also observed: patients who initially did not have access to any formal education, but learned (some) reading and writing skills later in life. We are therefore developing a tool to measure literacy in our multicultural memory clinic. In addition, it may be useful to inquire about literacy in a way that does not induce feelings of shame. For example, it may be helpful to formulate the question in the following way: "many individuals have not had the opportunity to go to school, what was that like when you were growing up?". By phrasing the question in this way, not having received any education is made the norm. In applying normative data, it is necessary to take into account that the (former) duration of primary school education may be different from the Netherlands; according to the Turkish system, five years of *İlköğretim Okulu* is equal to a Verhage level 2, whereas *(Madrasa) Ibtidaiya* in Morocco equals six years.

Based on the education level, and in particular literacy skills, a cognitive testing protocol can be selected. A relatively large number of cognitive tests is available in Europe for some cognitive domains, such as memory [229], while, in essence, not a single suitable test is available yet for some other cognitive domains, such as social cognition and language (naming). Table 1 provides an overview of some cross-cultural neuropsychological tests that are used by experts across Europe in neuropsychological assessment [229]. For some cognitive domains, in particular social cognition, it remains to be seen whether it is possible to test this function in a cross-cultural way: even individuals born in different countries across Europe differ in their abilities to recognize (supposedly universal!) emotional facial expressions [261]. Studies from abroad have also indicated that performance validity tests such as the TOMM may not be valid in individuals from other cultures than those these tests were originally designed for and validated in.

In selecting the instruments, it is important to be aware of specific elements of cognitive tests that may not be suitable for all patients. For example, the use of tests with black-and-white line drawings should best be avoided in patients with low education levels, as it has been found that these populations less accurately name [28,29] and remember [201] such stimuli. In addition, abstract symbols and elements that require skills learned in school (reading/writing/arithmetic) should probably be avoided. Furthermore, it is useful to know that doing something as fast as possible and to the best of someone's abilities may be perceived as mutually exclusive by people from some cultures; it is either the one or the other [12]. Last, non-verbal (intelligence) tests, that may at first glance seem suitable

to overcome language barriers, are often strongly influenced by someone's cultural and educational background [180].

Table 1. Overview of some of the cross-cultural neuropsychological tests used and published in Europe

Test	Cognitive function(s)
Rowland Universal Dementia Assessment Scale (RUDAS)	Global screening
Multicultural Cognitive Examination	Global screening
Adapted Mini Mental State Examination	Global screening
Adapted Montreal Cognitive Assessment	Global screening
Cross-Cultural Dementia Screening (CCD)	Extensive screening
European Cross-Cultural Test Battery:	Extensive test battery
RUDAS	Global screening
Recall of Pictures Test (RPT)	Memory
Enhanced Cued Recall	Memory
Semi-complex figure	Memory
Picture Naming (RPT)	Language
Animal and supermarket fluency	Language/executive functioning
Color Trails Test	Attention/executive functioning
Five Digit Test	Attention/executive functioning
Serial threes	Attention/executive functioning
Copying of simple figures	Visuoconstruction
Clock Drawing Test	Visuoconstruction
Clock Reading Test	Visuospatial functioning
TNI-93	Memory
TMA-93	Memory
WHO/UCLA adaptation RAVLT	Memory
Modified Visual Association Test	Memory
TFA-93	Executive functioning
Cross-Linguistic Naming Test	Language
Stick Design Test	Visuoconstruction

*Tests that were still under development in 2020: EMBRACED battery, literacy screening tool, cross-cultural naming test.

In using questionnaires to measure factors such as mood, anxiety, or coping alongside cognitive functioning, it is important to keep in mind that direct, literal translations often do not do justice to the measurement properties of the original questionnaire. Both the exact translation of the concepts in the questionnaire and the meaning of these concepts may differ across cultures. For some questionnaires, such as the 15-item Geriatric Depression Scale [244], versions are available that have been thoroughly studied and validated in diverse populations in the Netherlands. Informal translations of questionnaires of which the validity and reliability have not been studied should be used with caution.

2.5 The administration of test materials, use of norms, and interpretation

Patients who have never been in a formal testing situation before may have limited *test-wiseness*. Some patients may not understand why a neuropsychological assessment is necessary or how it may contribute to the overall diagnostic trajectory. Sometimes, it may be helpful to explain the assessment through a suitable metaphor, such as by comparing the neuropsychological assessment with a checkup of your car ('APK'). By explaining that we will examine all the parts—including those that, at first glance, seem to be functioning well—it may sometimes be possible to confer to patients why it is important to undergo seemingly simple or 'childish' neuropsychological tests. In our experience, it may be helpful

to explicitly mention before testing commences that some patients may state that they are not 'crazy' and that although tests may not look difficult, they are of great importance to us. Providing an explicit example of how such tests may go wrong, such as a patient who does not perceive half of the stimuli on a page due to hemispatial neglect, may help clarify to a patient why such tests are necessary.

During testing, it may frequently be unclear whether patients have understood all the instructions correctly. In such cases, it may be helpful to use the feedback method (described in more detail by Pharos [336]). Instead of merely asking whether someone has understood the instruction—to which most patients will respond with a polite "yes"—it can be useful to ask them the following question instead: "I want to make sure I explained everything well. Could you please summarize what I have just told you?". This method stresses the responsibility of the professional instead of questioning the comprehension capabilities of the patient and provides direct opportunities to determine which aspects may require clarification.

In terms of scoring and reporting of test results, we would like to refer to our suggestions regarding reporting findings for patients with a migration background in our previous case study in *Tijdschrift voor Neuropsychologie* [332]. In addition, there are two aspects we would like to elaborate on. First, it is important to take into consideration that the exact date of birth and therefore the patient's age may not be known for all patients—in those cases, the date of birth will often start with January first or July first. It may be useful to keep a broader age range in mind when applying normative data. Second, accurate normative data may not be available for a patient with a specific combination of (demographic) characteristics. In such cases, norms for other populations that are as comparable as possible are often used, such as lower educated elderly Turkish individuals instead of lower educated elderly Moroccans. It is important to describe and provide a rationale for the use of these norms in the report. In addition, it may be useful to obtain a better sense of the influence of different countries of origin by comparing the patient to multiple different norms—e.g. by comparing an Iraqi patient with both Dutch, Turkish, and Moroccan normative data. If the differences between these groups are substantial, it is likely that the interpretation of the scores requires even more caution.

2.6 Reporting back to the patient

Neuropsychologists who are trained in the Netherlands have often learned to primarily focus on the individual patient's needs, wishes, and motives in reporting back to the patients. Some patients, however, may come from a collectivist culture, in which group harmony may be valued above the individual. For example, at the department of Neurology of a large hospital in Ankara, Turkey, a dementia diagnosis was often shared with one or more of the patient's relatives, who were entrusted with the task of informing the patient in the way the family saw fit. It is clear that this is different from the guidelines and norms in the Netherlands. In addition to differences in individualism vs. collectivism, different explanatory models may be used—in addition to or instead of the biomedical perspective that is commonly held in the Netherlands [54,218]. In the case of dementia, cognitive impairment may be explained from a spiritual perspective, such as being possessed by evil spirits, as well as from a perspective of 'normal' aging, or it may be related to having experienced a physically and/or mentally straining life. By inquiring about the way patients

or caregivers explain their symptoms, it may be possible to provide psychoeducation that better matches the patient's beliefs and knowledge about brain diseases. Last, the way in which bad news is delivered to patients, such as an unfavorable prognosis, may differ between cultures. Some patients may feel that they are treated unnecessarily brusquely or rudely when they are informed that there is no hope of a cure; some patients may indicate that there is always the possibility of a (divine) miracle happening, after all. Similarly, in many cultures, it is common to 'read between the lines' (also known as high-context communication), and direct communication may be perceived as unnecessarily hurtful. In general, we advise to adhere to our professional code of ethics and provide the patient with the necessary information, while remaining aware of these differences in communication styles.

Explaining specific cognitive functions or processes may be difficult because of language barriers between the neuropsychologist and the patients and/or caregiver, but may also be hindered by the abstract nature of some neuropsychological concepts. A suitable metaphor to explain specific cognitive processes may prove useful. For example, a soccer coach who is guiding a team, or a mother taking care of a large family, may be used as examples to explain frontal or executive functioning, while a highway metaphor—with accidents at major traffic junctions or the use of alternative routes to circumvent roadblocks—might be used as a metaphor for mental speed and the way the brain handles vascular damage.

3 Conclusion

The neuropsychological assessment of culturally, educationally, and linguistically diverse patients in the memory clinic may benefit from specific techniques or methods, such as 1) adequate preparation by inquiring about language and (quality of) education in a sensitive way, 2) by properly instructing the interpreter, and 3) providing information in a sensitive way before and after the assessment. To embed culture-sensitive practices into routine clinical neuropsychological practice, the field may benefit from a quality standard for neuropsychological assessment of diverse individuals (for example, in collaboration with the Central Commission on Diversity & Psychology and Neuropsychology Section of the NIP, as well as the Dutch Association for Neuropsychology (NVN)), similar to the recently implemented 'Generieke Module Diversiteit' [337] in mental health care.

CHAPTER 4.2

Cross-cultural neuropsychological assessment in Europe: position statement of the European Consortium on Cross-Cultural Neuropsychology (ECCroN)

Sanne Franzen
on behalf of the European Consortium
on Cross-Cultural Neuropsychology

The Clinical Neuropsychologist, 2021. Advance online publication

Abstract

Introduction:
Over the past decades European societies have become increasingly diverse. This diversity in culture, education, and language significantly impacts neuropsychological assessment. Although several initiatives are under way to overcome these barriers—e.g. newly developed and validated test batteries—there is a need for more collaboration in the development and implementation of neuropsychological tests, such as in the domains of social cognition and language.

Methods:
To address these gaps in cross-cultural neuropsychological assessment in Europe, the European Consortium on Cross-Cultural Neuropsychology (ECCroN) was established in 2019.

Results:
ECCroN recommends taking a broad range of variables into account, such as linguistic factors, literacy, education, migration history, acculturation and other cultural factors. We advocate against race-based norms as a solution to the challenging interpretation of group differences on neuropsychological tests, and instead support the development, validation, and standardization of more widely applicable/cross-culturally applicable tests that take into account interindividual variability. Last, ECCroN advocates for an improvement in the clinical training of neuropsychologists in culturally sensitive neuropsychological assessment, and the development and implementation of guidelines for interpreter-mediated neuropsychological assessment in diverse populations in Europe.

Conclusions:
ECCroN may impact research and clinical practice by contributing to existing theoretical frameworks and by improving the assessment of diverse individuals across Europe through collaborations on test development, collection of normative data, cross-cultural clinical training, and interpreter-mediated assessment.

In this position paper, we provide a general overview of the challenges to and status of cross-cultural neuropsychological assessment in Europe, and subsequently present the standpoints and potential impact of the newly formed European Consortium on Cross-Cultural Neuropsychology (ECCroN). These standpoints reflect the emerging scientific evidence in cross-cultural neuropsychology in Europe as well as the combined clinical and research experience of the individual consortium members.

1 Europe: a continent with unique challenges to neuropsychological assessment

Over the past the past decades European societies have become increasingly diverse. Following decolonialization in the second half of the twentieth century, inhabitants of former European colonies immigrated to European countries, such as North Africans in France and Afro-Caribbean and South-Asian people in the United Kingdom [170]. Furthermore, in times of economic prosperity, European countries have traditionally relied upon a (low-educated) labor force recruited in countries outside and within the European Union to carry out low-skilled labor work [170,192]. In addition, refugees and asylum seekers have fled to Europe from the 1980s onwards [192,338]. Combined with those individuals born in European countries, including indigenous minorities such as the Sámi in the northern parts of Norway, Sweden, and Finland, Travelers in Ireland and the United Kingdom, and Romani people throughout Europe, this makes for a strikingly culturally and educationally diverse European population. Furthermore, Europe is characterized by remarkable linguistic diversity. In addition to the languages spoken by those from outside Europe, many of Europe's inhabitants are bilingual or multilingual, and foreign language learning is part of the school curriculum in most European countries [339]. Many countries have multiple official languages and/or nationally recognized/(co-)official dialects; for example, in Spain, people may speak Spanish, as well as other languages such as Catalan, Basque, Galician [340], or several other dialects.

Several of these populations—but particularly the "guest workers" who immigrated to Europe between 1950–1974—are at risk of developing cognitive impairment, due to a higher prevalence of medical conditions such as diabetes [47], stroke [47], hypertension [49] and dementia disorders [44]. Inevitably, neuropsychologists will therefore increasingly encounter culturally, linguistically, and educationally diverse individuals in clinical practice. However, this diversity significantly impacts neuropsychological assessment, and assessment practices therefore need to be adapted to suit these diverse populations—a need that has been internationally recognized (e.g. [222]).

A recent Delphi study [229] revealed that several initiatives are under way to address some of the most urgent issues in adult cross-cultural neuropsychology in Europe, such as the development of memory tests and screening tools that may support a culture-sensitive cognitive assessment. One important tool to result from these initiatives is the European Cross-Cultural Neuropsychological Test Battery [140]. The Delphi study also highlighted a need for more collaboration in the development, publishing, and implementation of neuropsychological tests developed in Europe, as well as a need for more research in the domains of social cognition and language in particular. In addition, it revealed pressing matters regarding training clinicians in cross-cultural neuropsychological

assessment and working with interpreters in interpreter-mediated assessments. However, these issues are not specific to adult cross-cultural neuropsychology, as the field of pediatric neuropsychology faces similar—as well as unique—challenges to cross-cultural neuropsychological assessment.

To address these gaps in cross-cultural neuropsychological assessment in Europe, ECCroN was established in late 2019 by 16 specialists from ten countries; founding consortium members represent the Netherlands, Denmark, Norway, England, Scotland, France, Spain, and Italy, as well as two specialists from the United States of America and the State of Palestine working on multinational projects with one or more European site(s). ECCroN is currently actively reaching out to specialists working with pediatric and adult diverse populations in other European countries and invites others working in European contexts to join the consortium by reaching out to the consortium members. ECCroN convenes in monthly to bimonthly web-based video conferences, as well as at European conferences, such as the biennial meeting of the Federation of European Neuropsychological Societies (FESN). In the next paragraphs, we outline the main standpoints of ECCroN.

2 Towards a broad definition and measurement of diversity: ECCroN recommends taking into account lifetime demographics and contextual factors

ECCroN proposes a broad definition of diversity in neuropsychology; instead of just studying those born in different countries or those of different ethnic groups, ECCroN recommends taking a broad range of variables into account, such as linguistic factors (e.g. dialect, age of second language acquisition), literacy, education, migration history, acculturation and other cultural factors, as well as other relevant social determinants of health (see e.g. [22]). For example, most neuropsychological tests have been developed for educated people and may not be suitable for individuals with low literacy skills, regardless of their country of origin. In addition, norms that are representative of the cultural, educational, and linguistic diversity in Europe are lacking for most tests. In the United States, many informative group level variables may be drawn from state or national databases, such as relevant indicators of educational quality—the length of the school term, the average number of school days the student attended, and the student to teacher ratio [22]. These factors provide additional value alongside traditional self-report variables, such as urban versus rural location of the school and whether single primary school lessons comprised children of several ages. As such regional or national data are often unavailable in Europe—even more so in those who immigrated from countries outside of Europe—collaborative approaches are needed to better measure and take diversity-related variables into account. A step in this direction would be to explore whether these variables are currently being measured by researchers and clinicians, and if so, how they are operationalized. This may lead to a recommended set of variables to consider in research and clinical assessment of diverse individuals across Europe—containing, for example, suitable measures of acculturation or educational quality. The ECCroN consortium members have started working towards this goal by structurally taking inventory of whether/how each of the aspects in the ECLECTIC framework [13] are currently being measured by the consortium members. This framework encompasses Education (level, quality, literacy), Culture and acculturation, Language (spoken and proficiency in the majority language), Economic issues, Communication

style, the Testing situation (including comfort and motivation), the conceptualization of Intelligence, and the Context of immigration.

3 ECCroN supports the use of widely applicable cross-cultural tests test as opposed to race-based norms

A recent study in a small sample of European experts in cross-cultural neuropsychology indicated that appropriate norms were not available for some—and sometimes even for none—of the available tests used in the clinic, or that norms were only available for some populations [229]. Race-based norms, which are commonly—but controversially [41]—applied in countries such as the United States, are not widely used in Europe. This should be seen in light of the historic events in Europe during World War II and subsequent European policies: EU member states generally have strict legislation regarding data collection by race—and to a lesser degree ethnicity—to prevent discrimination [341].

Given the marked diversity in Europe and the controversial nature of race-based norms, ECCroN advocates against race-based norms as a solution to the challenging interpretation of group differences on neuropsychological tests. Instead, ECCroN aims to focus on the development, validation, and standardization of more widely applicable/cross-culturally applicable tests. Many traditional neuropsychological tests, such as the Trail Making Test, are unsuitable for diverse populations, due to their reliance on school-based skills such as reading and writing (in the Latin alphabet) and the culturally-specific abstract reasoning skills they require [34,229,342]. In addition, tests emphasizing speed may be less suitable as cultural differences exist in time perception [20]—e.g. a good result may be considered as contingent upon a slow, thorough process [12,343]. Applying race-based norms on tests in which floor effects are likely to occur due to factors other than cognitive impairment will preclude valid conclusions on true cognitive functioning. For example, healthy Turkish immigrants who are illiterate often show floor performance on common tasks of visuoconstruction, such as Clock Drawing and figure copy [32]; applying a normative correction to such a performance would make it hard to document impaired tests performance and will inadvertently lead to the misclassification of persons with cognitive impairment as cognitively normal. In addition, caution should be exercised before using tests that are proclaimed to be "culture-free" but that have never been studied in diverse populations—the use of such nonadapted "culture-free" test may lead to diagnostic mistakes or misclassifications [145,344]. For example, applying Spanish or British norms for the Raven's Colored Progressive Matrices—a test that has historically been labeled "culture-free" due to its minimal linguistic requirements—to a sample of normally developing Moroccan children resulted in a substantial number of children being classified as having "below average" or "impaired" intelligence [344].

ECCroN therefore supports the use of more widely applicable, cross-cultural tests. Such tests should tap into the same cognitive ability in individuals across different cultures (construct validity) and be psychometrically sound, e.g. show clear differences in performance between persons with and without cognitive impairment (no floor effects). In addition, the influence of cultural factors such as acculturation on test performance should be minimal. These tests should use widely applicable stimuli instead of culture-specific ones—e.g. avoiding items like the igloo, pretzel, and beaver used in the Boston

Naming Test [18,19]—and should not rely on (school-based) skills that patients likely never acquired. The concepts and instructions of these widely applicable tests should be clear and easy to understand, even for those who are not used to being tested, i.e. have limited "test-wiseness" [178,231]. Although we recognize that standardized testing is important, ECCroN recommends to actively create an environment in which diverse patients feel comfortable and will perform optimally; in some cases, this means that additional explanations are necessary to ensure patients understand the need to undergo testing and the instructions for each individual test. ECCroN also recommends that researchers follow the adaptation and translation procedures outlined by the International Test Commission [167] when applying existing tests to a population the test was not designed for ECCroN is actively represented in the workgroup of the International Neuropsychological Society's Cultural Special Interest Group that is working on a neuropsychological comment on the ICT guidelines.

Several of such widely-applicable instruments have already been developed over the years across Europe, such as the aforementioned European Cross-Cultural Neuropsychological Test Battery (CNTB [140]), the Multicultural Cognitive Examination (MCE [202]), the Cross-Cultural Dementia Screening (CCD [57]), the computerized EMBRACED battery [345], the computerized Battery for Neuropsychological Evaluation of Children (BENCI [346]), an innovative verbal fluency-switching task (TFA-93 [203]) and a number of culturally—or regionally—appropriate picture-based memory tests, such as the Recall of Pictures Test (RPT [347]), modified Visual Association Test (mVAT [201]), TMA-93 [134], and TNI-93 [135]. A normative data and validation study was carried out for European majority groups, Pakistani/Indian, Polish, Turkish, and to a lesser extent Moroccan and Former Yugoslavian participants for CNTB (using multilingual research assistants or trained interpreters); for the CCD, the normative data and validation study was conducted among native Dutch, Moroccan-Arabic, Moroccan-Amazigh (Berber), Turkish, Surinamese-Creole, and Surinamese-Hindustani participants (assessed by bilingual, bicultural neuropsychologists); a general multicultural immigrant population as well as native French individuals were studied for the normative data and validity studies of the TNI-93, TMA-93, and TFA-93 (assessment in French). Normative data is increasingly collected for other tests; in addition, smaller normative data sets that were not formally published are available for some tests, such as a sample of predominantly Turkish individuals for the mVAT, which was validated in multicultural memory clinics across the Netherlands.

By using these more widely applicable tests that can be administered with an interpreter present, many cultural and linguistic effects can be minimalized (e.g. [21]). Ultimately, such tests may prove more feasible in diverse patient populations, and reducing cultural and linguistic effects will make the interpretation of the results of the neuropsychological assessment less challenging. The influence of education and literacy on test performance seem to be the most difficult to reduce in test design. In some cases, education-based norms will therefore remain necessary, although some recent paradigms using ecologically relevant material, some with low linguistic demand, show promise in that respect as well [348]. ECCroN consortium members are currently developing and validating tests measuring less well-studied cognitive domains in diverse individuals, such as language and social cognition, as well as brief tools that can be used to screen for cognitive impairment in general practice.

4 Becoming more sensitive to diversity: ECCroN recommends improvements in clinician training and the use of interpreters

Ideally, the European workforce of neuropsychologists would reflect the level of diversity within European societies; in current reality it is likely far from that ideal. For example, current selection criteria result in an underrepresentation of Black and Asian applicants in doctoral programs in (clinical) psychology [349], and Black and Asian individuals are underrepresented among National Health Services psychologists in comparison to the general population [350]. Although data on diversity among neuropsychologists is lacking in other European countries, both experts in cross-cultural neuropsychology [229] and the Cultural and Ethnic Diversity Taskforce of the European Federation of Psychologists' Associations [351] have previously recognized diversity among the professional workforce of (neuro)psychologists as an important issue. Currently, it is not clear which factors contribute to this underrepresentation of diverse (neuro)psychologists in Europe, and to what degree these issues vary across Europe; however, it is likely that these mechanisms will vary by country due to variation in factors such as entry criteria and selection procedures (e.g. no selection, selection based on grade point averages, selection based on assessment) and the accessibility of graduate and postgraduate education (e.g. tuition fees) across European countries. More research is urgently needed to shed light on the mechanisms behind the underrepresentation, before any targeted actions can be undertaken in the form of, for example, mentoring programs or changes to selection procedures.

However, even if diversity levels were to improve, it is unlikely that it will be possible to provide same-ethnicity providers to every patient; for example, major cities in the Netherlands like Rotterdam and Amsterdam represent more than 170 nationalities [352], while nationwide, there are only 161 neuropsychologists registered under the protected title of clinical neuropsychologist—who supervise a small subset of neuropsychologists among the 14,641 nationally registered health care psychologists [353]. While recognizing the potential benefits of assessments conducted by same-ethnicity neuropsychologists, such as outlined by e.g. Byrd et al. [354], ECCroN therefore advocates for a general improvement in the clinical training of all neuropsychologists in cross-cultural neuropsychological assessment, to ensure patient-friendly communication and correct administration and interpretation of cross-cultural neuropsychological tests. ECCroN is specifically investigating the development of a best practice that includes the minimal requirements for carrying out cross-cultural neuropsychological assessment, drawing inspiration from previous work by international (neuro)psychologists, such as the "Guidelines on multicultural education, training, research, practice, and organizational change for psychologists" by the American Psychological Association [188] and the work by Fujii [13]. In addition to a best practice, ECCroN is currently working towards cross-cultural clinical training at a European level, such as a European summer school or post-master course in cross-cultural neuropsychological assessment. To this end, ECCroN has started to collect and integrate existing training materials in Europe that were identified in a previous study [229]. We particularly endorse European-level training as a first step, as integrating cross-cultural neuropsychology training in all individual national neuropsychology curricula is challenging given the variation in the duration, level, and content of training in neuropsychology across European countries [223]. A European program ensures good accessibility, particularly for neuropsychologists working in countries in which cross-

cultural neuropsychology is less developed. This program may provide state-of-the-art knowledge through physical or virtual lectures held by ECCroN members, conveying the latest evidence-based practices from the international literature. Country specific add-ons to this European summer school or post-master course can subsequently be developed if needed. After this European program has been established, ECCroN aims to contribute to the integration of cross-cultural neuropsychology in national pre- and postgraduate training programs in neuropsychology.

Last, guidelines for interpreter-mediated neuropsychological assessment in diverse populations in Europe should be developed or adapted from existing guidelines for working with interpreters in psychological/medical practice, e.g. those of the British Psychological Society [225]. These guidelines should cover several aspects; for example, they may describe how to brief interpreters before the neuropsychological assessment about the aims of the assessment and its standardized test procedures [229]. It may also cover aspects such as the disadvantages of interpreter-mediated assessment via telephone, issues with regional variations in languages (e.g. Spanish in patients from South America) and issues with interpreters who are not certified [16,229].

5 The potential impact of ECCroN

ECCroN may impact research and clinical practice in several ways. First, it may accelerate improvements in assessment and subsequent diagnosis of diverse individuals in Europe. Such improvements are urgently needed; for example, previous European work has indicated that dementia is likely over-diagnosed in diverse individuals younger than 60 years and underdiagnosed in those older than 60 [75]. Populations that may particularly benefit from collaborative consortium efforts are those that are relatively small and scattered across Europe. For example, there is a large population of people from Former Yugoslavia in Germany, whereas this population is notably smaller in other European countries [338]; in such cases, multinational collaborations to validate tests or collect norms may be particularly helpful. Second, this consortium may facilitate the implementation of state-of-the-art knowledge and practices. Third, the ECCroN approach may serve as an example to other regions characterized by high levels of diversity; in fact, some of the instruments developed for diverse individuals in Europe are currently already implemented in other regions, such as the CNTB in Brazil [355]. Fourth, standardized training at the European level ensures that clinicians across Europe have access to high quality clinical training even where such training is unavailable or not part of the curriculum for neuropsychologists in the individual countries. Last, improvements in and standardization of the measurement of diversity-related variables provides an opportunity to examine theoretical assumptions regarding the influence of these variables on test performance in diverse individuals.

6 Conclusion

Here, we have raised several important challenges of cross-cultural neuropsychological assessment and assessed the practice landscape for diverse populations in Europe. Furthermore, we provide some solutions to existing barriers for culturally appropriate services. In sum, ECCroN aims to work towards a neuropsychological assessment that is carried out by neuropsychologists trained in cross-cultural assessment, with the help of

a well-instructed interpreter where required, and through using tests that are specifically suitable for patients with a wide variety of backgrounds, while taking into account the full spectrum of diversity-related variables in research and clinical practice. Such an approach allows European neuropsychologists to ultimately conduct neuropsychological assessments of diverse individuals that are in line with national professional and ethical codes of conduct (e.g. [356-358]). ECCroN will work to build on the momentum of existing partnerships within the collaboration to attract new members from across Europe, establishing measurable impact within the neuropsychology research and practice within Europe and beyond.

CHAPTER 4.3

Diversity in Alzheimer's Disease drug trials: the importance of eligibility criteria

Sanne Franzen
Jade E. Smith
Esther van den Berg
Monica M. Rivera Mindt
Rozemarijn L. van Bruchem-Visser
Erin Abner
Lon S. Schneider
Niels D. Prins
Ganesh M. Babulal
Janne M. Papma

Alzheimer's & Dementia, 2021. Advance online publication

Abstract

Introduction:
To generalize safety and efficacy findings, it is essential that diverse populations are well represented in Alzheimer's disease (AD) drug trials. In this review, we aimed to investigate participant diversity in disease-modifying AD trials over time, and the frequencies of participant eligibility criteria.

Methods:
A systematic review was performed using Medline, Embase, the Cochrane Library, and Clinicaltrials.gov, identifying 2247 records.

Results:
In the 101 included AD trials, participants were predominantly White (median percentage: 94.7%, interquartile range: 81.0–96.7%); and this percentage showed no significant increase or decrease over time (2001–2019). Eligibility criteria such as exclusion of persons with psychiatric illness (78.2%), cardiovascular disease (71.3%) and cerebrovascular disease (68.3%), obligated caregiver attendance (80.2%), and specific Mini-Mental State Examination scores (90.1%; no significant increase/decrease over time) may have led to a disproportionate exclusion of ethnoracially diverse individuals.

Conclusions:
Ethnoracially diverse participants continue to be underrepresented in AD clinical trials. Several recommendations are provided to broaden eligibility criteria.

1 Introduction

Although ethnoracially diverse individuals are at an increased risk of developing Alzheimer's disease (AD) dementia [44,73,359,360], these populations are systematically underrepresented in AD clinical trials [361-363]. To generalize safety and efficacy findings from drug trials to the general population, it is essential to include a diverse population, as differences in pharmacokinetics and pharmacodynamics across diverse populations may impact treatment effect and safety [364,365]; for instance, drug metabolism rates may differ [362]. The lack of diversity among clinical trial participants is often attributed to enrolling and retaining practices, such as recruitment strategies that do not account for factors that play a role in diverse populations, including mistrust and worry because of historical racism in medical research or the possibility of injury or complications [246].

Although recruitment factors should be taken into consideration, other explanations need to be considered as well, especially because a number of studies have indicated that people from underrepresented populations may be equally willing to participate in health research [366,367]. One important potential cause is that there are inherent features of AD-clinical trial eligibility criteria that lead to a disproportionate and systemic exclusion of underrepresented populations [368,369]. In 1997, Schneider et al. [369] demonstrated that applying the eligibility criteria of typical AD clinical trials to a Californian memory clinic population led to a systematic underrepresentation of people who are older, female, ethnoracially diverse, lower educated, and less wealthy; they provided several suggestions to improve provisional eligibility, such as a wider range of allowed scores on the Mini-Mental State Examination (MMSE [198] or by allowing more patients with (mild) behavioral and psychological symptoms to participate.

This systematic review aims to take a closer look at diversity in clinical trials and eligibility criteria. The first goal was to investigate the level of participant diversity in AD clinical trials in the decades after the publication of Schneider et al. [369] The second goal was to identify which eligibility criteria have been used and how these eligibility criteria were defined. Third, we aimed to assess whether the use of criteria related to cognitive and neuropsychiatric instruments such as the MMSE have changed over time, as these were highlighted by Schneider et al. [369] as particularly problematic. Last, we will discuss how some eligibility criteria may have affected diversity levels in AD clinical trials.

2 Methods

2.1 Search strategy

We performed a systematic review using Medline (which includes PubMed), Embase, the Cochrane Library, and ClinicalTrials.gov, without restrictions on the year of publication or location of the trial. Search terms included different terms for AD and mild cognitive impairment (MCI), terms referring to disease-modifying drugs, terms related to amyloid beta (Aβ) and tau, and different terms for phase II and phase III trials (for the complete lists of the search terms used, see Supplementary Text 1 in supporting information). Studies were included up to December 2019. Two independent authors screened all collected study data (JS and SF). Disagreement was resolved by a consensus agreement together with JMP. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)

guidelines [370] were followed, except for an assessment of the risk of bias—this step was omitted, as the aim was not to review or summarize the treatment effect reported in the included clinical trials.

2.2 Eligibility criteria

To be included in the review:

1. The study needed to be a planned, ongoing, completed, or early terminated phase II or phase III drug trial for patients with AD dementia, prodromal AD (early AD stage 3 [371]), or amnestic MCI (aMCI).
2. The experimental drug was a disease-modifying treatment. Disease-modifying was defined as targeting the pathogenic steps in the Aβ or tau pathways. This includes passive vaccination, monoclonal antibodies, agents disrupting accumulation or aggregation, and agents increasing clearance. As no agreed-upon standards are currently available that definitively delineate which drugs are considered disease-modifying, drug mechanisms were confirmed by consulting relevant literature (e.g. Galimberti & Scarpini [372]) and examining trial features (e.g. outcomes measuring amyloid clearance).

To adequately capture recent developments; collate study results; and provide a clearly delineated, concise set of recommendations, we focused on a homogeneous set of trials and excluded several other types of trials and study populations from this review. First, we excluded studies focusing on other forms of dementia. Second, we excluded AD prevention trials (e.g. lifestyle intervention trials) and studies in preclinical AD (early AD stages 1–2 [371]) as these types of trials present with unique challenges and eligibility criteria. Third, we excluded studies focused on symptomatic treatment of AD, including studies of acetylcholinesterase inhibitors—tacrine, donepezil, rivastigmine, and galantamine—and memantine. Fourth, we excluded trials investigating herbal and dietary treatments (e.g. vitamin supplements, olive oil, huperzine). Conference abstracts, dissertations, comments, editorials, book chapters, white papers, and reviews were also excluded.

2.3 Data extraction

For each included study, all available study protocol sources—that is, published papers or National Clinical Trial (NCT) database, European Union Drug Regulating Authorities Clinical Trial Database (EudraCT), and Australian New Zealand Clinical Trial Registry (ANZCTR) clinical trial registrations—identified in the search were used for data extraction. When available, the year that the study was first posted, the study phase, the investigational drug, the inclusion and exclusion criteria, the number of recruited participants, and participant demographics were recorded. Information was compiled from all available sources to create the most complete account of each study's design and study sample.

2.4 Data analysis

Participant eligibility criteria were divided into three main categories: 1) criteria related to medical conditions; 2) criteria related to undergoing specific study procedures, such as neuropsychological tests and brain scans; and 3) criteria based on diagnostic tests and questionnaire outcomes. Analyses were mostly descriptive. We used Cochran-Armitage trend tests (using the CATT package in R) to assess trends over time for binary variables,

that is, whether a criterion was used in the trial or not. Spearman correlations were used to analyze associations between the study start year and continuous variables.

3 Results

We identified 2247 records. The review process is summarized in the PRISMA flowchart in Figure 1. After deduplication, 1777 records remained; these records were screened on title and abstract. If the topic of the abstract fell within the criteria, but there was insufficient information on drug mechanism and/or trial phase, we reviewed the full text. A total of 506 records (clinical trial registrations or papers) were assessed in full for eligibility. A total of 17 NCT registrations, 35 EudraCT registrations, and one ANZCTR registration linked to published papers were retrieved manually. For three studies for which a published paper was available, we could not identify a clinical trial registration.

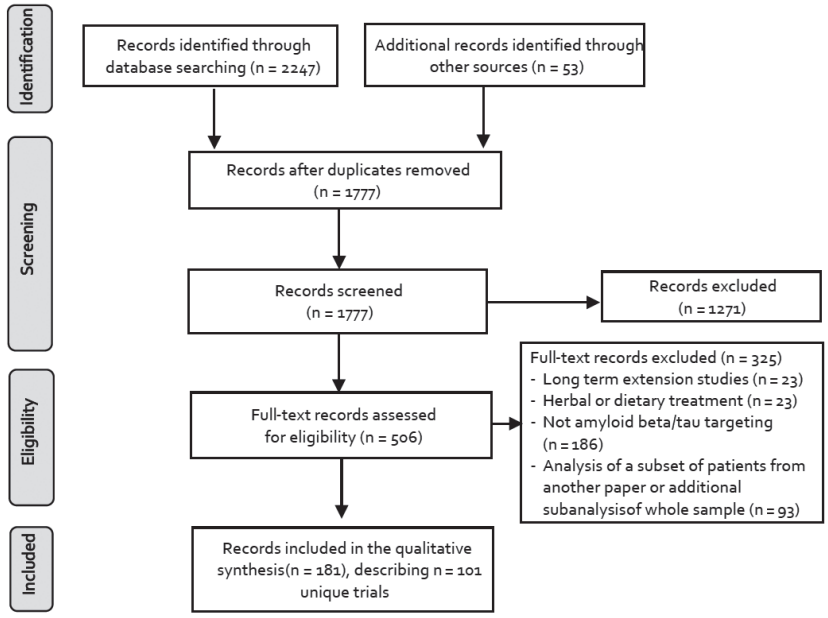


Figure 1. Results of database searches and selection process.

A total of 101 trials were included in this review. We extracted information about these trials from 181 unique papers and clinical trial registrations, as well as from 21 full protocols that were attached to the included papers or clinical trial registrations. The full protocols were not publicly available for the remaining trials. The sample consisted of 67 phase II trials and 34 phase III trials, investigating 47 different drugs. The studies covered 2001 to 2019, during which 79 studies had finished recruitment, and 22 studies had not yet commenced or were registered as active/recruiting. A listing of the included papers and clinical trial registration numbers is provided in Supplementary Table 1 in supporting information. Several of the eligibility criteria were more prevalent in studies for which a full protocol was available as opposed to studies for which a full protocol was not available (see Supplementary Text 2 in supporting information).

3.1 Diversity in clinical trial participants

Of the 101 trials, most had one or more study site(s) in North America (79.2%) or Europe (60.4%), and less frequently Asia (36.6%), or Oceania (32.7%); even fewer trials included study sites in South America (14.9%) or Africa (6.9%). Race/ethnicity data of the enrolled participants was available for less than half of the clinical trials (46 studies, 45.5%). Of these trials, 10 (9.9%) reported only the percentage of White participants without specifying percentages for any other ethnoracial groups, and four (4.0%) included White participants only. Race/ethnicity data was available for 58.2% (46/79) of the studies that were registered as completed or early terminated. When looking specifically at trials for which a published paper was available, 75.5% reported any race/ethnicity data (40/53). Different race/ethnicity categorizations were used across studies. Trials in Clinicaltrials.gov often reported race and/or ethnicity according to the National Institutes of Health/Office of Management and Budget (NIH/OMB) categories. Although few papers explicitly reported using the NIH/OMB categorization, a selection of these categories was often used in papers as well, whereas other categorizations were used very infrequently—one trial conducted across Asia, Europe, North America, and South America reported numbers for "Caucasian", "African", "Hispanic", "East Asian", and "West Asian" participants, and a paper about a trial conducted in the UK and Singapore reported the numbers of "Afro-Caribbean", "Asian", and "Caucasian" participants.

The median reported percentage of White participants in all studies was 94.7% (interquartile range [IQR]: 81.0–96.7%). This percentage of White participants was invariably high across both trials that did and those that did not use specific eligibility criteria (see Supplementary Table 2 in supporting information). Only seven studies reported the number of participants with a Latinx (Latina/o) ethnic background (median: 5.6%, IQR: 4.2–11.4%); specifically, 20.0% of the trials that included a North American site for which race/ethnicity data was available (7/35) reported the number of participants with a Latinx ethnic background. Data regarding (non-)Latinx background was often presented separate from the number of participants in each racial group; it was therefore unclear how many participants with a Latinx background were included across racial groups (e.g. Latinx–White). The median percentage of Black/African American participants was 1.2% (IQR: 0.4–1.7%), and the median percentage of Asian participants was 4.4% (IQR: 0.3–17.3%; NB: three studies from Asia had samples consisting of 100% Asian participants). The median percent of other or multiracial participants was 0.9% (IQR: 0.0–1.9%).

We found no statistically significant relationship between the percentage of White participants and the study start year ($\rho = -.26$, $p = .09$). Of the studies for which a published paper was available 47.2% (25/53) reported the number of people who did not meet the eligibility criteria. Only 17.0% (9/53) specified which criteria most frequently were the cause of participant exclusion. Although one study (NCT00105547) reported whether the excluded and included patients differed on age and sex, none of the studies reported whether included and excluded participants differed on race/ethnicity.

Of the studies reporting race/ethnicity, none explicitly referred to socioeconomic status (SES), while 41.3% (19/46) reported on the participants' education level. We extracted the mean education level of the total sample for each of these studies and calculated the average of the reported means across placebo and intervention groups for studies that did not report

the total sample mean. The average mean number of years of education across these studies was 13.3 years, and a higher mean level of education was significantly correlated with a higher percentage of White participants included in the trial ($\rho = .61$, $p = .02$).

3.2 Eligibility criteria

3.2.1 Criteria related to medical conditions

The frequency of exclusion criteria related to medical conditions is displayed in the first columns of Table 1, ranked from most prevalent (top) to least prevalent (bottom). In the remaining columns to the right, we present the prevalence of these medical conditions in several ethnoracial groups to provide context for the potential impact on ethnoracial diversity of participants. In addition to ethnoracial groups within the United States [373] (non-Latinx White, Latinx, non-Latinx Black, American Indian/Alaska Native), we have included prevalence estimates from the Indigenous Australian population [374] as an example to illustrate the potential impact of eligibility criteria on an international scale (see note to Table 1 for additional sources used to compile this table).

Non-AD neurological diseases and (major) psychiatric disorders were used as an exclusion criterion in more than three quarters of the included AD trials (Table 1, column 3), followed by cardiovascular disease (71.3%) and a history of cerebrovascular disease (68.3%). The last five columns of Table 1 demonstrate that the prevalence of some medical conditions is higher in either non-Latinx Black US residents, Latinx US residents, American Indian/Native Alaskan US residents, or Indigenous Australians than in non-Latinx White US residents or non-Indigenous Australians: diabetes, major psychiatric disease, cerebrovascular disease, renal disease, alcohol/substance use disorder, liver disease, higher weight/body mass index (BMI), and human immunodeficiency virus (HIV) diagnosis rates. For diabetes, studies sometimes referred to specific HbA1c levels, but these levels varied substantially from <6.0% to <9.0%; other studies included "insulin dependent" diabetes, "poorly controlled" diabetes, or merely "diabetes". Studies with a BMI criterion mostly required participants to have a minimum BMI of 18 or higher, but the upper cut-off value varied considerably from 28 to 40. Weight criteria specified a minimum weight of between 35 and 45 kg (≈77–99 pounds), mostly with a maximum of 120 kg (≈265 pounds). For hepatic disease, specific alanine transaminase (ALT; 1.5–3 times upper limit of normal, or ULN), aspartate transaminase (AST; 1.5–3 times ULN), and/or bilirubin (1.5–2.5 times ULN) cut-off levels were generally defined. For renal conditions, some studies referred to specific levels of creatinine clearance, whereas others only described "severe" renal disease, "impaired renal function", or specified dialysis requirement as the exclusion criterion.

3.2.2 Criteria related to study procedures

Caregiver attendance was the most prevalent criterion related to study procedures (80.2%, see Table 2), which often specified that the same caregiver had to attend all study visits and sometimes that the caregiver either had to live at the patient's home or had to visit a minimum number of times (range: <1–5 times/week) or hours per week (range: 4–24 hours/week). Some studies were more flexible, for example, by requiring the caregiver to accompany the patient only on key follow-up visits and allowing the patient to be accompanied by a "delegate" on the other visits. Written informed consent (52.5%) and a contraindication to undergoing positron emission tomography (PET)/magnetic resonance imaging (MRI; 51.5%) were used as a criterion in the majority of the included AD clinical trials.

Table 1. Frequencies of eligibility criteria related to medical conditions and prevalence of medical conditions in American and Australian ethnoracial groups*

	Criterion frequency in all trials (N = 101)		% in n-L white Americans	% in Latinx Americans	% in n-L Black Americans	% in American Indian and Alaska Native	% in Indigenous Australians†
Other neurological disease	81	80.2%	-	-	-	-	-
Psychiatric disorder	79	78.2%	6.9%	9.4%	9.7%	-	12% (9.6%)
Cardiovascular disease	72	71.3%	11.5%	8.2%	10.0%	14.6%	13% (1.2x)
Cerebrovascular disease	69	68.3%	2.6%	2.5%	3.9%	3.0%	-
Hachinski ischemia scale score >4	53	52.5%	-	-	-	-	-
Cerebrovascular evidence on MRI	48	47.5%	-	-	-	-	-
Childbearing/conception	62	61.4%	-	-	-	-	-
Unspecified systemic illness	62	61.4%	-	-	-	-	-
Alcohol or drug abuse	59	58.4%	8.4%	8.6%	7.4%	14.9%	18% (19%)
Vitals or lab abnormalities	53	52.5%	-	-	-	-	-
Infections/infectious diseases	50	49.5%	-	-	-	-	-
HIV status‡	26	25.7%	4.8‡	16.4‡	39.2‡	7.7‡	5.5‡ (4.5‡)
Liver disease	48	47.5%	1.7%	2.7%	1.1%	2.5%	15%–23% (1.4x–2.1x)
Autoimmune disease	47	46.5%	22.0%	16.8%	21.0%	30.6%	10.0% (1.1x)
Renal disease	46	45.5%	2.0%	2.2%	3.1%	-	3.0% (~3.7x)
Seizure disorder	44	43.6%	-	-	-	-	-
Cancer	41	40.6%	9.1%	4.2%	5.1%	7.1%	1.7% (1.5%)
Respiratory illness§	26	25.7%	7.5%§; 3.6%	6.0%§; 2.7%	9.1%§; 3.4%	9.5%; -	18% (1.9x) -
Endocrine dysfunction	25	24.8%	-	-	-	-	-
Brain/head trauma	25	24.8%	-	-	-	-	-
Diabetes¶	23	22.8%	8.6%¶; 13.0%	13.2%¶; 21.5%	13.1%¶; 19.6%	23.5% -	11% (3.3x) -
Weight or BMI cut-off	21	20.8%	31.0%	34.9%	38.0%	48.1%	37% (1.6x)
Gastrointestinal disease	18	17.8%	5.7%	4.3%	4.9%	8.3%	-
Excessive smoking (≥20 cigarettes per day)	9	8.9%	-	-	-	-	-
CNS inflammation	8	7.9%	-	-	-	-	-
Systemic inflammation	6	5.9%	-	-	-	-	-

Abbreviations: MRI = magnetic resonance imaging, HIV = human immunodeficiency virus, CNS = central nervous system, BMI = body mass index, n-L = non-Latinx

* 2018 US National Health Interview study data [373] and 2015 Australian Institute of Health and Welfare data are presented [374] (unless otherwise specified), providing prevalence rates for the following specific conditions within the broader categories specified in the first column: psychiatric disorders = moderate to severe depressive symptoms (USA [375]) versus feeling depressed (AUS); cardiovascular disease = any; cerebrovascular disease = stroke; alcohol or drug abuse = substance dependence or abuse (USA [376]) vs. lifetime risky alcohol consumption (AUS); infections – HIV status (USA [377]); autoimmune disease = arthritis diagnosis; renal disease = weak or failing kidneys (USA) vs. chronic kidney disease stages 3–5 (AUS); liver disease = any (USA) vs. abnormal ALT/GGT (AUS); cancer = any; weight or BMI = obesity; gastrointestinal disease = ulcers (duodenal, stomach, peptic).

† In parentheses: times increased risk as compared to non-Indigenous Australians or prevalence rate in non-Indigenous Australians

‡ Diagnosis rate per 100.000

§ Respiratory illness = current asthma (top) and chronic bronchitis (bottom)

¶ Diabetes = diagnosed (top) vs. diagnosed and undiagnosed combined (bottom [378])

Table 2. Frequencies of criteria related to undergoing study procedures

	Criterion frequency in all trials (N = 101)	
Caregiver attendance	81	80.2%
Written informed consent	53	52.5%
Contraindication to MRI/PET	52	51.5%
Adequate sensory abilities	42	41.6%
Language ability	35	34.7%
Residence in the community	35	34.7%
Caregiver consent	28	27.7%
Education requirement	19	18.8%
Reading or writing ability	19	18.8%
Determined likely to complete	15	14.9%
Recent hospitalization	4	4.0%

Of the 19 studies using an education criterion, eight studies also allowed a work history consistent with no intellectual disabilities. For language fluency, most studies required fluency in the test language (n = 11), in the "local" language (n = 11), or in English (n = 8), while four studies allowed fluency in one of a number of languages. One study allowed fluency in any language with sponsor approval, as long as 1) staff were also fluent in that language, and 2) required study documents were available in that language. A subset of studies (14.9%) included a criterion whether patients or patient–caregiver dyads were likely to complete the study in the opinion of the investigator; an operationalization of this criterion was not provided.

3.2.3 Criteria related to diagnostic tests and questionnaires

Cognitive tests, batteries, or screeners were used as an inclusion criterion in nearly all studies, with little variety in the tests that were used; the MMSE score was a criterion in over 90% of the studies (Table 3). Aside from the MMSE, a handful of other screening tests/short batteries were used, such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS [379]), the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog [380]), and the Montreal Cognitive Assessment (MoCA [204]). Additionally, some studies used memory-specific tests: the Free and Cued Selective Reminding Test (FCSRT [381]), tests from the Wechsler Memory Scale–Revised (WMS-R [382]), and the International Shopping List Test (ISLT [159]). One study used different cut-off scores for the test they used (WMS-R) to correct for education (0–7, 8–15, and ≥16 years); none of the other studies described different cut-offs based on demographic or sociocultural characteristics known to impact cognitive test performance (e.g. age, sex, ethnicity, quality of education, acculturation, etc.).

In addition to cognitive tests, roughly one-third of the trials used the Clinical Dementia Rating (CDR [70]) global score as a criterion. A similar proportion of studies used a measure of psychiatric symptoms as part of the eligibility criteria. For depression, the 15-item version of the Geriatric Depression Scale (GDS [245]) was used most often, as well as the Hamilton Depression Rating Scale [383]. The allowed range of scores for the GDS was relatively homogeneous across studies: the majority of studies (n = 22, 88% of studies with GDS) included patients with a score below 6 or 7, one study used the original 30-item version and used a cut-off score of ≤10, and two studies using a cut-off of <8 did not specify whether the long or short version of the GDS was used. The Columbia Suicide Severity Rating Scale

[384] was used a few times, but the majority of studies with a suicide risk criterion left the interpretation of this criterion to the opinion of the investigator (in contrast with depressive and cognitive symptoms).

Table 3. Frequencies of neurocognitive and neuropsychiatric screening tests and measures

	Criterion frequency in all trials (N = 101)	
Cognitive tests		
MMSE	91	90.1%
Memory-specific test*	7	6.9%
RBANS	4	4.0%
ADAS-Cog	3	3.0%
MoCA	1	1.0%
Global & functional measures		
CDR	36	35.6%
Eastern Cooperative Oncology Group status	1	1.0%
FAQ	1	1.0%
Psychiatric Assessments	25	24.8%
Geriatric Depression Scale	6	5.9%
Hamilton Depression Rating Scale	1	1.0%
Other depression instrument	5	5.0%
C-SSRS	14	13.9%
Other/unspecified suicide / self-harm risk scale		

Abbreviations: MMSE = Mini-Mental State Examination, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; ADAS-Cog = Alzheimer's Disease Assessment Scale-Cognitive Subscale; MoCA = Montreal Cognitive Assessment; CDR = Clinical Dementia Rating; FAQ = Functional Activities Questionnaire; C-SSRS = Columbia Suicide Severity Rating Scale
* Includes Free and Cued Selective Reminding Test (FCSRT), Wechsler Memory Scale-Revised (WMS-R), and International Shopping List Test (ISLT)

3.2.4 Diagnostic tests and screeners: the use of the MMSE, CDR, and GDS over time

Additional Cochran-Armitage trend analyses of the use of the MMSE revealed that the study start year did not differ between studies with or without an MMSE-eligibility criterion ($Z = 0.14$, $p = .89$); that is, the MMSE cut-off scores were not used significantly less (or more) often with time. As displayed in Figure 2, the cut-off score for the MMSE increased over time (MMSE lower limit $\rho = .53$, $p < .001$; MMSE upper limit $\rho = .48$, $p < .001$). Furthermore, the range of allowed MMSE scores narrowed over time ($\rho = -.44$, $p < .001$). Similar to the MMSE, the Cochran-Armitage trend test showed that there was no statistically significant increase or decrease in the use of the GDS by study year ($Z = 0.0$, $p = .99$); the CDR, however, was used significantly more frequently in later years ($Z = -2.48$, $p = .01$).

4 Discussion

In this systematic review, we aimed to 1) investigate the level of participant diversity in AD clinical trials targeting Aβ and tau; 2) identify which eligibility criteria have been used and how these criteria were defined; and 3) discover whether the use of criteria related to cognitive and neuropsychiatric instruments changed over time. The results showed that study samples were predominantly composed of White individuals, and ethnoracial diversity levels did not show a significant increase (or decrease) over time. Some of the most frequently reported criteria were the exclusion of participants with non-AD neurological disease, psychiatric illness, cardiovascular and cerebrovascular disease, obligated caregiver attendance, and cognitive impairment as defined by a specific score on the MMSE. The MMSE was used in an overwhelming majority of cases as the main cognitive eligibility criterion and was used consistently over time, with cut-off scores increasing over the years, but with the range of allowed scores decreasing over the years. The criteria related to

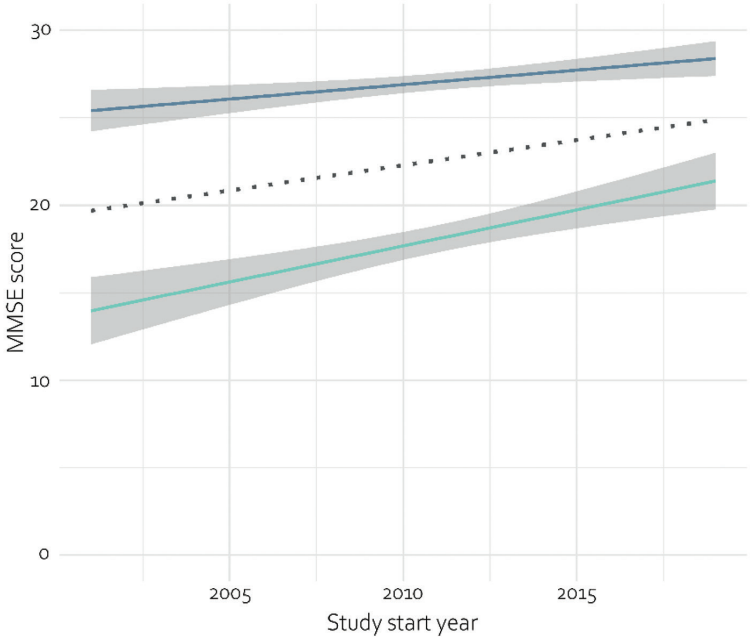


Figure 2. Changes in MMSE upper- and lower cut-off scores (midpoint in dotted line).

medical conditions and study procedures often were not well operationalized and cut-off scores were often wide ranging. In addition to these main aims, our goal was to discuss how these eligibility criteria may have affected diversity levels. In the following paragraphs, we will discuss the main outcomes of this review and provide recommendations for future clinical trials, an overview of which can be found in Table 4.

We could not retrieve race/ethnicity data for more than half of the studies included in this review; for those studies for which a paper was published, a little over three quarters reported race/ethnicity data. This is somewhat higher than in a review of cholinesterase inhibitors and memantine randomized controlled trials (59.2% [361]). The studies that reported race/ethnicity data included an overwhelming majority of White participants (≈95%), and no significant increase or decrease in this ratio was observed over time. For most trials, data regarding Latinx ethnicity was not reported, and in the handful of cases in which it was described, it was presented separately from the numbers by racial group. It was therefore not possible to determine how many Latinx versus non-Latinx participants were included, and whether these proportions may have changed over time. However, based on the studies that did report the number of Latinx participants, as well as the data from Black, Asian, and other racial groups, it seems unlikely that Latinx participants were well represented. This lack of diversity, as well as the underreporting of Latinx background are particularly notable for studies with a North American site (79.2%), given the rapidly increasing diversification of the United States during this review period. Whitfield et al. [385] describe how, as the ratio of White participants to other ethnoracial groups increases, the statistical power to detect group differences decreases drastically, and samples will typically have to include a larger proportion of diverse ethnoracial participants than a representative sample of the general population (e.g. more than 15% Black participants

Table 4. Issues with eligibility criteria of clinical trials and recommendations

Issue/criterion	Recommendations
Overarching issues	
- Race and ethnicity often were not reported	- Improve reporting
- Current race/ethnicity definitions not globally suitable	- Critically examine and improve definitions of race/ethnicity
- It is unclear how many diverse patients are invited, screened, and excluded	- Improve reporting
- It is unclear which criteria lead to exclusion	- Improve reporting
- Criteria from phase II copied to and expanded on in phase III	- Revisit/revise all criteria in moving from phase II to phase III
Criteria related to medical conditions	
- Imprecise/unspecific definitions of medical conditions	- Use validated, internationally recognized clinical classifications (of disease staging)
- Variation in cut-offs for specific medical conditions	- Organize expert consensus meetings to determine appropriate cut-offs in AD-research
- It is unclear if race corrections should be used or not	- Organize expert consensus meetings to determine whether and when to apply race corrections
- Exclusion of all patients with a medical condition regardless of past/present health status	- Include more patients who can safely participate, e.g. persons living with HIV who are medically stable and have a non-detectable viral load
- Questionable safety of drugs for patients with medical conditions due to exclusion	- Use expansion cohorts to study safety
Criteria related to study procedures	
- Language fluency as a barrier to participation	- Allow fluency in any language if adapted materials and staff speaking that language are available
- Lower educated individuals often excluded	- Allow persons with a work history consistent with no intellectual disabilities (ID) to participate - Investigate other ways to screen for ID
- Risk of compliance stereotyping if 'likely to complete' is not defined	- Define 'likely to complete' before trial
- Caregiver attendance as a barrier to participation	- Allow others to accompany patient on subset of visits - Plan appointments outside business hours - Explore remote interviewing options
- Written informed consent as a barrier in persons with limited literacy/education	- Explore alternatives for written informed consent, such as video informed consent
Criteria related to neurocognitive and neuropsychiatric measures	
- MMSE is unsuitable for diverse populations	- Consider alternative, more widely applicable tests - Use different MMSE cut-offs depending on education and other relevant variables
- CDR may be biased due to cultural differences	- Consider adaptations to the instrument/questions - Provide additional training to staff

in the sample). As it stands, the limited percentage of ethnoracially diverse individuals precludes sufficiently powered analyses of safety and efficacy across ethnoracial groups. In addition, currently used racial/ethnic categories themselves may need to be revised to fully represent global diversity—for example, categorizing all individuals from Europe, North Africa, and the Middle East as "White" does not do justice to the diversity within and between persons originating from these regions.

Our results showed that trials targeting Aβ or tau in AD often provide unclear definitions of their eligibility criteria; these imprecise definitions, such as "diabetes" or "impaired renal function" (not further specified), likely result in the exclusion of all or most patients with a specific medical condition. When specific ranges on indices of certain medical conditions were provided, such as BMI or ALT/AST levels, the allowed ranges differed substantially between studies. There thus seems to be a lack of consensus on how these conditions are best defined in the context of Aβ and tau trials. These ill-defined eligibility criteria may particularly affect the inclusion of underrepresented populations that are characterized by health disparities. Kim et al. [386] made several suggestions to broaden inclusion criteria in

oncology trials that may provide inspiration for AD trials. One of these recommendations is to include persons living with HIV (PWH) based on current and past CD4+ and T-cell counts instead of excluding all PWH—unless antiretroviral therapy is expected to interact with the investigational product. Additionally, one might take into consideration whether PWH are medically stable and whether they have a (non-)detectable viral load. Furthermore, Kim et al. [386] provided examples of how to improve the clarity of the definitions used in clinical trials eligibility criteria, such as the use of validated clinical classifications (of disease staging) as opposed to more generic definitions.

With regard to the impact of criteria related to medical conditions on the inclusion of ethnoracially diverse groups specifically, it is still uncertain if, how, and when race corrections should be used to evaluate various clinical laboratory results as indicators of specific medical conditions, such as indicators of kidney functioning [387] and several other common laboratory values [388]. Although such race corrections could potentially make the process of inclusion in clinical trials more inclusive, they may also inadvertently perpetuate or amplify existing disparities [389]. The field is in need of expert guidance to reach a consensus on whether and when to apply these race corrections.

Criteria related to undergoing study procedures were commonly part of the eligibility criteria. In the following paragraphs, the eligibility criteria related to language, education, caregiver attendance, written informed consent/reading and writing abilities, and whether patients are considered likely to complete the study, are discussed in more detail, specifically in the context of the inclusion of diverse individuals.

First, language requirements, such as fluency in the English language, were included in more than one third of the clinical trials. Depending on their definition, specific language requirements may lead to disproportionate exclusion of individuals from underrepresented populations. The lack of guidance on how to handle language barriers in clinical trials was acknowledged as a problem by multicenter research ethics committees in the UK [390]. A more inclusive solution may be to allow fluency in any preferred language, as long as the required test materials are available in that language and there is a staff member available who speaks the language to the degree necessary for cognitive testing—as was allowed in one trial (NCT00676143). This would, however, require the development/adaptation and validation of test materials across a number of languages. In addition, it may be worthwhile to investigate if assessment with experienced formal interpreters could be a viable option at study sites where the population is exceptionally diverse.

Regarding education, a minimum of six years of formal education was often used as a criterion—sometimes stating this was to ensure that patients with intellectual disabilities were not included. This criterion is problematic for several reasons; first, many diverse elderly patients across the world did not receive any formal education during childhood due to reasons other than intellectual disabilities—such as a lack of financial means or a large geographic distance to educational facilities (e.g. in first-generation immigrants in Europe). Second, mandatory primary education across the world has historically been variable—although some countries required six years of primary education, others may have required only four or five. Therefore, years or level of education cannot serve as a suitable proxy for intellectual disabilities in diverse patients. Some studies acknowledged

these barriers by allowing people with a work history consistent with no intellectual disabilities to participate in the study. Future studies should focus on developing ways to screen for intellectual disabilities that do not result in the exclusion of patients without intellectual disabilities who had limited access to formal education.

Several studies included a criterion that patients should be likely to complete the study. However, the interpretation of this criterion often was not defined, requiring the investigator to make this judgment call. Although such a criterion may be necessary to prevent costly missed visits in clinical trials, especially for studies using PET-ligands, likeliness to complete should be well defined at the outset. For example, a protocol may state that the patient and caregiver should complete a first run-in period of a specific number of screening visits fully compliant with the specified study procedures and in line with a specified time schedule. If this criterion is left undefined, it may prove problematic, as studies have indicated that participant selection may be influenced by implicit bias of the clinicians, that is, compliance stereotyping [391].

More than three quarters of the studies required some form of caregiver participation, often explicitly stating caregivers had to engage in frequent contact with patients—one study required caregivers to spend at least 24 hours per week with the patient. In some diverse ethnoracial groups, the main caregiver is often an adult child, rather than a spouse [55,392,393], and previous research has indicated that adult children are less likely than spouses to be eligible to participate alongside patients in dementia clinical trials [394]. Adult–child caregivers are more likely to still be active in the workforce [393], potentially limiting their opportunities to engage in frequent study visits due to the practical and financial burden of missed work. Researchers may provide more flexibility by allowing others to accompany patients on a subset of visits; by having appointments taking place outside of weekday business hours; or by exploring options for remote administration of interviews, such as over the phone or via video calls [393].

More than half of the AD clinical trials in this review explicitly required written informed consent. Although this currently seems to be the standard, requiring written informed consent will lead to the exclusion of people with low literacy skills—either because these patients will not be asked, or because they will be hesitant to sign a document they have difficulty understanding. Globally, ≈781 million adults are illiterate, with a high prevalence in lower- and middle-income countries [395], although disparities in literacy are also prevalent in some underrepresented populations in high-income countries. For example, so-called "guest workers" in Europe often received little if any formal education [51,179], and Latinx adults—and to a lesser degree Black and American Indian/Alaska Native adults—in the United States were overrepresented in the "below basic" level on the National Assessment of Adult Literacy [396]. To facilitate the enrollment of underrepresented populations, informed consent procedures will have to be tailored to patients and caregivers with low literacy skills. Over two decades ago, the US Food and Drug Administration (FDA) described the possibility of non-written consent procedures in illiterate English-speaking subjects, in which an impartial third party cosigns the consent document, preferably with a videotape recording [397]. A recent study in a different medical field (cardiology/endocrinology) has indicated that using a video informed consent procedure can increase the enrollment of patients from underrepresented populations [398]. As an additional example, in India,

audiovisual recording of the informed consent procedure has been mandatory since 2013, and standard operating procedures have consequently been developed [399]. AD research would benefit from efforts to incorporate alternatives to written informed consent developed in other research areas that include diverse and vulnerable populations, as well as from initiatives examining the feasibility of integrating such approaches in AD research.

Regarding cognitive screening tests and questionnaires, we found that the MMSE was used almost invariably as an inclusion criterion, and its use remained stable over time, with cut-off scores even increasing over the years. This is notable, given the fact that Schneider et al. [369] warned about the use of the MMSE in dementia trials in 1997. There is an abundant literature describing how MMSE-scores are substantially influenced by literacy and education [400-402] and likely also by cultural background [401]. In particular the subtests of orientation to time and place, serial 7s, figure copy, writing, and reading will be substantially influenced by someone's educational and cultural background [403]. Developing alternatives to written informed consent will only solve half of the problem as long as the cognitive tests used for screening and to measure primary and secondary outcomes require reading and writing skills. Moving forward toward more valid and inclusive global clinical trials will entail using other cognitive tests that are more suitable for diverse populations. For instance, the Rowland Universal Dementia Assessment Scale (RUDAS [60])—a test to assess the general level of cognitive impairment—or the International Shopping List Test [159]—for the inclusion of patients with memory impairment specifically—may be relevant options for further study. Before any instrument is selected for a clinical trial, it is imperative that a thorough review of the literature is carried out to determine whether the instrument is a valid and reliable measure of cognition in all groups that are to be included in the trial. As selection bias is often present in reliability/validity studies—for example, by excluding persons with low education levels or limited language fluency—it may be necessary to specifically check the demographic characteristics of these original study samples to ensure they reflect the intended trial sample. At a minimum, trials can be made more equitable by using different cut-off scores for groups with different levels of education in cognitive screeners and memory tests, as was done by one trial in this review (NCT00890890).

In addition to the MMSE, this study showed a rise in the use of the CDR as an inclusion criterion. The CDR has considerable merits, but researchers and clinicians need to be aware of possible cultural differences that may bias the results, such as 1) downplaying of cognitive symptoms out of respect for older family members, 2) different perceptions of what "normal" daily functioning may entail, 3) the need for adaptations to questions relating to hobbies that may be uncommon in some groups—for example, crossword puzzles—and social or cultural practices, 4) the potential influence of traditional gender roles, and 5) the potential influence of limited literacy on some activities of daily life [404]. Aside from the extensive training that is already needed to administer the CDR in a reliable and valid way in the general population, it is likely that additional training and/or adaptations to the instrument itself are needed to make it more suitable for the assessment of diverse populations across the globe.

In addition to these specific recommendations pertaining to criteria related to medical conditions, undergoing study procedures, and cognitive screeners and questionnaires,

some general recommendations may further improve inclusion of underserved populations in AD clinical trials. In the design phase, the FDA specifically recommends revisiting and revising the criteria when moving from a restrictive phase II to a more inclusive phase III trial [386,405]. Furthermore, they encourage the inclusion of samples known as "expansion cohorts" in trials—consisting of patients with specific comorbidities that may not fit the inclusion criteria for the main study—to determine the safety of doses in these populations as well [386]. Aside from changes to the trial design, more insight can be gained into the mechanisms behind the underrepresentation of diverse patients in clinical trials, if studies were to report the ethnoracial characteristics of all patients that 1) were considered for eligibility, 2) were invited, 3) were screened, and 4) were excluded/screen failed. In addition, reports should provide specifications regarding the eligibility criteria that were most often the reason for exclusion.

Although not technically part of the CONSORT (Consolidated Standards of Reporting Trials) guidelines [406], a short summary of the main reasons for exclusion may provide valuable insights to researchers on the eligibility criteria that have the strongest effect on eligibility. This information was only provided in a handful of studies in this review, and none of the studies specified whether there was a disproportionate exclusion of patients from underrepresented populations. It therefore remains unclear whether there was a disproportionate exclusion of patients from these groups based on overly strict eligibility criteria, or whether these patients were not invited in the first place or did not consent to study participation after invitation. For example, patients from underrepresented populations may experience geographical, financial, or logistical barriers that prevent them from participating in research [405,407]. Additionally, recruitment strategies need to be tailored to suit the needs of underrepresented populations, such as by investing in community-outreach programs, trust-building initiatives, and cultural-sensitivity training [246,390,408,409]. Financial support from funding agencies and/or the trial sponsor to facilitate such initiatives may be needed. In addition, more general financial or regulatory incentives from funding organizations or governmental bodies to actively enroll patients from underrepresented populations may further improve inclusion, for example, similar to the changes in the field of pediatrics, in which the Pediatric Research Equity Act (PREA) now requires manufacturers to complete studies in children if a substantial number of children is expected to use the drug [410].

Although this review specifically examined race/ethnicity, we acknowledge that race is a social construct and that health disparities are often driven by social determinants of health, such as education, literacy, socioeconomic status, racially patterned social stress, and access to care [411-413]. Although some trials in this review with race/ethnicity data reported the education level of the included participants, none mentioned SES. This limited reporting of social determinants of health is in line with a previous review in symptomatic treatment of AD, in which no studies reported on variables such as lifetime occupation, individual/household income, or wealth, and few studies on education [414]. It remains unclear how these variables may have affected enrollment of diverse participants in the trials included in this review; however, participants are often recruited in memory clinics, and these facilities may not be accessible to some underrepresented groups, for example because of limited health literacy [415], or because medical care is expensive and insufficiently covered by insurance [416].

Several limitations to this review should be mentioned. Although we did not exclude studies based on the language in which the record was written, our study did not identify any articles that were not written in English. Therefore, some local trials may have been missed. Second, race/ethnicity data was not available for a substantial number of studies, and the full protocols describing all eligibility criteria were only available for about one fifth of the included trials. As can be seen in the supporting information, the frequencies of the eligibility criteria may differ between studies with and without a full protocol available, and the rates we presented in this review may be an underestimation of the actual frequencies. For example, it seems unlikely that only slightly more than half of the clinical trials required written informed consent, particularly as the studies without such a criterion did not describe any alternative consent requirements. Likewise, trials that did not report race/ethnicity data may have included even fewer diverse participants—or, less likely, more—than the studies that did report race/ethnicity data. Third, in this review, we presented data from diverse ethnoracial populations in Australia and the United States alongside the frequencies of the eligibility criteria related to medical conditions to provide the reader with a better sense of the potential impact on diversity in clinical trials. These populations cannot be seen as directly representative of all underrepresented populations across the world, and given that these data were obtained in the general population, health disparities may actually be even more systemic and striking when zooming in on elderly populations specifically. For example, the prevalence of overweight and obesity in indigenous populations in Australia is 35% in those aged 15 to 17, but rises to 80% in those 55 and over [374]. Although we only showed data from the United States and Australia, similar health disparities are observed in populations outside those two countries, such as across different ethnoracial groups in Europe—particularly in the prevalence of diabetes, stroke, hypertension, and cardiovascular disease [47,417,418], but also in kidney disease [419,420]. Fourth, it is important to note that the data based on Latinx American samples is based on a pan-Latinx construction of this population. These studies did not account for the significant within-group variance that has important implications for health disparities and cognitive test performance (e.g. origin/nativity [Mexican, Puerto Rican, etc.], acculturation). Fifth, we only focused on Aβ and tau trials in this review. Although many of these recommendations can likely also be applied to other types of trials across neurodegenerative diseases, such as lifestyle trials like World-Wide FINGERS [421], some of these trials will come with their own unique challenges—such as a lack of suitable cross-cultural instruments measuring social cognition, language, and behavioral changes in frontotemporal dementia trials [229] as well as issues regarding the applicability of the diagnostic criteria for primary progressive aphasia subtypes across global languages, such as Chinese [422]. Last, we were unable to determine the direct effect of each criterion on the representation of diverse individuals using inferential statistics. Several factors precluded such analyses, such as the fact that some criteria were used either very infrequently or invariably (e.g. the MMSE, Supplementary Table 2), as well as the fact that race/ethnicity data was not reported for each global region/country specifically, precluding any comparisons of the makeup of the study samples with a priori disease estimates in the general populations in these countries/regions. The contribution of each individual eligibility criterion to the underrepresentation of diverse individuals across trials therefore remains unclear—even more so given the underreporting of the main reasons for exclusion.

Both federal law (Public Health Service Act §492B [410]) and NIH policy [423] require studies involving human subjects to address the inclusion of "minorities", and Alzheimer Europe [424] similarly calls upon researchers, ethics committees, and funders to address inequity in research. This review illustrates that there is a continuous, systemic underrepresentation of ethnoracially diverse groups in AD clinical trials. To generalize safety and efficacy data of AD clinical trials to the general population, more diverse individuals need to be enrolled, and modifying or changing the eligibility criteria in AD clinical trials may play a key role in reaching this goal.

Acknowledgments and funding
The authors would like to thank Wichor Bramer from the Erasmus MC University Medical Center Rotterdam for his help in developing the search strategy and Jolien Franzen for her contribution to the results section. We also acknowledge the ABOARD consortium for supporting the work on diversity in primary and secondary prevention of Alzheimer's disease in the Netherlands. SF and JMP report a grant from The Netherlands Organisation for Health Research and Development/Alzheimer Nederland (ZonMw Memorabel; grant number 733050834). JES reports a personal James B. Reynolds scholarship for foreign study. LSS reports a grant from NIH (P30 AG066530). GMB reports grants from NIH/NIA (R01AG068183, R01AG067428, A2021142S) and the BrightFocus Foundation (A2021142S). MRM is supported by grants from the NIA/NIH (R13 AG071313-01, R01AG065110-01A1, NIH/NIA 5U19AG024904-14, NIH/NIA R01AG066471-01A1), NIH/NIMH (U24MH100931-03), NIH/NIA (5R24AG065163), National Science Foundation, the Genentech Health Equity 2020 Fund (G-89294), and the Alzheimer's Association (AARGD-16-446038).

Conflict of interest
RLvB-V, JES, ELA, GMB, JMP, EvdB have nothing to disclose (aside from the funding reported in the "Funding Information" section). SF received support to attend meetings/conferences from Alzheimer Nederland and the Erasmus Trustfonds. She served as the executive committee member of the Cultural Diversity & Psychology commission of the Dutch Association of Psychologists and serves as executive committee member to the Diversity and Disparities PIA of ISTAART. NDP is consultant to Boehringer Ingelheim, Amylyx, and Aribio (payments made to his institution). He is co-PI of studies with EIP Pharma and Fuji Film Toyama Chemical. He serves on the DSMB of Abbvie's M15-566 trial (payment to his institution). He is CEO and co-owner of the Brain Research Center, the Netherlands. He is also on the scientific program committee of the Alzheimer's Association. LSS reports grants/contracts by Eisai, Eli Lilly, Roche/Genentech, Biogen, Biohaven, Novartis, and Washington University/NIA-DIAN-TU paid to the institution. In addition, he reports consulting fees from Abbott, AC Immune, Avraham Ltd, Boehringer Ingelheim, Cognition Therapeutics, Cortexyme, Eisai, FujiFilm, Immunobrain Checkpoint Ltd, Neurally Inc, Neurim Ltd, Neuronix Ltd, Samus, Takeda, vTv. MRM is the past president of the Hispanic Neuropsychological Society and standing member of the NIH NIA-T Study Section. MRM also has a relationship with the Alzheimer's Association Harem Community and Academic partnership. MRM received support from NIH/NIA to attend meetings. MRM received payment/honoraria for being panelist/chair/speaker/plenary speaker at the following events: AAIC Neuroscience Next 2020; mid-year conference of the International Neuropsychological Society 2019; Latinos and Alzheimer's Disease Symposium 2019; International Association of Forensic Mental Health Services Conference

2017; International Neuropsychological Society Annual Conference 2012; International Neuropsychological Society Annual Conference, 2012; 37th Annual Conference of the International Neuropsychological Society Annual Conference, 2010 American Academy of Clinical Neuropsychology annual meeting [2021 delayed due to COVID]; Harvard MGH Psychology Assessment Center Seminar 2021; University of Washington Department of Neurology Grand Rounds [delayed due to COVID 2020]; Annual Conference of the Pacific Northwest Neuropsychological Society, 2020; Annual Conference of the Council of University Directors of Clinical Psychology, 2020; National Academy of Sciences/Simons Foundation: The Science & Entertainment Exchange, 2019; Emory University HIV & Aging Conference; Brown University Alpert Medical School, Department of Psychiatry and Human Behavior Grand Rounds, 2019; Wisconsin Alzheimer's Institute/University of Wisconsin School of Medicine & Public Health 16th Annual Alzheimer's Disease Update Conference, 2018; 38th annual meeting of the National Academy of Neuropsychology, 2018; Council of Science Editors, Technica Editorial Services Webinar. The Peer Review Ecosystem: Where Does Diversity & Inclusion Fit In? 2018; 12) Colloquium Presentation, Dept. of Psychology, Ohio University, Athens, OH, 2018.

Chapter 4.3 Supplementary material

Supplementary text 1: Review search strategy

embase.com

('Alzheimer disease'/de OR 'dementia'/de OR 'mild cognitive impairment'/de OR (Alzheimer* OR dementia* OR (mild* NEAR/3 cogniti* NEAR/3 (impair*))):ab,ti) AND ('drug therapy'/de OR 'Alzheimer disease'/de/dm_dt OR psychopharmacotherapy/de OR 'psychotropic agent'/exp OR 'immunotherapy'/exp OR 'amyloid beta protein'/de OR 'enzyme inhibitor'/exp OR (drug* OR agent* OR psychopharmacotherap* OR pharmac OR inhibitor* OR (monoclonal* NEAR/3 antibod*) OR immunotherap* OR immun*-therap* OR amyloid-β OR β-amyloid OR beta-amyloid OR aβ OR a-β OR amyloidβ OR amyloid-beta):ab,ti) AND ('phase 2 clinical trial'/exp OR 'phase 3 clinical trial'/exp OR ((study OR trial*) NEAR/10 (phase-2 OR phase-2a OR phase-2b OR phase-2-a OR phase-2-b OR phase-3 OR phase-ii OR phase-iia OR phase-iib OR phase-ii-a OR phase-ii-b OR phase-iii)):ab,ti) NOT ([Conference Abstract]/lim) NOT ([animals]/lim NOT [humans]/lim) NOT ([Conference Abstract]/lim) AND [english]/lim NOT ('systematic review'/de OR 'meta analysis'/de OR ((systematic NEAR/3 review*) OR meta-analys* OR metaanalys*):ti)

Medline Ovid

(Alzheimer Disease/ OR Dementia/ OR (Alzheimer* OR dementia* OR (mild* ADJ3 cogniti* ADJ3 (impair*))).ab,ti.) AND (drug therapy/ OR Alzheimer Disease/dt OR exp Psychotropic Drugs/ OR exp Immunotherapy/ OR exp Amyloid beta-Peptides/ OR exp Enzyme Inhibitors/ OR (drug* OR agent* OR psychopharmacotherap* OR pharmac OR inhibitor* OR (monoclonal* ADJ3 antibod*) OR immunotherap* OR immun*-therap* OR beta-amyloid OR amyloid-beta).ab,ti.) AND (Clinical Trial, Phase II/ OR Clinical Trial, Phase III/ OR ((study OR trial*) ADJ10 (phase-2 OR phase-2a OR phase-2b OR phase-2-a OR phase-2-b OR phase-3 OR phase-ii OR phase-iia OR phase-iib OR phase-ii-a OR phase-ii-b OR phase-iii)).ab,ti.) NOT (news OR congres* OR abstract* OR book* OR chapter* OR dissertation abstract*).pt. NOT (exp animals/ NOT humans/) NOT (news OR congres* OR abstract* OR book* OR chapter* OR dissertation abstract*).pt. AND english.la. *NOT (Systematic Review/ OR Meta-Analysis/ OR ((systematic ADJ3 review*) OR meta-analys* OR metaanalys*).ti.)*

Cochrane CENTRAL

((Alzheimer* OR dementia* OR (mild* NEAR/3 cogniti* NEAR/3 (impair*))):ab,ti) AND ((drug* OR agent* OR psychopharmacotherap* OR pharmac OR inhibitor* OR (monoclonal* NEAR/3 antibod*) OR immunotherap* OR immun* next therap* OR amyloid next β OR β next amyloid OR beta next amyloid OR aβ OR a next β OR amyloidβ OR amyloid next beta OR *mab):ab,ti) AND ((phase next 2 OR phase next 2a OR phase next 2b OR phase next 2 next a OR phase next 2 next b OR phase next 3 OR phase next ii OR phase next iia OR phase next iib OR phase next ii next a OR phase next ii next b OR phase next iii):ab,ti)

Clinicaltrials.gov

Condition or disease: Alzheimer OR dementia

Other terms: antibodies OR inflammatory OR "rage antagonists" OR "calcium channel" OR amyloid OR tau OR psychotropics OR statins OR "hmg coa"

Additional Criteria: Phase: Phase 2 OR Phase 3

Supplementary text 2: Analyses of studies with and without a full protocol available
We compared studies with and without a full protocol using Fisher's Exact test (without correcting for multiple testing). Studies for which a full protocol was available more often contained criteria related to childbearing/conception ($p < .001$), a cardiovascular disease history ($p < .05$), Hachinski scores ($p < .05$), brain trauma ($p = .01$), respiratory illness ($p < .05$), infection ($p < .001$), HIV-status ($p < .05$), autoimmune disease ($p = .001$), cancer ($p < .001$), vital and lab abnormalities ($p < .05$), a contraindication to undergoing an MRI scan ($p < .01$), sensory abilities ($p < .01$), written informed consent ($p < .05$), language ($p < .05$), caregiver consent ($p < .05$), and whether patients are likely to complete the study ($p < .001$).

Supplementary Table 1. List of included trials and data sources

	Trial Registry			Trial Details					Additional Sources	
DMT	**NCT#**	**EudraCT**	**ANZCTR**	**Phase**	**Year**	**Status**	**Sample size**	**Region (number of countries)***	**Protocol**	**Paper**
AADvac1	NCT02579252	2015-000630-30		2	2015	Complete	208	Europe (8)		-
ABvac40	NCT03461276			2	2018	Recruiting	-	Europe (4)		-
Aducanumab	NCT02477800	2015-000966-72		3	2015	Terminated	1647	Asia (3), Europe (8), North America (2), Oceania (1)		-
	NCT02484547	2015-000967-15		3	2015	Terminated	1638	Asia (1), Europe (10), North America (2)		-
	NCT03639987	2018-002102-31		2	2018	Terminated	500	Europe (2), North America (2), Oceania (1)		-
AFFITOPE AD02	NCT01117818	2009-016504-22		2	2009	Complete	335	Europe (6)		[425]
Amilomotide (CAD106)	NCT00733863			2	2008	Complete	58	Europe (4)		[426]
	NCT00795418			2	2008	Complete	31	North America (1)		[426]
	NCT01097096	2009-012394-35		2	2009	Complete	121	Europe (8), North America (2)	Yes	[427]
AN1792 (AIP 001)	NCT00021723			2	2001	Complete	372	Europe (?), North America (1)		[428]
Atabecestat	NCT02260674	2014-002159-24		2	2014	Complete	114	Europe (6)		-
Avagacestat	NCT00810147	2008-005929-11		2	2008	Complete	209	Europe (3), North America (1)		[429]
	NCT00890890	2009-010067-16		2	2009	Terminated	263	Europe (4), North America (2)	Yes	[430]
Azeliragon	NCT00141661			2	2005	Complete	67	North America (1)		-
	NCT00566397			2	2007	Complete	402	North America (1)		[431]
	NCT02080364			2	2014	Terminated	880	Africa (1), Europe (2), North America (2), Oceania (2)		-
Lecanemab	NCT01767311	2012-002843-11		2	2013	Active	800	Asia (2), Europe (7), North America (2)		-
	NCT03887455	2018-004739-58		3	2019	Recruiting	-	Asia (2), Europe (6), North America (2)		-
Bapineuzumab		2004-004120-12		2	2004	Complete	26	Europe (2)		[432]
	NCT00112073			2	2005	Complete	234	North America (1)		[433]
	NCT00174525			2	2005	Unknown	-	North America (1)		-
	NCT00574132			3	2007	Complete	1331	Europe (2), North America (2)	Yes	[434]
	NCT00575055			3	2007	Complete	1121	North America (1)	Yes	[434]

Supplementary Table 1. Continued.

	Trial Registry			Trial Details					Additional Sources	
DMT	**NCT#**	**EudraCT**	**ANZCTR**	**Phase**	**Year**	**Status**	**Sample size**	**Region (number of countries)***	**Protocol**	**Paper**
	NCT00676143	2007-005995-14		3	2008	Terminated	1100	Africa (1), Asia (1), Europe (16), North America (2), Oceania (2) South America (2)	Yes	[435]
	NCT00663026			2	2008	Complete	79	North America (1)		-
	NCT00667810	2007-005994-79		3	2008	Terminated	901	Africa (1), Asia (2), Europe (16), North America (3), Oceania (2), South America (2)	Yes	[435]
	NCT01254773			2	2010	Complete	146	North America (1)		[436]
Bryostatin-1	NCT02221947			2	2014	Terminated	9	North America (1)		[437]
				2	2019	Complete	147	North America (1)		[438]
CHF 5074	NCT01602393	2010-024270-19		2	2011	Complete	51	Europe (1) North America (1)		-
	NCT01723670			2	2012	Withdrawn	0	North America (1)		-
Colostrinin				2	2004	Complete	105	Europe (1)		[439]
Crenezumab	NCT01397578			2	2011	Complete	91	Europe (2), North America (1)		[440]
	NCT01343966	2010-021926-37		2	2011	Complete	448	Europe (4), North America (2)		[441]
	NCT02670083			3	2016	Complete	813	Asia (3), Europe (21), North America (4), Oceania (1)	Yes	-
	NCT03114657	2016-003288-20		3	2017	Complete	806	Africa (1), Asia (5), Europe (16), North America (3), Oceania (1), South America (3)		-
Daratumumab	NCT04070378			2	2019	Recruiting	15	North America (1)	Yes	-
Donanemab	NCT03367403			2	2017	Active	266	North America (2)		-
Elenbecestat	NCT02322021			2	2014	Active	71	North America (1)		-
	NCT02956486	2016-003928-23		3	2016	Active	950	Asia (2), Europe (12), North America (2), Oceania (1), South America (1)		-
	NCT03036280	2016-004128-42		3	2017	Terminated	950	Africa (1), Asia (5), Europe (13), North America (3), South America (1)		-
ELND005	NCT00568776			2	2007	Complete	353	North America (2)		[442]
Etanercept	NCT01068353	2009-013400-31		2	2010	Complete	41	Europe (1)		-

Supplementary Table 1. Continued.

	Trial Registry			Trial Details					Additional Sources	
DMT	**NCT#**	**EudraCT**	**ANZCTR**	**Phase**	**Year**	**Status**	**Sample size**	**Region (number of countries)***	**Protocol**	**Paper**
EVP-0962	NCT01661673			2	2012	Complete	52	North America (1)		-
Gantenerumab	NCT01224106	2010-019895-66		3	2010	Active	799	Asia (1), Europe (16), North America (3), Oceania (1), South America (3)		[443]
	NCT02051608	2013-003390-95		3	2013	Active	389	Asia (2), Europe (16), North America (2), Oceania (1), South America (1)		-
	NCT03443973	2017-001365-24		3	2017	Recruiting	750	Asia (3), Europe (11), North America (3), South America (2)		-
	NCT03444870	2017-001364-38		3	2017	Recruiting	750	Asia (3), Europe (7), North America (2), Oceania (1), South America (3)		-
Gosuranemab	NCT03352557	2017-002901-37		2	2017	Active	654	Asia (1), Europe (6), North America (1), Oceania (1)		-
IVIg	NCT00299988			2	2006	Terminated	24	North America (1)		-
	NCT00812565			2	2008	Complete	58	North America (1)		-
	NCT00818662			3	2009	Complete	390	North America (2)		[444]
	NCT01300728			2	2011	Active	52	North America (1)		-
	NCT01524887	2011-000914-21		3	2012	Terminated	508	Asia (1), Europe (4), North America (2), Oceania (1)		-
	NCT03319810			2	2017	Complete	5	North America (1)	Yes	-
Lanabecestat	NCT02245737	2014-002601-38		3	2014	Terminated	2218	Asia (2), Europe (9), North America (3), Oceania (1)	Yes	-
	NCT02783573	2015-005625-39		3	2015	Terminated	3800	Asia (4), Europe (10), North America (3)	Yes	-
LMTM	NCT00515333			2	2007	Complete	323	Asia (1), Europe (1)	Yes	[445]
	NCT01689246	2012-002866-11		3	2012	Complete	891	Asia (4), Europe (9), North America (2), Oceania (1)	Yes	[446]
	NCT01689233	2012-002847-28		3	2012	Complete	761	Europe (9), North America (2), Oceania (1)	Yes	[447]
LY2886721	NCT01561430	2011-005217-37		2	2012	Terminated	70	Asia (1), Europe (3), North America (1)		-
LY3202626	NCT02791191			2	2016	Terminated	316	Asia (1), North America (2), Oceania (1)		-

Supplementary Table 1. Continued.

	Trial Registry			Trial Details					Additional Sources	
DMT	**NCT#**	**EudraCT**	**ANZCTR**	**Phase**	**Year**	**Status**	**Sample size**	**Region (number of countries)***	**Protocol**	**Paper**
NPO31112	NCT00948259			2	2009	Complete	30	Europe (1)		-
PBT2	NCT00471211			2	2007	Complete	78	Europe (1), Oceania (1)		[448]
			ACTRN12611001008910	2	2011	Complete	42	Oceania (1)		[449]
Ponezumab	NCT00722046			2	2008	Complete	194	Asia (1), Europe (2), North America (2), Oceania (1)	Yes	[450]
	NCT00945672	2009-011172-30		2	2009	Complete	36	Europe (1)		[451]
Rosiglitazone	NCT00428090			3	2007	Complete	862	Asia (5), Europe (9), North America (3), Oceania (1), South America (2)		[452]
Semagacestat	NCT00244322			2	2005	Complete	51	North America (1)		[453]
	NCT00594568			3	2008	Complete	1537	Africa (1), Asia (3), Europe (10), North America (2), Oceania (1), South America (2)		[454]
	NCT00762411			3	2008	Complete	1108	Asia (4), Europe (10), North America (3), South America (1)		-
Semorinemab	NCT03289143	2017-001800-31		2	2017	Active	457	Europe (10), North America (2), Oceania (1)		-
	NCT03828747			2	2019	Recruiting	260	Europe (3), North America (1)		-
Sodium selenate			ACTRN12611001200976	2	2009	Complete	40	Oceania (1)		[455]
Solanezumab	NCT00329082			2	2006	Complete	52	North America (1)		[456]
	NCT00905372			3	2009	Complete	1012	Asia (1), North America (2), South America (2)	Yes	[457]
	NCT00904683			3	2009	Complete	1040	Asia (3), Europe (8), North America (1), Oceania (1)	Yes	[457]
	NCT01148498			2	2010	Complete	55	North America (1)		-
	NCT01900665			3	2013	Terminated	2129	Asia (1), Europe (7), North America (2), Oceania (1)	Yes	[458]
	NCT02760602	2016-000108-27		3	2016	Terminated	26	Asia (2), Europe (8), North America (3)	Yes	-
Solanezumab / gantenerumab	NCT01760005	2013-000307-17		3	2013	Recruiting	490	Asia (2), Europe (7), North America (4), Oceania (1), South America (3)		-

Supplementary Table 1. Continued.

	Trial Registry			Trial Details					Additional Sources	
DMT	**NCT#**	**EudraCT**	**ANZCTR**	**Phase**	**Year**	**Status**	**Sample size**	**Region (number of countries)***	**Protocol**	**Paper**
Tarenflurbil (r-flurbiprofen)	NCT00105547			3	2005	Complete	1649	North America (1)		[459]
	NCT00322036			3	2006	Terminated	800	Europe (10), North America (2)		-
				2	2003	Complete	189	Europe (1), North America (1)		[460]
Thalidomide	NCT01094340			2	2010	Unknown	25	North America (1)		[461]
Thiethylperazine	NCT03417986			2	2018	Active	100	Europe (1)		-
Tideglusib	NCT01350362			2	2011	Complete	306	Europe (6)		[462]
Tilavonemab	NCT02880956	2016-001634-10		2	2016	Active	454	Asia (1), Europe (7), North America (2), Oceania (2)		-
Tramiprosate (homotaurine)				2	2002	Complete	58	North America (1)		[463]
	NCT00088673			3	2004	Complete	1052	North America (2)		[464, 465]
UB-311	NCT02551809			2	2015	Complete	43	Asia (1)	Yes	-
Vanutide cridificar (ACC-001)	NCT00479557	2006-002061-39		2	2007	Complete	86	Europe (3)	Yes	[466]
	NCT00498602			2	2007	Complete	160	North America (1)		[466]
	NCT00752232			2	2008	Complete	40	Asia (1)		[467]
	NCT00959192			2	2009	Complete	32	Asia (1)		[467]
	NCT01227564			2	2010	Complete	63	North America (1)		[468]
	NCT01284387			2	2011	Complete	126	North America (1)		[469]
Varoglutamstat (PQ912)	NCT02389413	2014-001967-11		2	2015	Complete	120	Europe (7)		[470]
Verubecestat	NCT01739348	2011-003151-20		3	2011	Terminated	1958	Asia (3), Europe (12), North America (2), Oceania (2), South America (2)		[471]
	NCT01953601	2012-005542-38		3	2012	Terminated	1454	Africa (1), Asia (2), Europe (13), North America (2), Oceania (2), South America (2)		[472]
Zagotenemab	NCT03518073			2	2018	Active	285	Asia (1), North America (2)		-

* Central American countries were included in North America. Russia was categorized as European. Puerto Rico is a US territory, however we counted it as a separate North American country for the purpose of this table

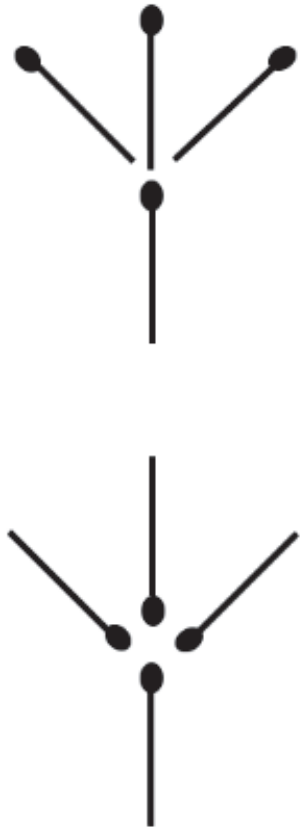
Supplementary Table 2. Median percent white for trials with and without specific eligibility criteria

Criterion	Used as criterion, median %white (n trials)	Not used as criterion, median %white (n trials)
Medical conditions		
Other neurological disease	95.80% (35)	92.50% (11)
Psychiatric disorder	95.45% (38)	88.10% (8)
Cardiovascular disease	94.50% (31)	96.10% (15)
Cerebrovascular disease	96.15% (30)	87.15% (16)
- Hachinski ischemia scale score >4	94.30% (30)	96.25% (16)
- Cerebrovascular evidence on MRI	95.45% (24)	94.20% (22)
Childbearing/conception	91.40% (27)	96.30% (19)
Unspecified systemic illness	95.95% (32)	91.95% (14)
Alcohol or drug abuse	91.95% (22)	95.45% (24)
Vitals or lab abnormalities	94.80% (23)	94.50% (23)
Infections/infectious diseases	90.60% (23)	96.30% (23)
HIV status	83.70% (11)	95.80% (35)
Liver disease	92.50% (21)	95.10% (25)
Autoimmune disease	95.45% (26)	94.20% (20)
Renal disease	92.50% (17)	95.10% (29)
Seizure disorder	95.80% (23)	94.10% (23)
Cancer	95.10% (21)	94.30% (25)
Respiratory illness	95.10% (11)	94.50% (35)
Endocrine dysfunction	92.80% (12)	95.45% (34)
Brain/head trauma	90.60% (15)	96.20% (31)
Diabetes	94.80% (9)	94.50% (37)
Weight or BMI cut-off	95.95% (6)	94.20% (40)
Gastrointestinal disease	94.50% (9)	94.80% (37)
CNS inflammation	96.25% (8)	94.40% (38)
Undergoing study procedures		
Caregiver attendance	94.10% (39)	96.30% (7)
Written informed consent	94.50% (31)	96.10% (15)
Contraindication to MRI/PET	94.95% (22)	94.40% (24)
Adequate sensory abilities	94.95% (20)	94.20% (26)
Language ability	96.15% (20)	91.95% (26)
Residence in the community	96.20% (19)	90.60% (27)
Caregiver consent	95.10% (21)	90.60% (25)
Education requirement	96.25% (8)	94.20% (38)
Reading or writing ability	95.35% (12)	94.55% (34)
Determined likely to complete	95.10% (9)	94.50% (37)
Cognitive and neuropsychiatric measures		
CDR	95.45% (8)	91.25% (38)
Geriatric Depression Scale	87.15% (12)	94.95% (34)
Hamilton Depression Rating Scale	95.10% (6)	94.65% (40)

* The following variables were excluded because ≤5 studies reporting ethnicity data used this criterion: 'systemic inflammation', 'excessive smoking', 'Recent hospitalization', 'other depression instrument', 'CSSRS', 'other suicide/self-harm risk scale', 'MoCA', 'Eastern Cooperative Oncology Group Status', 'FAQ', 'ADAS-Cog', 'RBANS', 'Memory-specific test'. Similarly, the MMSE was removed as ≤5 studies reported not using this criterion.

CHAPTER 5

GENERAL DISCUSSION



5 General discussion

Cross-cultural neuropsychological assessment is challenging due to barriers in language, education, culture, and several other factors. Given the increasing diversity caused by global migration, neuropsychologists need to become aware of and prepared for the intricacies of cross-cultural neuropsychological assessment. My dissertation provided a roadmap towards more sensitive neuropsychological assessment in diverse patient populations, with an emphasis on European—and specifically Dutch—memory clinics. The first part of the dissertation summarized the major gaps in cross-cultural neuropsychology, providing the points of departure for subsequent improvements as described in the second part. In the final part of this dissertation, I focused on the implementation of our findings in the clinical and research practices of tomorrow. In this general discussion, I will critically review the main findings per section and subsequently describe my perspective on methodological issues and on avenues for future research.

5.1 Main findings

Chapter 2: State of the art of cross-cultural neuropsychological assessment

5.1.1 There is a lack of appropriate tests and norms for culturally, educationally, and linguistically diverse populations, particularly for cognitive domains other than memory

Although a neuropsychological assessment including the CCD, RUDAS, and adapted versions of existing tests, such as a Turkish auditory verbal learning test, was sufficient to rule out cognitive impairment in the case study presented in chapter 2.1, these tests alone in most cases are insufficient to determine a profile of impaired and intact cognitive functions. Both chapter 2.2 and 2.3—the review describing the available neuropsychological tests and the Delphi study—highlighted a lack of appropriate neuropsychological tests. In addition, chapter 2.2 revealed that few studies reported using a rigorous cultural and linguistic adaptation procedure. This is in line with the findings from a review examining cultural adaptations of health scales for Turkish, Arabic, and Surinamese groups [473].

Memory tests with cross-cultural potential were relatively widely available (chapters 2.2 and 2.3), and experts perceived this domain as relatively easy to assess with the available tests (chapter 2.3). The overall diagnostic properties of memory tests were satisfactory (AUC .74–.99; chapter 2.2). Given the emphasis on AD in the literature included in chapter 2.2, it is perhaps not surprising that diagnostic accuracy was highest for memory tests—even more so given the issues with lack of blinding/circularity of reasoning in a number of these studies. Chapter 2.3 presents several tests specifically suitable for the diverse population of Europe: the TMA-93, TNI-93, Enhanced Cued Recall, the Objects test of the CCD, the WHO/UCLA adapted version of the Rey Auditory Verbal Learning Test, and the modified Visual Association Test (presented in chapter 3.2). This relatively large number of suitable memory tests might be interpreted as a sign that culture does not impact performance on memory tests to the same degree as tests of other cognitive domains. Interestingly, although total performance on memory tests may indeed be similar across diverse individuals, cross-cultural differences may still exist in the strategies used to store and remember information; for example, although elderly American individuals relied more heavily on categories in a (free recall) memory test than their Chinese peers,

their total scores were comparable [474]. It would be interesting to investigate whether such differences in the employment of specific memory strategies also occur in diverse populations in Europe.

Cognitive domains other than memory were studied less often, and diagnostic accuracy for these cognitive domains was also more variable (chapter 2.2). We identified a remarkable lack of (variety in) language tests studied in low educated, culturally diverse dementia populations. Chapter 2.3 similarly stressed the need for more research/development into language, which ultimately led to the development of the NAME presented in chapter 3.3. This lack of suitable language tests perhaps is not surprising, given that language tests in particular cannot simply be translated from one language to another. Although the NAME can be used to assess patients of a wide variety of linguistic backgrounds, language-specific tests are likely needed to measure the unique morphological, phonological, and syntactical characteristics of specific languages. For example, Arabic, a language family spoken by patients from across North Africa and the Middle East, is known for its nonconcatenative morphology [475]—a manner of forming words in which a root is modified instead of stringing morphemes together. In addition, vowels are often left out of written text. This results in a highly contextual language; for example, the letters k-t-b (كتب ;ك-ت-ب) in Arabic can be read as 'kutub' (books) or 'kataba' (he wrote) depending on the context. Language tests developed for Arabic speakers should be sensitive to errors in such aspects of the Arabic language. Other global languages are faced with their own unique challenges, such as aspects of tone and orthography in languages spoken in China [476]. Errors in these language characteristics currently remain undetected and unassessed in the multicultural memory clinic. Collaborations with native speakers, possibly in the country of origin, will be needed for the development of assessment tools and techniques. Furthermore, technological advances may allow for the identification of errors through automatic speech analysis in the future (as was applied in the case study by Taiebine & El Alaoui Faris [475]).

New test paradigms are likely needed to assess the cognitive domain of attention/mental speed. Even when patients are urged to work as fast as they can, performance on currently available tests is influenced by culture/acculturation despite the use of tests that are specifically designed with diverse populations in mind, such as the Five Digit Test [21]. The international literature suggests this may be due to differences in attitudes towards time and/or the exposure to timed testing in the educational system across cultures [20,477]. A more ecologically valid approach with high face validity may be required to ensure patients truly respond as fast as they can. Several real-life situations are imaginable that would elicit an equally fast and attentive response across diverse populations, such as responding to a fire alarm. Perhaps such contexts provide a better opportunity to measure mental speed/attention than the current paradigms.

Tests of executive functions tend to be relatively complex on a conceptual level (chapter 2.3). Furthermore, perhaps due to the *supramodal* nature of executive functions [478], tests of executive functioning, even more so than other domains, seem to require intact functioning across multiple cognitive domains. For example, the Zoo Map test of the Behavioral Assessment of the Dysexecutive Syndrome requires 1) intact language and memory to understand and remember the instructions, 2) intact sustained attention and working memory during the task, 3) intact visuoperception and gnosis (including reading)

to read instructions, the map itself, and to identify targets, and 4) intact fine motor skills to draw the route, in addition to the executive functions it is supposed to measure. This complexity likely contributes to the difficulty of designing tests of executive functioning that can be applied in culturally, educationally, and linguistically diverse individuals. In addition, the abstract reasoning skills that tests of executive functioning often rely on may make test development even more challenging, given the issues with abstract reasoning in culturally and educationally diverse individuals [33]. A useful approach may be to study tests with more promising face validity and ecological validity. However, it is important to ensure such tests are appropriate for the specific target population. For example, consultation with diverse individuals revealed that one of the 'ecologically valid' tests considered for inclusion in the TULIPA battery, the Pillbox Test [479], likely was not suitable for older diverse individuals given their lack of experience using a pillbox. Additionally, a study showed that, although the Pillbox Test has evident face validity, its ecological validity was surprisingly low compared to another test with low face validity—the Push-Turn-Taptap test [480]. Studies utilizing novel approaches such as virtual reality environments (as suggested in chapter 2.2) might present new opportunities to investigate this domain in diverse populations, but researchers will have to consider all aspects of validity in developing such tests.

A paradigm shift may also be necessary for visuospatial perception, including visuoconstruction; traditional tests of visuoconstruction are not suitable for (educationally) diverse individuals, often displaying poor psychometric characteristics (chapter 2.2). The Stick Design Test, although more acceptable to patients [112,481], has occasionally displayed ceiling effects [482] as well as substantial variation in test scores between countries. For example, whereas healthy controls in Brazil obtained a median score of 12 (SE: 0.04) on this test, the controls in Nigeria had a mean score of only 8.2 (SD: 3.1). Whether this test is indeed cross-culturally applicable therefore remains to be seen; other alternatives likely need to be sought as well. Innovative techniques, such as the use of virtual reality or eye-tracking may provide an interesting new perspective on measures of visual processing. A substantial number of eye-tracking studies have investigated cultural differences in scene/object perception, with heterogeneous results [483]. A study in a diverse population in Europe highlighted the challenges associated with such paradigms. For example, several studies have demonstrated how individuals from collectivist—specifically, East Asian—cultures may pay more visual attention to contextual details (holistic processing), as opposed to people from individualistic cultures, who supposedly focus more on prominent focal features (analytic processing) [484]. However, these traditional theories could not directly be applied to a diverse sample of Asian—in this case, Indian—and British participants in Europe [485].

Social cognition was identified as a cognitive domain that is very challenging to assess and was designated a priority for future research (chapter 2.3). However, developing cross-culturally applicable tests for social cognition is likely challenging, given the substantial influence of culture on sociocognitive processes. Cross-cultural differences in social cognition already occur early in social development [486] and seem to persist into later life—although developmental trajectories are similar [487]. In healthy adult populations, even the recognition of emotional expression of supposedly 'universal' emotions varies by country [261]. Cultural differences explained almost 21% of the variance in performance

on an emotion recognition test and 25% on a faux-pas test [261], which, according to the authors, is an almost ten times larger effect than cultural effects on memory and attention. Similar cross-cultural differences were found in other studies, such as those investigating the Reading the Mind in the Eyes test measuring theory of mind [488]. These differences do not necessarily influence the tests' ability to discriminate patients with specific types of dementia from controls (see e.g. [489]), provided that norms are available for the target population. However, they do call into question the construct validity of these tests; newly designed or adapted tests may be needed for a more valid assessment. Tests of emotion recognition and theory of mind using faces as stimuli—such as the Ekman 60 Faces Test or Reading the Mind in the Eyes test—may need to include photographs of individuals from the same ethnoracial background as the target population, given the differences in performance depending on whether the individuals portrayed have a similar or different ethnoracial background [490,491]. Similarly, theory of mind and social norms tasks need to be adapted to suit local social rules and norms, such as in the adaptation of the Social Norms Questionnaire [492]. As it stands, however, the assessment of social cognition in culturally, educationally, and linguistically diverse populations in Europe mostly is uncharted territory.

A final aspect of cognitive functioning that was not included in the studies in this dissertation—but was identified as an important gap in assessment in chapter 2.3—is intelligence. In clinical practice, I have seen several examples of physicians and neuropsychologists who interpreted limited premorbid education levels in diverse individuals as a sign of intellectual impairment—as was the case with mr. A in chapter 2.1. Although I acknowledge that estimating intelligence is difficult given the lack of adequate tools available, I find this very disconcerting. Although the historical contributions of non-verbal tests of intelligence, such as the Raven Progressive Matrices [493], to our field should be acknowledged, clinicians and researchers should by now be aware of its limitations in cross-cultural settings [344,494,495] and should be very hesitant to administer such tests in diverse populations. Nonetheless, these non-verbal intelligence tests are still used, as evidenced by several chapters of a Dutch compendium on culture and psychodiagnostic assessment [496]. Strikingly, the Raven Progressive Matrices test was even recommended as an instrument to measure intelligence in diverse populations in the Netherlands by the national institution tasked with the evaluation of (neuro)psychological tests, the 'Commissie Testaangelegenheden Nederland' [497].

Chapter 3: Improvements to the field of cross-cultural neuropsychological assessment

5.1.2 Neuropsychological assessment of diverse individuals is feasible, as long as the test battery is not too long and factors complicating the neuropsychological assessment are taken into consideration

The newly composed TULIPA test battery, which includes tests such as the CCD, RUDAS, mVAT, NAME, and subtests of the CNTB, was generally feasible, with the exception of the CCD Dots subtest B (chapter 3.1). Interestingly, the pilot version of the CCD originally contained three tests of executive functioning: the Sun-Moon test, a Card Sorting Test, and the 'Hands test'—a version of the CCD Dots test in which, instead of dots, the digits were represented by means of the fingers on two hands. Ultimately, however, the Card

Sorting Test proved too difficult; in addition, the hands in the 'Hands test' were replaced with dots on 'dominoes' in the final version of the CCD as these stimuli were presumed to be simpler to count and also more widely applicable across cultures [58]. Based on our findings, however, it seems the 'dominoes' with dots on them were not intuitive to all patients visiting memory clinics. Given the limited predictive validity of the Dots test beyond the Objects and Sun-Moon subtests of the CCD [58], I would therefore recommend clinicians to only use the CCD Dots test if there are convincing reasons to do so in clinical practice. Instead, clinicians may focus on covering a wider range of cognitive domains, including naming (language) and visuospatial or constructive abilities, or replacing the Dots test with the Five Digit Test included in the TULIPA test battery. Although not all tests have demonstrated validity in diverse populations in the Netherlands specifically (yet), these tests at least showed promising psychometric properties in international research and showed high completion rates in chapter 3.1. Another recommendation is to interpret all CCD subtests separately instead of applying the risk of cognitive impairment score—the neuropsychologists in the focus groups in chapter 3.1 agreed the individual scores better matched their clinical impression of the patient than the overall score.

Chapter 3.1 also shows that it is important to monitor factors that can complicate the neuropsychological assessment of diverse (older) adults visiting memory clinics, such as fatigue, depression, and suboptimal effort. The potential consequences of invalid neuropsychological assessment are manifold; patients may not receive any or an inappropriate diagnosis—in turn resulting in suboptimal care—patients may be frustrated that the (often) lengthy neuropsychological assessment was inconclusive, and valuable time of patients and clinicians is misspent. Neuropsychologists should adapt their clinical practice where possible to optimize testing conditions. Furthermore, neuropsychologists could consider using screening questions, such as those sometimes used in studies of the GDS [244], to examine whether such factors are present before formally starting the cognitive assessment.

5.1.3 Neuropsychological assessment of diverse individuals can be improved substantially by including more widely applicable, colored stimuli in test design

The findings in chapter 3.2 highlight the importance of using colored stimuli as opposed to black-and-white line drawings in the assessment of diverse individuals; we therefore used colored stimuli for the NAME in chapter 3.3 as well. Although the literature in the Netherlands—dating from before the CCD was developed—recommended the use of the Location Learning Test, Visual Association Test and SAN-test for memory and naming [52], our studies indicate that such tests are less suitable for (educationally) diverse populations due to their black-and-white line drawings.

In addition to whether items are portrayed in black-and-white, another important factor to take into consideration is the appropriateness of the stimuli themselves. Although the hedgehog in the mVAT, for example, does not seem culturally appropriate, the reliability analysis of the mVAT indicated that only the combination of the syringe and leaf had to be replaced. This likely has to do with the scoring of this test: as no points are deducted for misnamed/misidentified objects in the mVAT, it is not necessarily a problem if the patients mistake one item for something else, such as a hedgehog for a brush or a "prickly animal". In

contrast, in developing the NAME we made an effort to only include culturally appropriate items, as misidentification of the objects would hinder optimal performance in this case.

5.1.4 Appropriate measurement of caregiver burden is needed to identify caregivers of diverse patients in need of intercultural support

Chapter 3.4 highlighted the high levels of caregiver burden experienced by caregivers of diverse patients visiting our multicultural memory clinic. Our goal was to specifically use an instrument that could capture both 'positive' and 'negative' aspects of care; however, our findings showed that the positive subscale of the Caregiver Strain Index-Expanded provided limited additional information. In an effort parallel to our own, other researchers in the Netherlands developed a Turkish version of the Self-Perceived Pressure from Informal Care instrument [498], investigating the degree to which caregivers can balance "the demands of the caregiving situation in comparison with the personal interests of the caregiver" [499]. This instrument has promising psychometric properties, with the limitation that it was mostly studied in female caregivers. Although several studies of diverse populations in the Netherlands emphasize that there is often one female primary caregiver [327,500], our own study sample in chapter 3.4 seems to call this into question, as 30% of the adult child caregivers in our study were male. Similarly, other Dutch researchers have recently called for a more nuanced view of caregivers of diverse patients that does justice to the complexity of and variation in care in diverse families [501]. Mixed methods approaches are likely needed in which interviews or focus groups inform the design of quantitative scales or surveys to measure caregiver burden and caregiver roles in a large, representative population. This may then inform policies regarding intercultural care.

Chapter 4: Implementation: diversity in clinical practice and research

5.1.5 European collaboration is needed to address issues in cross-cultural neuropsychology for all diverse populations in Europe

Chapters 2.3 and 4.2 both highlight the need for collaboration in the development and implementation of (widely applicable) neuropsychological tests. In the Delphi study in chapter 2.3, a hierarchy of priorities was established—first tests of social cognition and language, second tests of executive functioning, visuospatial functioning, working memory, and orientation, and last, memory, mental speed, and attention. However, priority ratings for all cognitive domains were relatively high, indicating that neuropsychologists experienced a need for test development across all cognitive domains. Although ECCroN (chapter 4.2) recommends the development of widely applicable neuropsychological tests, it is unclear if it will ultimately be possible to develop one test battery that can be used to assess everyone in the European population, including individuals without a migration background. Although several studies, such as those of the RUDAS [206] and the CNTB [140] indeed attempted to do so by including a European 'majority' population, our findings for the NAME (chapter 3.3) indicate that there may be a trade-off—at least for some cognitive domains—between how widely applicable versus how difficult a test can be made. For example, in the case of the NAME, a more difficult item may also have a more variable age of acquisition, word frequency, or familiarity across languages, and adding such an item may make the NAME less widely applicable—but more sensitive in detecting naming impairment early on in the disease for some populations.

In addition to neuropsychological tests, chapter 4.2 also stresses the need for improvements to clinical training in cross-cultural neuropsychology, with the possibility of developing a European best practice or clinical training program; given the wide variety in clinical training models across European countries [223,502], European-level guidelines and training courses may be an optimal first step. Similar efforts have been ongoing in the field of dementia in Latin America, where a best practice guideline in Spanish was recently developed by the Latin American and Caribbean Consortium on Dementia, the Multipartner Consortium to Expand Dementia Research in Latin America, the Global Brain Health Institute, and the Alzheimer's Association [503]. There are several opportunities for future research in this area, which are described in more detail in the section on *Future directions.*

5.1.6 Inclusion of diverse individuals in research requires a change in consent procedures, in the eligibility criteria, and in the instruments used as screening and outcome measures

The review in chapter 4.3 illustrates that there is a continuous, systemic underrepresentation of ethnoracially diverse groups in global clinical trials. This issue certainly is not limited to clinical trials. For example, a review of studies including patients with a dementia diagnosis published in top journals showed that study populations were often from North America and Europe, invariably including a predominantly White study sample, with a median of 89% White participants (IQR 78–97% [504]). This IQR is remarkably similar to our findings in Aβ and tau trials. A limitation of our study in chapter 4.3 was that we were unable to determine the direct effect of each criterion on the representation of diverse individuals using inferential statistics. Several factors precluded such analyses, such as the fact that some criteria were used either very infrequently or invariably (e.g. the MMSE, Supplementary Table 2), as well as the fact that race/ethnicity data was not reported for each global region/country. A recently published study of preclinical AD (in the USA), however, did demonstrate why fewer diverse participants may be included in study samples; there were significant differences in 1) the referral sources by racial/ethnic group (e.g. local vs. centralized recruitment), and 2) the number of ethnoracially diverse participants who met the eligibility criteria [505]. Specifically, diverse individuals often screen-failed on the MMSE (in the case of Latino/a participants), CDR (Black participants), and on a logical memory test (Black participants [505]). Issues with such measurement tools are not limited to tests used to screen for cognitive impairment, but probably also occur in outcome measures used in dementia research. More research aimed at alternative screening- and outcome measures is long overdue.

5.2 Methodological considerations

There are several limitations/methodological considerations relevant to the studies in this dissertation, beyond those discussed in the individual discussions of the papers. In the following paragraphs, I will discuss the accuracy of the diagnoses, issues of blinding and circular reasoning, the scarcity of normative data, the validity of questionnaires, the representativeness of study samples, and limitations to neuropsychological data and statistical techniques.

5.2.1 The accuracy of the diagnosis in diverse individuals and issues with blinding and circular reasoning

Diagnosing dementia in diverse individuals is challenging. A Danish hospital registry study revealed an underdiagnosis of dementia in diverse individuals younger than 60 years, and an overdiagnosis of dementia in those over 60 [75]. The authors attributed these findings in part to diagnostic challenges such as language barriers and inadequate cognitive instruments. The majority of the diagnoses of the diverse patients in this Danish study were registered as 'dementia not otherwise' specified (48%), followed by AD and VaD. Not a single individual with a diverse background had a registered diagnosis of FTD [75]. These findings reflect the difficulties in determining the etiology of the dementia in diverse populations, which was also an issue in this dissertation (chapters 3.2, 3.3, and 3.4). A subset of our patients remained without a formal diagnosis, even when using additional biomarkers. Several factors likely contribute to this issue. First, it is probably in part due to multimorbidity, such as the presence of diabetes, depression, and cognitive complaints in the same individual; a large case-control study in a multicultural memory clinic in Amsterdam revealed that symptoms of depression (56%) and diabetes (47%) were often present in diverse individuals visiting memory clinics—significantly more often than in the Dutch control group [506]. Second, cognitive tests and appropriate norms are missing for some cognitive domains and some diverse groups, leading to difficulties in the differential diagnosis (see also the sections on *The scarcity of normative* data and *Future directions* below). Ideally, the patients from our multicultural cohort would receive long-term follow-up to determine the accuracy of the diagnosis. This would also provide an opportunity to study whether performance on these tests declines as the disease progresses. Together with formal studies to establish test-retest reliability and potential learning effects, such studies may ultimately determine whether these tests will be suitable for longitudinal studies, including clinical trials.

Our diagnostic accuracy studies of the mVAT and NAME also carried a risk of circular reasoning due to a lack of blinding—that is, neuropsychologists who reported on the results of the neuropsychological assessment were not blinded to the score on the NAME or mVAT, and the test scores may therefore have influenced the final diagnosis. This may have been relatively more problematic for the mVAT, as it is a modification of an existing measure that is often used by Dutch neuropsychologists; most neuropsychologists are aware that ceiling effects often occur in this test, and that a small number of errors therefore likely indicate memory impairment. Issues of blinding may have had less impact on the NAME, as the neuropsychologists had no information available on the performance of healthy controls on this test, and therefore could not know what range of scores should therefore be considered 'normal'. Although the diagnosis was not based solely on the neuropsychological assessment and imaging data was often available, it cannot be ruled out that blinding issues impacted our findings in these two studies.

5.2.2 Normative data and assessment of diversity-related variables

A lack of norms was highlighted as a major issue by the experts in the Delphi study in chapter 2.3. Although norms are available for some tests from the CNTB and the CCD, they do not cover patients of all ages and all nationalities. For example, the norms for the CNTB for Moroccans are very limited (n = 14, mean age: 58, standard deviation: 8 [140]); similarly, although a substantial number of Moroccans were included in the normative data study

of the CCD (n = 232), no Moroccans of 80 or older participated [58]. An additional problem with norms is that they are often stratified by variables such as ethnocultural group, age, and education. This may result in very small sample sizes per stratum. An interesting alternative would be to use approaches such as (non-parametric) continuous norming [507], in which statistical models are used that take into account all characteristics relevant to performance on the test. This may improve statistical power.

The European Consortium on Cross-Cultural Neuropsychology (ECCroN), which was established in 2019 to improve cross-cultural neuropsychological assessment in Europe, provides an optimal platform to collect normative data for smaller populations that are scattered across Europe (chapter 4.2). Furthermore, it may also facilitate a comparison of the performance of intra-EU immigrants with their peers in their country of origin to see if the original norms apply (for example, comparing Italian labor immigrants in the Netherlands with Italians living in Italy). In the collection of norms and interpretation of the entire neuropsychological assessment, it is important to take into account a variety of diversity-related variables, but it currently is not clear how these variables should be measured in the European context (chapter 4.2). A good example is acculturation, "the processes by which groups or individuals adjust the social and cultural values, ideas, beliefs, and behavioral patterns of their culture of origin to those of a different culture" [508]. There seems to be little consensus on the best way to measure acculturation, with several dimensions described in the literature, such as acculturation conditions, acculturation orientations, acculturation outcomes, acculturation attitudes, and acculturation behaviors [509]. A study from the USA highlighted that several different acculturation factors may be relevant to investigate regarding the relationship between acculturation and neuropsychological outcomes in diverse populations [510], such as language aspects, social aspects (including discrimination), and familism. In Europe, the main approaches to assessing acculturation in cross-cultural assessment are 1) to use years of residence in the host country, 2) to use adapted acculturation scales from the USA, or 3) to calculate the cultural distance based on the difference on Hofstede's dimensions of national cultures between the dominant culture and the culture of the country of origin of the patient [21]. In line with our European colleagues, we mainly used years living in the Netherlands and the score on the language subscale of an adapted version of the Short Acculturation Scale for Hispanics [215] for this purpose. Whether these variables sufficiently capture acculturation in the European context, however, requires more research.

Another example is quality of education, which can be measured in a number of ways. In the USA, approaches that have been used include investigating reading level, student-to-teacher ratios, classroom size, and length of the academic year [22,39]. In recognition of this important variable, we have developed a screening tool for literacy as part of the TULIPA test battery, with adequate feasibility (chapter 3.3); a formal analysis of its validity will soon follow. Although quality of education and acculturation are two examples, several other variables are currently not well-defined and measured in cross-cultural neuropsychological research in Europe, such as aspects of language (e.g. proficiency, bilingualism). ECCroN is currently reviewing the ways in which these variables are measured to work towards harmonized methods of data collection in Europe.

5.2.3 The validity of questionnaires

In 2017, we started with the translation/adaptation of questionnaires to use in the assessment of Amazigh (Moroccan-Berber), Moroccan-Arabic, and Turkish patients. Translating and adapting these questionnaires was challenging. For example, for a Tarifit translation (northern Moroccan dialect from the Rif mountains), our questionnaires were first translated by a professional translation agency; however, the translator frequently selected words from Standard Arabic when a direct translation to Tarifit was not possible. In the end, four native speakers with a background in medicine/psychology had to be consulted before the drafts were sufficiently suitable to the lower educated and elderly Moroccan population in our clinic. It remains to be evaluated, however, whether these adapted questionnaires, such as the adapted/translated Geriatric Depression Scale (GDS [244]) are equivalent across individuals with cognitive impairment with different ethnic backgrounds. In a previous study of the GDS, differential item functioning was found for some items in some of the included diverse groups [244]. Research from other countries also highlights issues with equivalence, which are hypothesized to be due to cross-cultural differences in the concept of depression [511]. In addition, the choice of words is particularly important in this matter; for example, some languages are suggested to not have a word for "depressed" [512]. Furthermore, in individuals with dementia, interactions may occur between reporting of depressive symptoms and cognitive impairment that may further complicate the matter; for example, the GDS had low sensitivity and specificity for depression (as diagnosed by a clinical interview) in patients with AD, whereas the instrument performed well in patients with MCI [513]. Given these issues, the large number of individuals with a high score on the GDS in chapter 3.2 is difficult to interpret. It remains unclear whether these very high levels of self-reported depressive symptoms are indeed indicative of severe depression, or whether they can be interpreted in another way; for example, they may be considered a 'cry for help' reflecting cultural differences in idioms of distress [514,515], or they may reflect overreporting due to factors such as symptom misinformation, inattentive responding, or due to the order of administration of tests [516].

5.2.4 Representativeness of the sample

Clinical samples often are not representative of the general population and results may therefore not be generalizable to the population at large [517,518]. This undeniably also applies to the patient samples included in our studies (chapters 3.1, 3.2, 3.3, and 3.4). For example, even though the numbers of Moroccan and Turkish elderly are approximately equivalent in the Netherlands [519], a proportionally larger number of Turkish individuals was included in our studies (chapter 3.1: 1.7x as many, chapter 3.2: 2.6x, chapter 3.3: 2.1x, chapter 3.4: 1.9x). Similarly, the populations included in our study are higher educated than the general population. This is not necessarily problematic as long as these instruments are mainly used in memory clinic populations. However, issues with representativeness should be taken into consideration should the tests be used in e.g. epidemiological studies investigating incidence, prevalence, or risk factors for dementia. For example, the associations between cognitive and functional outcomes may differ in epidemiological versus memory clinic settings [517]. In addition, the findings regarding caregiver burden (chapter 3.4) should also be interpreted as specific to a memory clinic context. The level of caregiver burden in caregivers of patients who did not visit memory clinics is unknown. It is not unlikely that caregiver burden is higher if patients are not accessing services due to e.g. limited (health) literacy, lack of familiarity with the health system, or shame/stigma.

It could also be, however, that only those individuals with the highest caregiver burden decide to visit the memory clinic.

5.2.5 Limitations to neuropsychological data and subsequent statistical analyses
Neuropsychological data is often skewed due to floor or ceiling effects, and transformations to the data and/or nonparametric analyses are often required. In addition, the effect of demographics and other relevant variables may be non-linear in nature, as was apparent for age and education on NAME performance (see chapter 3.3, Supplementary Figures 3A–3C). Statistical techniques such as the Generalized Additive Models (GAM) used in chapter 3.3 may be helpful because they allow for nonlinear effects in the data; however, these techniques are still relatively new—being first conceptualized in the 1990s—and consensus is not always available on best ways to report and interpret effect sizes and p-values. In addition, overfitting may be an issue in GAMs, especially in smaller datasets [520], and little consensus seems to exist on the best way to determine whether overfitting has occurred.

5.3 Future directions

In the following section, I describe what I envision are the next steps for research and clinical practice. I will successively discuss my perspective on 1) the phase leading up to the diagnostic assessment, 2) the diagnostic trajectory itself, and the implementation of our findings in 3) clinical practice, and 4) research with diverse populations.

5.3.1 The prediagnostic phase: dementia awareness and detection in diverse populations
Previous work has highlighted that diverse individuals may not recognize the initial stages of dementia as a disease, attributing these symptoms to normal aging or other causes instead [54,308]. Likewise, general practitioners report that diverse patients first present with dementia-related symptoms at a late stage [521]. In our community education program for diverse individuals, we frequently encountered instances of mislabeling of cognitive symptoms; for example, one participant reported how her 85-year-old husband was purposely irritating her by praying in the wrong direction (not facing Mecca), or by making errors in the prayer ritual. On further inquiry, she had not considered that such behavior may also be a sign of MCI or dementia. To increase dementia awareness among diverse individuals, researchers from the applied university of Rotterdam ('Hogeschool Rotterdam') developed an information sheet listing such possible presenting symptoms of dementia [522]. Although these initiatives contribute to dementia awareness, large-scale projects to reach a wider population are needed. Therefore, together with colleagues from the National Aging Research Institute of Australia, Samvedna care in India, and colleagues from the Latin American and Caribbean Consortium on Dementia, we applied for and received funding to create a global repository of culturally and linguistically appropriate dementia resources (GENIE) in a co-creation process with diverse individuals. In this repository, caregivers and clinicians will be able to find high-quality information about dementia in a wide variety of languages, allowing better services even for small/scattered communities.

As previously mentioned, diverse individuals may also experience barriers to accessing care, such as due to language barriers or a lack of familiarity with health care services [52,308]. In general, diverse individuals in the Netherlands visit general practitioners more often than native Dutch individuals [523]. It therefore seems likely that, in addition to a

lack of dementia awareness and problems with late consultation of general practitioners, issues also arise in referral by general practitioners. In a mixed-methods study, general practitioners reported barriers in providing services to diverse individuals due to a lack of—or unfamiliarity with—suitable screening tools for dementia for diverse populations and a lack of knowledge about culturally appropriate services and communication tools [521]. The study report mixed findings regarding whether patients and caregivers themselves would like to be referred to a memory clinic.

The combination of a lack of dementia awareness, barriers to accessing dementia diagnostic services, and issues with referrals may explain the diagnostic delay and late presentation at memory clinics reported in some studies [506,524]. Future research should focus on initiatives to promote dementia awareness and examine how to improve diagnostic services in, and collaboration between, primary and secondary care providers. Some of these issues will be addressed as part of the new national ABOARD consortium focusing on primary and secondary prevention of Alzheimer's disease.

5.3.2 The diagnostic trajectory: towards a comprehensive cross-cultural neuropsychological assessment in the Netherlands and Europe

There are numerous possibilities for future studies addressing the diagnostic trajectory of diverse individuals. Given the focus of most of my work on memory clinics, I will therefore structure this section by systematically describing opportunities for future research regarding each of the core clinical criteria for all-cause dementia as described by McKhann et al. [96].

Criterion 1 and 2: "*There are cognitive or behavioral (neuropsychiatric) symptoms that interfere with the ability to function at work or usual activities; the functioning represents a decline from previous levels of functioning.*"

Ideally, impairment in basic and instrumental activities of daily living (ADL and iADL) would be measured systematically. However, little is known about the validity and reliability of traditional scales measuring functional impairment in diverse populations. Even in the general population, studies of instruments measuring iADL are often of suboptimal quality [525]. Those studies that have investigated (i)ADL-scales across countries generally did not include the culturally, educationally, and linguistically diverse populations that are the focus of this dissertation (e.g. [526]). Remarkably, cultural adaptations were even needed to adapt a Dutch instrument for use in the UK [527], as everyday activities like "using a coffee maker and a dishwasher" were identified as less appropriate for the UK; according to the authors, these activities were not common practice among older adults living there [527]. Diverse elderly individuals in particular may always have been dependent on others for some iADL due to reasons other than cognitive impairment, such as doing finances/administrative tasks due to problems with literacy or limited (Dutch) language proficiency [52]. Additionally, responsibilities regarding everyday household tasks may have been transferred to others in case of physical impairment. Qualitative studies may be needed to determine which activities should be included in scales that can quantitatively measure functional impairment in diverse individuals in the Netherlands. As an example, researchers in an Indian study asked professional experts, field workers, and village leaders in rural India which activities elderly individuals would still be expected to engage in, and developed an (i)ADL-scale based on these activities [528]. This scale includes items

inquiring whether patients still express their opinions on important family matters (such as marriages) and whether they are able to remember important festivals (Diwali and Holi). In addition to developing more suitable functional impairment scales, it may be necessary to look for other ways to assess this aspect of dementia, such as through a home visit by an occupational therapist [521].

Criterion 3: "*These cognitive/behavioral symptoms are not explained by delirium or major psychiatric disorders"*

Some of the neuropsychiatric symptoms reported by patients or caregivers may be challenging to interpret in diverse populations due to differences in explanatory models, presentation, and experiences of disease. For example, there are several examples of 'culture-bound syndromes' [93] or 'cultural concepts of distress' [251], such as *nervios* in individuals from the Americas or *brain fag* in Western Africa. In the latter, individuals experience distress from thinking too much, with symptoms including headache and an experience of a worm crawling in the head; in the former, an idea is 'stuck to one's mind', which is said to lead to a slow deterioration of the mind, including possible panic attacks and dissociative features [529]. Such culture-specific presentations may be mislabeled, misdiagnosed and mistreated if they are not properly recognized. An example from my own clinical practice is when patients report feeling or perceiving the presence of a deceased relative. It is very challenging to determine whether such symptoms should be interpreted 1) as hallucinations due to dementia or delirium, 2) as (culturally-mediated) experiences of grief, or 3) as a general psychotic episode (unrelated to grief). The interpretation becomes even more challenging if their relative passed away relatively long ago, or if the perceptual experience includes multiple senses (e.g. visual, auditory, and tactile). Although some literature tries to provide recommendations on how to discriminate between these different causes [530], clinical practice remains challenging. More research should be directed at the way (neuro)psychiatric symptoms are initially presented in memory clinic populations, as well as at culture-sensitive tools to measure behavioral and psychological symptoms. For example, to my awareness, no studies exist that address whether the measurement instruments that are currently used in native Dutch populations, such as the Neuropsychiatric Inventory (NPI), are psychometrically sound in diverse populations in the Netherlands as well.

Criterion 4 and 5: "*Cognitive impairment is detected and diagnosed through a combination of 1) history taking from the patient and a knowledgeable informant and 2) an objective cognitive assessment; the cognitive/behavioral impairment involves a minimum of two cognitive domains"*

During my research project, I contributed to a project aimed at improving informant-based history taking led by the OLVG hospital in Amsterdam. In this study, we investigated the value of an informant-based questionnaire to measure cognitive decline, the Informant Questionnaire on Cognitive Decline (IQCODE [242]). This is the first structured informant-based questionnaire that is validated in diverse memory clinic populations in the Netherlands [243]. The study replicated findings from a previous study demonstrating the added value of combining IQCODE with the RUDAS in the diagnosis of dementia in an Arabic population in Lebanon [531]. The IQCODE seems to be a suitable tool to assess proxy-rated cognitive impairment.

The studies presented in this dissertation have mainly contributed towards better assessment of individual cognitive domains that cannot be assessed with such screening tests. The *Main findings* section already covered some remaining gaps and opportunities for future research for specific cognitive domains. Here, I will highlight some additional ideas. A logical first step to improve neuropsychological assessment would be to analyze the reliability and validity of the individual tests included in the TULIPA battery, so that all cognitive domains can then be adequately assessed. Good coverage of all cognitive domains will contribute towards identifying the underlying etiology of the dementia, which may be particularly important for neurological and neurodegenerative diseases that have been studied less frequently in diverse populations, such as Parkinson's disease dementia, Lewy body dementia, or frontotemporal dementia (FTD). Although most research is focused on AD (chapter 2.2), research and clinical practice regarding these other neurodegenerative diseases naturally is facing similar issues in cross-cultural neuropsychological assessment [532,533]. To address issues in FTD research, I am currently co-leading an international workgroup studying gaps in the literature on diversity in FTD research covering all disciplines, including genetics, epidemiology, and neuropsychology. In addition to this workgroup, our new project 'Cross-cultural neuropsychological assessment of social cognition' in collaboration with the University of Paris, will examine and develop suitable measures of social cognition in diverse European settings, making use of technological advances such as VR.

In addition to test development, extensive normative data need to be collected for the TULIPA tests so that clinicians can determine whether an individual patient's performance is impaired. These tests and/or norms can be published formally by a publisher, as we have sought to do for the mVAT (in 2021) and potentially for the NAME (2022). Collecting data in healthy controls also allows for analyses of differential item functioning across diverse individuals, which should provide valuable information on the generalizability across groups. Similarly, detailed analyses of the individual subtests and items may further improve the tests' user-friendliness (as described in chapter 3.1). Qualitative analyses of neuropsychological data in native Dutch patients have already been proven to be valuable, such as the number of word clusters and number of switches between those clusters in animal verbal fluency (language [534]); however, we were unable to replicate these findings in a study of animal verbal fluency in diverse patients from the multicultural memory clinic (results not published).

5.3.3 Implementation of our findings in clinical practice

5.3.3.1 Training of clinicians and use of interpreters

Since the multicultural memory clinic opened in 2015, we have received a substantial number of questions from professionals across the Netherlands on how to improve diagnostics and care for their diverse patients. Some of these clinicians saw only a handful of diverse patients a year, while others assessed numerous patients with a variety of diverse backgrounds. In some centers, formal interpreters were available, but many others lacked funding, resulting in assessments with an informal interpreter. Although the CCD is designed so it can—in theory—be administered with audio recorded instructions, my personal experience in clinical practice is that interpreters are often necessary to provide additional instructions/clarifications, as well as to understand which cognitive functions contribute to abnormal performance—e.g. did the patient forget the instruction

(memory), did he/she not understand the instruction (language), or are there dysexecutive symptoms? Unfortunately, the number of TULIPA tests that can be administered without an interpreter is limited. Assessing patients with an informal interpreter has, throughout this dissertation, been advised against in most cases (e.g. chapters 2.3 and 4.1). However, it would be interesting to formally compare the outcomes of the neuropsychological assessment in carefully matched samples of patients assessed with an informal vs. a formal interpreter present, to examine the degree to which the type of interpreter may influence the outcome. It would also be interesting to study whether a thorough briefing of the interpreter (formal or informal) can improve the quality of the interpreter-mediated assessment. To improve interpreter-mediated assessment, ECCroN is currently developing specific guidelines on interpreting during neuropsychological assessment. For example, how should interpretation take place during category fluency tests—given that live translation may disrupt the patient's thought process?

It is also unclear what level of training and experience is required to successfully carry out a cross-cultural assessment. Although diversity seems to be receiving more attention in clinical psychology training in the Netherlands, such as in the development of training programs [535] and mental health care standards [337], a lot of ground still needs to be covered in neuropsychology. Some efforts have been made in the field of dementia, such as the development of a supplement on cross-cultural assessment in the neuropsychological guidelines for MCI and dementia [536]. Similarly, diversity is now explicitly addressed in the national health care standards for dementia [537], as well as in national standards for dementia for general practitioners [538]. However, the information included in these standards is brief; training to increase cultural competency is likely needed. Therefore, in addition to the clinical recommendations provided in chapter 4.1, we are currently focusing on the development of a best practice and (post-master) masterclass in cross-cultural neuropsychology. It may be beneficial to base such clinical training initiatives on models of intersectionality instead of focusing only on a specific aspect of diversity. Intersectionality "acknowledges that aspects of diversity are not simply cumulative, but also determine, in dynamic, mutual interactions, the position that a person occupies in society, and as such also shapes their experiences of exclusion, inclusion, power or disadvantage" [535]. An intersectionality approach allows for an analysis of the interplay of these factors in the neuropsychological assessment. For example, how does the combination of being young, high educated, female, and Dutch impact the assessment of a patient who is older, low educated, male, and Turkish? Such an intersectionality approach also aligns well with the ECCroN goals outlined in chapter 4.2.

Although beyond the scope of my dissertation, a last point that is important to raise regarding clinical practice is the limited diversity among memory clinic staff in the Netherlands, which was also raised as an important issue in cross-cultural neuropsychology in Europe (chapter 4.2). For example, estimates indicate that only 2%–4% of medical specialists have a culturally diverse background ("allochtoon" [539]). Given this limited diversity, my dissertation mainly focused on approaches to interpreter-mediated diagnostic assessments with memory clinic professionals of any cultural or linguistic background. However, some patients may have a preference for ethnic matching—depending on factors such as the patient's sex, ethnicity, education level, and years in the Netherlands [540,541]. A more diverse workforce is needed to meet the needs of this patient group. A first step towards

this goal—which requires efforts from national institutions, however—would be to gather data on the representation of diverse individuals among Dutch neuropsychologists and to subsequently develop strategies to improve representation.

5.3.3.2 Populations requiring further study: underrepresented diverse subgroups

The patients included in the studies in this dissertation are predominantly first-generation immigrant patients from North Africa, Turkey, and the Middle East—populations that were targeted in particular because of their low levels of education and limited proficiency in Dutch. Patients from former Dutch colonies, such as current-day Suriname, Indonesia, and the Dutch Antilles were represented in some of the study samples, but overall received less attention—a limitation reflected in the general literature on diverse elderly individuals in the Netherlands [542]. Future studies should identify the common and unique issues in neuropsychological assessment of these populations, as well as the optimal test battery for this population. In addition, the next decades will see a rise in patients who are descendants of first-generation immigrants, who may be more fluent in the Dutch language, but nonetheless may experience cultural differences. More research is needed to determine which tests and norms should be used to test these populations. It would also be interesting to examine whether the neuropsychological tests that are currently used in low-educated older adults who are born in the Netherlands are sufficiently suitable, or whether they would require adaptations similar to the ones made to improve tests for low-educated diverse populations; for example, tests such as Clock Drawing and Trail Making may be challenging for Dutch individuals with limited numeracy and literacy.

5.3.4 Implementation of our findings in research

5.3.4.1 Research literacy, informed consent, and participation

Throughout my research project, we explored several avenues to improve the enrollment of diverse individuals in research. Some of these efforts revolved around 'research literacy', which encompasses "1) knowledge of research concepts, 2) attitudes towards research, 3) self-efficacy in the ability to weigh participation decisions, 4) increased motivation to explore research options, and 5) participation in research" [543]. There is a need for an improvement in approaches that explain research, as well as rules and regulations pertaining to research in the Netherlands. Study information and informed consent procedures are often written in difficult language that is not appropriate for the level of language proficiency of all patients. Informed consent forms for clinical trials generally read more like legal documents than information tailored to caregivers—let alone persons with dementia. For a planned randomized controlled trial (RCT) focused on solution-focused brief therapy, we therefore developed a patient information letter in easy-to-read Dutch (A2), which could also be played in the form of a video dubbed in the patient's language. We also received approval for interpreter-mediated informed consent procedures. As described in chapter 4.3, such approaches may facilitate enrollment of diverse individuals.

5.3.4.2 Improving clinical trials

Although chapter 4.3 focused on disease-modifying trials, efforts to broaden eligibility criteria should not be limited to those trials, but should instead also include other types of research, such as lifestyle interventions and primary prevention trials. In an effort to improve eligibility criteria, investigators from the National Institute on Aging and UsAgainstAlzheimer's are launching a new working group on 'Health Equity and Eligibility Criteria'. This working group

will have a focus on broadening the eligibility criteria. An important target for future research should be the identification of (clinical) screening and outcome measures that are more suitable for diverse populations. The Alzheimer's and Dementia journal family has recently launched an effort towards this goal through their Gaps and Goals project, which aims to review the literature on specific tools (such as the MMSE) with regard to their applicability in diverse populations. More research will likely be needed that examines the reliability of (repeatedly) administering promising existing cross-cultural screeners and test batteries, such as the RUDAS or CNTB, in clinical trial settings.

Changing the eligibility criteria, however, will only address one of a number of issues in research participation. Gilmore-Bykovskyi et al. [544] describe mechanisms of exclusion on several levels. First, there may be individual/interpersonal barriers to participation, which, in addition to eligibility criteria, may include aspects such as logistical and language barriers to participation. Second, they describe barriers at the level of teams and institutions, regarding aspects such as a lack of training/clinical competency, a lack of cultural humility, limited opportunities to remunerate participants, and limited investment in activities to promote trustworthiness. Last, the authors address systems and structural norms—a lack of accountability, funding agencies that are allowed to consistently under-enroll diverse populations, journals lacking standards for reporting subgroup differences, inconsistent reporting (similar to our finding in chapter 4.3), and regulatory standards, and several other factors. Although the situation in the Netherlands cannot be compared directly with the USA, it is likely that many of these issues also occur in the Dutch context.

5.2.4.3 Combining qualitative and quantitative approaches

As previously described in paragraph 5.2, it remains to be seen whether the questionnaires, such as those researchers are required to administer by major funding agencies, are equivalent across diverse populations. For example, although quality of life measures like the EuroQol-5D are often available in many languages, it is not clear whether these questionnaires adequately capture the way diverse (older) individuals experience and value quality of life. Qualitative studies may be needed to examine what quality of life means for diverse individuals; quantitative tools may subsequently be developed. As an example, a qualitative study about perceptions of successful aging showed that first-generation immigrants, like their peers born in the Netherlands, prioritize health and being active and engaged; however, this study also highlights cross-cultural differences, such as in perceptions regarding social networks and filial obligations [545]. Transforming these findings into scales for quantitative research will help push the field forward.

5.4 Conclusion about neuropsychological assessment in the multicultural memory clinic

The projects featured in this dissertation have contributed to fairer neuropsychological assessment for patients with a culturally, linguistically, and educationally diverse background, by demonstrating the pitfalls of existing neuropsychological tests and through the development and study of more suitable tests. Furthermore, my research has led to a number of recommendations to improve neuropsychological assessment and research, particularly the design of clinical trials, on an international scale. Although much more research is needed in this area, I am hopeful that the current momentum in the field of diversity and disparities will continue to propel us forward towards more equitable diagnosis and care in the memory clinic.

CHAPTER 6

SUMMARIES

Samenvatting in gewoon Nederlands (niveau A2)

Samenvatting in formeel Nederlands

English summary

6.1 Samenvatting in gewoon Nederlands (niveau A2)

Ik wil dat iedereen deze tekst kan lezen en begrijpen. Ik gebruik daarom veel gewone woorden. Soms gebruik ik een moeilijk woord. De moeilijke woorden zien er anders uit. Ze staan *niet recht.* Na elk moeilijk woord vertel ik wat het woord betekent.

Ik deed vier jaar een groot onderzoek. Bij *onderzoek* zoekt iemand meer informatie en gegevens. Bijvoorbeeld om problemen op te lossen. In deze tekst vertel ik meer over mijn onderzoek. Mijn onderzoek ging over mensen met een *diverse achtergrond*. Mensen die *divers* zijn, zijn allemaal een beetje anders. Ze spreken bijvoorbeeld een andere taal. Sommige mensen die divers zijn, spreken maar een klein beetje Nederlands. Sommige mensen die divers zijn, zijn niet in Nederland geboren. In mijn onderzoek waren dat bijvoorbeeld mensen uit Marokko, Turkije, Suriname en Kaapverdië.

Oudere diverse mensen in Nederland hebben meer kans om problemen te krijgen met het denken. Ze gaan bijvoorbeeld dingen vergeten. Of ze kunnen niet lang een boek lezen of televisie kijken. Soms kunnen ze hele gewone dingen niet meer doen. Zoals koken of apparaten gebruiken. Sommige van deze mensen hebben *dementie*. Dat is een ziekte van de hersenen. Bij deze ziekte worden sommige delen van de hersenen kleiner. Dit komt omdat er steeds meer heel kleine deeltjes van de hersenen kapot gaan (*hersencellen*).

Er zijn veel soorten dementie. Alzheimer is 1 vorm van dementie. Maar problemen met het vergeten kunnen ook komen door andere redenen. Bijvoorbeeld als een deel van de hersenen plotseling kapot gaat na een groot ongeluk. Dat is geen dementie. Want de problemen bij dementie beginnen meestal heel langzaam. En de problemen worden langzaam erger.

Mensen die last hebben van vergeten gaan soms naar de huisarts. De huisarts stuurt deze mensen naar het ziekenhuis. Wij kijken in het ziekenhuis of iemand problemen heeft met denken. Zoals vergeten of problemen om op woorden te komen. We doen veel testen achter elkaar. Deze testen heten samen het *neuropsychologisch onderzoek*. De *neuropsycholoog* vertelt aan de patiënt hoe het neuropsychologisch onderzoek moet. De neuropsycholoog stelt ook veel vragen. De patiënt moet de testen zelf maken.

De testen in het ziekenhuis zijn gemaakt voor mensen uit Nederland. In de testen zitten plaatjes die diverse mensen soms niet kennen. Zoals een krakeling. We kunnen die testen daarom niet gebruiken bij iedereen. We moesten daarom andere, betere testen vinden. En als die testen er niet waren, moesten we ze zelf maken. Dat heb ik gedaan in mijn onderzoek. Eerst ging ik kijken welke testen er waren in Nederland. Maar ook in de rest van de wereld. Ik keek of die testen goed genoeg waren of niet. Dat is deel 1 van dit boek. Toen gingen wij zelf de goede testen ook in Nederland gebruiken. We maakten zelf nieuwe testen als er geen goede testen waren. Dat is deel 2 van dit boek. Aan het einde hadden we heel veel testen. Ik wist niet of patiënten die wel allemaal wilden en konden doen. Met zoveel testen duurt het neuropsychologisch onderzoek best lang. Patiënten worden er misschien wel heel moe van om zoveel testen te doen. Of ze krijgen er hoofdpijn van. Daar ging ik ook onderzoek naar doen.

In het onderzoek hebben we een paar problemen opgelost. Maar we zijn nog niet klaar. Andere mensen moeten met het onderzoek verder gaan. Daarom wilde ik ook informatie

geven aan andere mensen die onderzoek doen. Zo kunnen zij het onderzoek beter maken. Ook gaf ik advies aan andere mensen die neuropsychologisch onderzoek doen. Ik gaf bijvoorbeeld informatie hoe je op een fijne manier vragen kan stellen. Dan voelen mensen zich fijn in het gesprek. Dat is deel 3 van dit boek.

Wat kwam er uit mijn onderzoek?

Deel 1. Problemen en testen die we hebben
Eerst schreef ik over een Turkse man met de ziekte van *Parkinson*. Mensen met Parkinson kunnen vaak niet meer zo goed bewegen. Sommige mensen met Parkinson gaan ook langzamer denken. De Turkse man moest allemaal testen doen voor mensen uit Nederland. Hij deed deze testen niet goed. Hij was bijvoorbeeld niet lang naar school geweest. Hij kon niet zo goed lezen en schrijven. Omdat de testen niet goed gingen, mocht de man geen operatie. Wij hebben toen andere testen gedaan in ons ziekenhuis. Deze testen waren speciaal voor Turkse mensen. Deze testen kon hij wel goed. De man mocht toch een operatie.

In mijn onderzoek leerde ik ook dat er al veel testen zijn om vergeten te meten. In Europa en de wereld. Maar we hebben nog niet veel testen om te kijken of mensen bijvoorbeeld op woorden kunnen komen. Daarom gingen we nieuwe testen maken.

Deel 2. Oplossingen die ik heb bedacht
De patiënten begrepen meestal wel wat ze moesten doen bij het neuropsychologisch onderzoek. Ook konden ze de testen meestal wel afmaken (hoofdstuk 3.1). De patiënten vonden 1 test heel moeilijk. Vooral mensen die niet goed konden tellen. We moeten deze test daarom alleen gebruiken bij mensen die naar school zijn geweest. Veel patiënten zeiden dat ze moe werden van de testen. Ook waren er patiënten die erg verdrietig waren. Neuropsychologen moeten dus goed kijken of patiënten niet te moe worden van het neuropsychologisch onderzoek. En of patiënten niet te verdrietig zijn op de dag van het onderzoek.

In Nederland was er al een test om vergeten mee te meten. Een test om vergeten mee te meten heet ook een *geheugentest*. In mijn onderzoek gebruikte ik de geheugentest 'Visuele Associatietest'. In die test zitten plaatjes. De plaatjes zijn tekeningen die met een pen getekend waren. Ze hebben geen kleur, maar zijn zwart en wit. We weten dat mensen die niet naar school zijn geweest het soms moeilijk vinden om deze plaatjes te herkennen. Daarom hebben we de test veranderd. Nu gebruiken we foto's met kleur en niet meer zwart met witte tekeningen. Nu kan de test ook gebruikt worden voor diverse mensen (hoofdstuk 3.2).

Ik maakte een hele nieuwe test om te kijken of mensen het moeilijk vinden om op woorden te komen (hoofdstuk 3.3). Mensen zeggen dan vaak "dinges", of "je weet wel". Ze weten het goede woord even niet meer. Dat gebeurt heel vaak bij mensen met Alzheimer. De nieuwe test die ik had gemaakt werkte goed. Mensen met Alzheimer konden vaak niet zeggen wat er op de plaatjes stond. Mensen zonder Alzheimer konden dat wel! Als iemand veel fouten maakt op deze test, kan het dus zijn dat hij Alzheimer heeft.

Ook keken we hoe het ging met *mantelzorgers.* Dat zijn mensen die zorgen voor iemand met dementie. Zoals bijvoorbeeld de man, vrouw of de kinderen. Hoe zwaar is het voor mantelzorgers om te zorgen voor iemand met dementie? Dat wilden we onderzoeken (hoofdstuk 3.4). De mantelzorgers met een diverse achtergrond vonden het best zwaar. Veel van de mensen met dementie vergaten heel veel. Ook moesten mantelzorgers de mensen met dementie bij bijna alles helpen. Misschien vonden ze het daarom zwaar.

Deel 3. Informatie voor andere mensen die onderzoek doen
In het laatste deel van mijn onderzoek gaf ik informatie aan mensen die onderzoek doen en neuropsychologen. Ik schreef hier drie teksten over.

Ik gaf informatie aan neuropsychologen over neuropsychologisch onderzoek bij mensen met een diverse achtergrond (hoofdstuk 4.1). Bijvoorbeeld hoe je goede vragen kan stellen over welke talen iemand spreekt. Of waar iemand thuis problemen mee heeft. Sommige mensen met dementie vinden het bijvoorbeeld moeilijk om te bidden. Ik gaf informatie over hoe het bidden moet. Zo kan de neuropsycholoog goede vragen stellen.

Ik begon samen met neuropsychologen in Europa een nieuwe groep (hoofdstuk 4.2). Deze groep probeert het neuropsychologisch onderzoek bij diverse mensen beter te maken in heel Europa. Wij gaan bijvoorbeeld testen met elkaar delen. Ook gaan we meer samen werken aan nieuwe testen. En we gaan samen neuropsychologen in Europa nieuwe dingen leren en informatie geven.

Als laatste wil ik graag dat meer mensen met een diverse achtergrond mee mogen doen met onderzoek. Bijvoorbeeld onderzoek naar nieuwe medicijnen voor dementie. Nu moeten mensen die mee willen doen heel goed Nederlands spreken. Mensen moeten ook goed kunnen lezen en schrijven. Ze moeten testen doen met moeilijke vragen. Bijvoorbeeld: in welke provincie zijn we nu? Veel mensen die niet uit Nederland komen, weten dat niet. Ze hebben dat niet geleerd op school. Deze regels zijn dus niet eerlijk. Ik bekeek alle onderzoeken die vroeger met medicijnen bij Alzheimer zijn gedaan. Aan die onderzoeken deden bijna geen diverse mensen mee. Dit moet beter worden. In mijn tekst gaf ik veel informatie hoe je onderzoek naar medicijnen kan verbeteren. Zo wordt het onderzoek eerlijk en kan iedereen meedoen.

6.2 Samenvatting in formeel Nederlands

In dit proefschrift geef ik een overzicht van hoe het ervoor stond op het gebied van cross-culturele neuropsychologie in 2017; tevens presenteer ik verschillende nieuwe neuropsychologische tests ter verbetering van de cross-culturele neuropsychologie. Ten laatste voorzie ik het veld van aanbevelingen voor de klinische praktijk en wetenschappelijk onderzoek op zowel nationaal als internationaal niveau.

Hoofdstuk 1 introduceert enkele thema's die relevant zijn voor de cross-culturele neuropsychologie en beschrijft hoe deze zaken zich verhouden tot diverse populaties in Europa en, meer specifiek, in Nederland. De casusbeschrijving van een persoon met Parkinson met een diverse achtergrond in hoofdstuk 2.1 onderstreept het belang van zorgvuldige selectie van testmateriaal en een weloverwogen daaropvolgende rapportage; het gebruik van minder geschikte cognitieve tests in combinatie met suboptimale verslaglegging kan (onbedoeld) resulteren in suboptimale diagnostiek en behandeling. Hoofdstuk 2.2 onderzoekt welke tests er beschikbaar zijn om dementie mee vast te stellen bij laagopgeleide populaties met een culturele en taalkundig diverse achtergrond, waar hoofdstuk 2.3 een overzicht presenteert van neuropsychologische tests die momenteel gebruikt worden in Europese landen voor cross-culturele neuropsychologische diagnostiek. Beide hoofdstukken laten zien dat er veel geheugentests beschikbaar zijn die mogelijk cross-cultureel kunnen worden ingezet, terwijl tests die sociale cognitie, taal—in het bijzonder benoemen—of executieve functies meten schaars zijn. Europese samenwerking is waarschijnlijk nodig voor de ontwikkeling, validering en normering van deze testen, maar ook om andere aspecten van het neuropsychologisch onderzoek te verbeteren, zoals de training van clinici in cross-cultureel neuropsychologisch onderzoek en het gebruik van tolken.

De artikelen in hoofdstuk 2 vormden de basis voor het samenstellen van een neuropsychologische testbatterij, met daarin veelbelovende tests om de cognitieve functies te meten in diverse populaties: de TULIPA batterij. In hoofdstuk 3.1 beschrijf ik de ontwikkeling van deze testbatterij en de haalbaarheid (*feasibility)* van het afnemen van deze batterij in een geheugenpoli populatie. Ik toonde aan dat de individuele tests in deze testbatterij over het algemeen goed uitvoerbaar waren bij mensen met een diverse achtergrond—met uitzondering van de Stippentest van de CCD—maar de volledige batterij is waarschijnlijk te lang voor de meeste patiënten. Neuropsychologen zullen daarom een selectie moeten maken van de meest relevante instrumenten per patiënt. Daarnaast blijken secundaire, complicerende invloeden op het neuropsychologisch onderzoek, zoals depressieve klachten en vermoeidheid, vaak aanwezig te zijn. Dergelijke factoren moeten zodoende goed worden gemonitord tijdens het testen. Hoofdstuk 3.2 en 3.3 richten zich op specifieke testen die deel uitmaken van de TULIPA batterij. Het is bekend uit de wetenschappelijke literatuur dat zwart-witte lijntekeningen—die vaak gebruikt worden in neuropsychologische tests—problemen opleveren wanneer deze benoemd moeten worden door mensen met een laag opleidingsniveau. Daarom richtten beide hoofdstukken zich op het gebruik van kleurenfoto's. Ten eerste pasten we een bestaande geheugentest aan die gebruik maakt van zwart-witte lijntekeningen als stimuli—resulterend in de *modified Visual Association Test,* de mVAT (hoofdstuk 3.2). De vervanging van lijntekeningen door kleurenfoto's resulteerde in een betere prestatie op deze geheugentest dan op het zwart-

witte origineel, zelfs bij gezonde mensen zonder cognitieve klachten. Ik kom zodoende tot de conclusie dat neuropsychologen de geheugencapaciteit van hun (laagopgeleide) patiënten zullen onderschatten als zij tests gebruiken met zwart-witte lijntekeningen als stimuli (hoofdstuk 3.2). Daarnaast ontwikkelden wij een nieuwe benoemtaak met kleurenfoto's—de *Naming Assessment in Multicultural Europe*, ofwel NAME (hoofdstuk 3.3). De psychometrische kwaliteiten van deze test blijken hoopgevend; de NAME is daardoor mogelijk een waardige vervanger voor de (minder geschikte) Boston benoemtaak in deze populatie.

De diagnostiek in de geheugenpoli richt zich echter niet alleen op de patiënt zelf; in dit proefschrift richt ik me daarom ook op mantelzorgers. Ik onderzocht de mate van (over)belasting van mantelzorgers met een diverse achtergrond die onze multiculturele geheugenpoli bezochten (hoofdstuk 3.4) en vergeleek deze ervaren belasting met die van mantelzorgers van in Nederland geboren ouderen uit het Alzheimercentrum. Deze studie toont aan dat de ervaren mantelzorgbelasting hoog was. De ervaren mantelzorgbelasting was geassocieerd met zowel objectieve als door de informant gerapporteerde cognitieve problemen. Deze studie kan dienen als fundering voor vervolgonderzoek dat zich richt op psychosociale interventies door te onderzoeken wie er het meeste risico hebben om een ernstige mantelzorgbelasting te ervaren.

In het laatste deel van dit proefschrift werk ik toe naar verbeteringen in het onderzoek en de klinische praktijk van de toekomst. Ten eerste geef ik advies aan neuropsychologen hoe zij neuropsychologisch onderzoek kunnen verrichten dat meer sensitief is voor diversiteitsaspecten (hoofdstuk 4.1). Ten tweede presenteer ik de standpunten en korte- en lange termijn doelen van het Europees Consortium voor Cross-Culturele Neuropsychologie (ECCroN; hoofdstuk 4.2). Samen met Europese collega's richtte ik dit consortium op met als doel om problemen in cross-cultureel neuropsychologisch onderzoek in Europa (zoals beschreven in hoofdstuk 2.3) aan te pakken. In het kort zijn de standpunten van dit consortium als volgt: 1) ECCroN raadt aan om een breed scala aan diversiteit-gerelateerde variabelen in ogenschouw te nemen in het onderzoek en de klinische praktijk; 2) ECCroN stelt zich als doel samen te werken aan de ontwikkeling, validatie en implementatie van breed toepasbare neuropsychologische tests; 3) ECCroN maakt zich hard voor een verbetering in de opleiding van neuropsychologen en voor verbeteringen in neuropsychologisch onderzoek met een tolk.

Neuropsychologisch onderzoek wordt niet alleen gebruikt in de klinische praktijk, maar speelt ook een belangrijke rol in wetenschappelijk onderzoek, in het bijzonder medicatieonderzoek. De afgelopen jaren informeerden meerdere mantelzorgers met een diverse achtergrond bij ons naar mogelijkheden voor hun naaste met dementie om deel te nemen aan ziekte-beïnvloedend medicatieonderzoek bij dementie. Vanwege de in- en exclusiecriteria die gebruikt worden in medicatieonderzoek—in het bijzonder de prestatie op de MMSE, maar ook andere aspecten zoals vereisten aan taal of opleidingsniveau—werden deze patiënten vrijwel zonder uitzondering uitgesloten van deelname aan medicatieonderzoek. Deze problematische in- en exclusiecriteria leidden ons ertoe dit aspect verder te onderzoeken. We verrichtten een systematische analyse van de criteria die gebruikt werden in medicatieonderzoek bij de ziekte van Alzheimer en geven vervolgens aanbevelingen hoe deze criteria verbreed en verbeterd kunnen worden

(hoofdstuk 4.3). We onderzochten ook de mate van diversiteit in deze klinische trials. Eén van onze belangrijkste bevindingen is dat in medicatieonderzoek (inderdaad) een groot aandeel van de deelnemers 'wit' was (~95%). Dit percentage toont geen verandering met het verstrijken van de jaren. Inclusiecriteria die vaak worden gebruikt zijn psychiatrische aandoeningen, cardio- en cerebrovasculaire aandoeningen, vereisten wat betreft de aanwezigheid van de mantelzorger en de scores op de MMSE; het gebruik van deze criteria leidt hoogstwaarschijnlijk tot een disproportionele exclusie van mensen met een diverse achtergrond in deze medicatieonderzoeken—temeer omdat veel van de gebruikte criteria vaak slecht gedefinieerd waren.

6.3 English summary

In this dissertation, I provide an overview of the state of the field in 2017, present new neuropsychological tests developed to improve cross-cultural neuropsychology, and provide recommendations for clinical practice and research.

Chapter 1 introduces issues relevant to cross-cultural neuropsychology and describes how these issues relate to diverse populations in Europe and the Netherlands specifically. The case study of a diverse Parkinson's disease patient presented in chapter 2.1 underlines the importance of careful test selection and subsequent reporting; the use of less appropriate tests in cognitive testing in combination with suboptimal reporting may inadvertently result in substandard diagnosis and treatment of cognitive impairment. Chapter 2.2 investigates which tests are available to diagnose dementia in low educated, culturally, and linguistically diverse individuals, while chapter 2.3 contributes a list of tests that are currently used in European countries for cross-cultural neuropsychological assessment. Both chapters show that memory tests with cross-cultural potential are more widely available, whereas tests measuring aspects of social cognition, language—particularly naming—and executive functioning are scarce. European collaborations may be needed to develop and validate new tests and collect normative data; furthermore, collaboration may also be needed to improve other aspects of cross-cultural neuropsychological assessment, such as training of clinicians and the use of interpreters.

The results from the studies presented in chapter 2 contributed to the compilation of a neuropsychological test battery consisting of the tests that showed promise in measuring cognitive functioning in diverse individuals: the TULIPA test battery. In chapter 3.1, we describe the development of this test battery and examine its feasibility in a memory clinic population. In brief, we find that the individual tests in this test battery were generally feasible in diverse individuals—with the exception of the Dots Test of the CCD—but the full battery is likely too long for most patients. Neuropsychologists should therefore make a selection of the most relevant instruments for each patient. In addition, factors complicating the neuropsychological assessment, such as depressive symptoms and fatigue, are often present and should be carefully monitored during testing. Chapters 3.2 and 3.3 describe specific tests that are part of the TULIPA battery. As the scientific literature indicates that low educated diverse individuals may have difficulty naming the black-and-white line drawings used in many neuropsychological tests, both chapters focus on the use of colored photographs. First, we modified an existing memory test that originally used black-and-white line drawings as stimuli—resulting in the modified Visual Association test, or mVAT (chapter 3.2). By substituting the line drawings with colored photographs, we see a significantly better performance on the mVAT than its black-and-white counterpart, even in healthy controls. I conclude that memory capacities of low educated diverse individuals are underestimated if they are assessed with tests using black-and-white line drawings as stimuli (chapter 3.2). Second, we developed a new naming test with colored photographs—the Naming Assessment in Multicultural Europe, or NAME (chapter 3.3). The NAME displays promising psychometric characteristics and may be a worthy substitute for the (less suitable) Boston Naming Test.

The diagnostic trajectory in the memory clinic does not revolve solely around the patient, however; in this dissertation, I therefore also focus on caregivers. I examined caregiver burden levels in diverse caregivers visiting our multicultural memory clinic (chapter 3.4) and compared these burden levels with caregivers of native Dutch patients from the Alzheimer Center. This study reveals that caregiver strain levels in these groups are high and are associated with both proxy-rated and objective measures of cognition. This work may pave the way for studies aimed at psychosocial interventions by identifying those most at risk of experiencing severe caregiver burden.

In the last part of this dissertation, I aim to take steps to improve future research and clinical practice. First, I provide guidelines for neuropsychologists on how to use a more diversity-sensitive approach in neuropsychological assessment (chapter 4.1). In addition, I present the standpoints and short- and long-term goals of the European Consortium on Cross-Cultural Neuropsychology (chapter 4.2), which I co-founded to address the issues in cross-cultural neuropsychological assessment in Europe (as described in chapter 2.3). In brief, the consortium's standpoints are as follows: 1) ECCroN recommends taking a broad range of diversity-related variables into account in research and clinical practice; 2) ECCroN aims to collaborate on the development, validation, and implementation of widely applicable neuropsychological tests; 3) ECCroN advocates for an improvement in the clinical training of neuropsychologists and improvements in interpreter-mediated assessment.

Last, neuropsychological assessment is not only used in clinical assessment, but also plays an important role in research, in particular in drug trials. Over the years, diverse caregivers visiting our multicultural memory clinic sometimes inquired about opportunities for their loved ones with dementia to participate in disease-modifying dementia clinical trials. However, these patients are invariably ineligible for participation due to the entry requirements of most trials—in particular the MMSE score, but also aspects such as language and education requirements. Instead of focusing on enrollment issues that were widely examined in the international literature, we therefore systematically reviewed the eligibility criteria used in clinical trials and provide recommendations on how to broaden these criteria (Chapter 4.3). We also investigate the level of participant diversity in these clinical trials. We find that, across the included trials, participants are predominantly white (~95%) and this percentage shows no increase or decrease over time. Criteria that are often used in these trials were related to psychiatric illness, cardiovascular and cerebrovascular disease, caregiver attendance, and MMSE scores, the use of which likely results in a disproportionate exclusion of diverse individuals—particularly since many eligibility criteria were not well-defined.

CHAPTER 7

APPENDICES

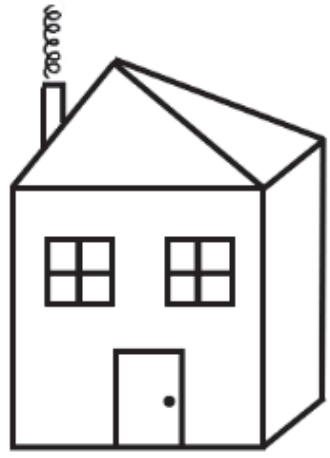
Acknowledgments/dankwoord

Curriculum vitae

List of publications

PhD portfolio

List of abbreviations



7.2 Curriculum vitae

Sanne Franzen was born on the fourth of December, 1990 in Bilthoven. In 2009, she graduated with honors ('cum laude') from secondary school. She then started her studies in (Neuro)psychology at Utrecht University. After obtaining her BSc degree in Psychology in 2012 ('cum laude'), she worked as a volunteer in a Tibetan-Buddhist monastery in Nepal for several months, after which she started her MSc degree in Neuropsychology at Utrecht University. She completed her thesis about the relationship between metabolic syndrome and global brain volumes in patients with schizophrenia and (schizo)affective disorders at the department of Psychiatry of the University Medical Center in Utrecht. In 2014, she started her clinical internship in Neuropsychology at the department of Neurology of the Erasmus Medical Center, and she received her Master's degree in September of 2014 ('cum laude'). She then stayed on in the Erasmus MC, working as a neuropsychologist, as well as in the position of research assistant on the FTD-RisC study and as a rater in dementia clinical trials. When the multicultural memory clinic opened in late 2015, ms. Franzen was assigned as its coordinator. In 2017, Janne Papma and Sanne Franzen obtained a ZonMw Memorabel grant, allowing ms. Franzen to formally start her PhD in the multicultural memory clinic, working closely with her supervisors Dr. Esther van den Berg and Dr. Janne Papma. While working on her research, she continued to see patients for neuropsychological assessment in the outpatient clinic. After completing her PhD, ms. Franzen has entered a clinical training program to become a licensed health care psychologist ('GZ-opleiding'), combining this clinical position with her postdoctoral research.

7.3 List of publications

The multicultural memory clinic

Uysal-Bozkir, Ö., **Franzen, S.**, & Goudsmit, M. (2022). Neuropsychological Assessment of Moroccan Patients in the Netherlands. In F. Irani (Ed.), *Cultural Diversity in Neuropsychological Assessment: Developing Understanding through Global Case Studies*. Routledge.

Nielsen, T.R., **Franzen, S.**, Goudsmit, M., & Uysal-Bozkir, Ö. (2022). Cross-cultural Neuropsychological Assessment in the European Context: Embracing Maximum Diversity at Minimal Geographic Distances. In F. Irani (Ed.), *Cultural Diversity in Neuropsychological Assessment: Developing Understanding through Global Case Studies*. Routledge.

Franzen, S., van den Berg, E., Ayhan, Y., Satoer, D.D., Türkoğlu, Ö., Genç Akpulat, 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 Multicultural Memory Clinic. Advance online publication. https://doi.org/10.1017/S135561772100148X

Franzen, S., Eikelboom, W.S., van den Berg, E., Jiskoot, L.C., van Hemmen, J., & Papma, J.M. (2021). Caregiver Burden in a Culturally Diverse Memory Clinic Population: The Caregiver Strain Index-Expanded. *Dementia and Geriatric Cognitive Disorders.* Advance online publication. https://doi.org/10.1159/000519617

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.* Advance online publication. https://doi.org/10.1080/13854046.2021.1981456

Franzen, S., Smith, J.E., van den Berg, E., Rivera Mindt, M., van Bruchem-Visser, R.L., Abner, E., Schneider, L.S., Prins, N.D., Babulal, G.M., & Papma, J.M. (2021). Diversity in Alzheimer's Disease Drug Trials: The Importance of Eligibility Criteria. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association.* Advance online publication. https://doi.org/10.1002/alz.12433

Franzen, S., van den Berg, E., & Papma, J.M. (2021). Cross-culturele Neuropsychologische Diagnostiek bij Volwassenen en Ouderen op de Geheugenpoli [Cross-cultural Neuropsychological Assessment of Adult and Elderly Memory Clinic Patients]. *Tijdschrift voor Neuropsychologie, 16(1),* 14–28.

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),* 815–839. https://doi.org/10.1093/arclin/acaa083

Goudsmit, M., van Campen, J.P.C.M., **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.

Franzen, S., van den Berg, E., Goudsmit, M., Jurgens, C.K., van de Wiel, L., Kalkisim, Y., Uysal-Bozkir, Ö., 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 2019, 47(4–6),* 345–354. https://doi.org/10.1159/000501151

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 Very Low Educated Multicultural Memory Clinic Population. *Dementia and Geriatric Cognitive Disorders Extra, 8(2),* 290–305.

Franzen, S., van den Berg. E. & Papma, J.M. (2018). Neuropsychologische Diagnostiek bij niet-Westerse Migranten [Neuropsychological Assessment of non-Western Immigrants]. *Tijdschrift voor Neuropsychologie, 13(1),* 59–67.

The Alzheimer Center

Jiskoot, L.C., Poos, J.M., Vollebergh, M.E., **Franzen, S.**, van Hemmen, J., Papma, J.M., van Swieten, J.C., Kessels, P.C., & van den Berg, E. (2021) Emotion Recognition of Morphed Facial Expressions in Presymptomatic and Symptomatic Frontotemporal Dementia, and Alzheimer's Dementia. *Journal of Neurology,268(1),* 102–113. https://doi.org/10.1007/s00415-020-10096-y.

Poos, J.M., van den Berg, E., Visch-Brink, E., Eikelboom, W.S., **Franzen, S.**, van Hemmen, J., Pijnenburg, Y., Satoer, D., Dopper, E., van Swieten, J.C., Papma, J.M., Seelaar, H., Jiskoot, L.C. (2021) Exploring Abstract Semantic Associations in the Frontotemporal Dementia Spectrum in a Dutch Population. *Archives of Clinical Neuropsychology.* https://doi.org/10.1093/arclin/acab022.

van den Berg, E. Poos, J.M., Jiskoot, L.C., Montagne, B., Kessels, R.P.C., **Franzen, S.**, van Hemmen, J., Eikelboom, W.S., Heijboer, E., de Kriek, J., van der Vlist, A., van Swieten, J.C., Seelaar, H., Papma, J.M. (2021) Impaired Knowledge of Social Norms in Dementia and Psychiatric Disorders: Validation of the Social Norms Questionnaire-Dutch Version (SNQ-NL). *Assessment.* Advance online publication. https://doi.org/10.1177/10731911211008234

van den Berg, E., Poos, J.M., Jiskoot, L.C., Heijnen, L.M., **Franzen, S.,** Steketee, R.M.E., Meijboom, R., de Jong, F.J., Seelaar, H., van Swieten, J.C., & Papma, J.M. (2020) Differences in Discriminability and Response Bias on RAVLT Delayed Recognition in Behavioral Variant Frontotemporal Dementia and Alzheimer's Disease. *JINS: Journal of the International Neuropsychological Society,* https://doi.org/10.1017/S1355617720000375

van den Berg, E., **Franzen, S.,** van Hemmen, J., Poos, J.M., & Jiskoot, L.C. (2020) Opinie: Diagnostische Dilemma's in de Neuropsychologische Diagnostiek bij Gedragsvariant Frontotemporale Dementie. [Opinion Piece: Diagnostic Dilemmas in the Neuropsychological Assessment of Behavioral Variant Frontotemporal Dementia] *Tijdschrift voor Neuropsychologie, 15(1),* 1–18.

Jiskoot, L.C., Panman, J.L., Meeter, H.H., Dopper, E.G.P., Donker Kaat, L., **Franzen, S.,** van der Ende, E.L., van Minkelen, R., Rombouts, S.A.R.B., Papma, J.M., & van Swieten, J.C. (2019) Longitudinal Multimodal MRI as a Prognostic and Diagnostic Biomarker in Presymptomatic Familial Frontotemporal Dementia. Brain, 142(1), 193–208.

Jiskoot, L.C., Panman, J.L., van Asseldonk, L., **Franzen, S.**, Meeter, H.H., Donker Kaat, L., van der Ende, E.L., Dopper, E.G.P., Rimman, R., van Minkelen, R., van Swieten, J.C., van den Berg, E., & Papma, J.M. (2018) Longitudinal Cognitive Biomarkers Predicting Symptom Onset in Presymptomatic Frontotemporal Dementia. Journal of Neurology, 265(6), 1381–1392.

Meeter, H.H., Gendron, T.F., Sias, A.C., Jiskoot, L.C., Russo, S.P., Donker Kaat, L., Papma, J.M., Panman, J.L., van der Ende, E.L., Dopper, E.G., **Franzen, S.,** Graff, C., Boxer, A.L., Rosen, H.J., Sanchez-Valle, R., Galimberti, D., Pijnenburg, Y.A.L., Benussi, L., Ghidoni, R., Borroni, B., Laforce, R., del Campo, M., Teunissen, C.E., van Minkelen, R., Rojas, J.C., Coppola, G., Geschwind, D.H., Rademakers, R., Karydas, A.M., Öijerstadt, L., Scarpini, E., Binetti, G., Padovani, A., Cash, D.M., Dick, K.M., Bocchetta, M., Miller, B.L., Rohrer, J.D., Petrucelli, L., van Swieten, J.C., & Lee, S.E. (2018) Poly(GP), Neurofilament and Grey Matter Deficits in C9orf72 Expansion Carriers. Annals of Clinical and Translational Neurology, 5(5), 583–597.

7.4 PhD portfolio

Conferences

International conferences	Location	Type	ECTS
AAIC (2017)	London	Attendance	1.5
FESN (2017)	Maastricht	Oral presentation	1
International Congress on Multidisciplinary Approach to Elderly Health and Care (2018)	Ankara	Invited speaker	1.5
AAIC (2018)	Chicago	Poster presentation (2x)	1.5
'Avondcongres Dementiediagnostiek bij ouderen met een migratie-achtergrond' (2018)	Brussels	Invited speaker	0.5
INS Annual Meeting (2019)	New York	Poster presentation	1
Symptom Validity Assessment conference (2019)	Amsterdam	Attendance	1
AAIC (2019)	Los Angeles	Oral presentation	1.5
FESN (2019)	Milan	Oral presentation	1
AAIC (2020)	Virtual event	Poster & oral presentation	1.5
INS World Conference (2020)	Virtual event	Attendance	1
Alzheimer Europe Conference (2020)	Virtual event	Oral presentation	1
INS Annual Meeting (2021)	Virtual event	Attendance	1
Alzheimer's Association Health Disparities Conference (2021)	Virtual event	Attendance	1
AAIC (2021)	Hybrid event	Poster & oral presentation	1.5
Nordic Meeting in Neuropsychology (2021)	Copenhagen	Invited speaker	1.5
Vlaams Forum voor Diagnostiek (2021)	Virtual event	Invited speaker	0.5
INS World Conference (2022)	Hybrid event	Invited panelist	0.5
National conferences/meetings			
NVN conference (2019, 2021)	Amsterdam	Attendance	0.5
NIP symposium neuropsychological guidelines for dementia diagnosis (2019)	Utrecht	Invited speaker	0.5
Amsterdam Solution Focused Community (2018, 2020)	Amsterdam	Attendance	0.3
Symposium 'Onderzoek bij mensen in kwetsbare situaties' (2019)	Utrecht	Attendance	0.3
NGN conference (2021)	Amsterdam	Invited plenary speaker	0.5

Training

Course	Institute, location	ECTS
'A culture-sensitive approach using the Cultural Interview and Cultural Formulation Interview' (2017)	RINO, Amsterdam	0.3
'Delivering patient education about depression to individuals with limited literacy' (2018)	Pharos, Utrecht	0.15
Preventing Failed Interventions in Behavioral Research (2018)	NIHES, Rotterdam	1.4
'Solution-focused treatment' (2018)	RINO, Amsterdam	2.35
'Diversity, Multiculturalism, and professionalism' (2018)	Nuance door Training en Advies (NTA), Rotterdam	0.15
'The Arabic and Eritrean cultures' (2019)	TVcN, Utrecht	0.15
Biomedical English Writing (2019–2020)	MolMed, Rotterdam	2.0
CPO-course patient oriented research (2019)	Erasmus, Rotterdam	0.3
Contextually Valid Executive functioning Assessment (2019)	INS, New York	0.1
Caregiver burden from an empirical perspective (2019)	INS, New York	0.1
BROK-recertification (2020)	NFU, Rotterdam	1.5
Using R for Statistics in Medical Research (2021)	NIHES, Virtual	1.4
Research integrity (2021)	Erasmus MC	0.3

Teaching

Lectures, presentations, and workshops

- Invited presentation 'Neuropsychological assessment of diverse populations in Europe', Cognitive Neuroscience Seminar, Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University USA (2022, 0.3 ECTS).
- Invited speaker 'Neuropsychological Assessment of Diverse Older Populations in Europe', scientific session Cognition PIA (2021, 0.3 ECTS).
- Invited speaker 'Diversity in Alzheimer's disease drug trials: The importance of eligibility criteria', the Health Equity and Eligibility Criteria in Alzheimer's Disease and Related Dementias Research Workshop (National Institute on Aging & UsAgainstAlzheimer's, 2021, 0.3 ECTS).
- Co-organizer and moderator of the webinar 'Towards more inclusive Alzheimer's research and communication: lessons learned from translational research in diverse populations in Australia' (2021, 0.5 ECTS).
- Invited lecture Diversity in Neuropsychology (2021; Bachelor Psychology UU, 0.5 ECTS).
- Lecture Cross-Cultural Neuropsychology (2017; Master Neuroscience and Cognition, UvA, 0.3 ECTS).
- Annual lecture Dementia (2018–2021; Bachelor Psychology EUR, 1 ECTS).
- Quarterly lectures 'Dementia and Future Directions' for general practitioners in training (2018–2020, Erasmus MC, 1 ECTS).
- Annual lecture 'Dementia diagnosis in the multicultural memory clinic' for geriatricians and geriatricians in training (2020–2021, Department of Geriatric Medicine Erasmus MC, 0.5 ECTS).
- Presentations about cross-cultural neuropsychological assessment for neurologists, geriatricians, neuropsychologists, dementia 'casemanagers', and general practitioners (Fonds Achterstandwijken, HAGRO Rotterdam). Total of 11 presentations (2 ECTS).
- Presentations about solution-focused treatment for neurologists and rehabilitation neuropsychologists (2018, 2 presentations, 0.5 ECTS).
- Workshops for a mixed audience (lay and professional) about cross-cultural dementia diagnosis and (solution-focused) dementia care (Alzheimer Netherlands volunteer day, Science meets city event (2x), Deltaplan dementie (2x), Mix-and-Match meeting; 2017–2021, 1.5 ECTS).
- Lay education sessions about dementia & end-of-life care as part of project 'Gezond Ouder Worden', including design of educational materials (6 sessions, 2017–2019, 2 ECTS).

Supervision

- Supervision of research interns (2017–2021, 13.5 ECTS): MSc theses (five students), applied university bachelor theses (two students), BSc thesis (one student), voluntary research interns (three students).
- Supervision of 30 clinical interns (2017–2021), as part of position as neuropsychologist (see Other).

Chair positions, board memberships & professional affiliations

- Co-founder and chair of regional network 'Culturele Dementiezorg Rotterdam (quarterly meetings between 2017–2021, 3 ECTS).
- Co-founder and (informal) chair of the European Consortium on Cross-Cultural Neuropsychology (2019-present, 2 ECTS).
- Elected executive committee member of the Diversity and Disparities PIA of ISTAART, including monthly executive committee meetings, organization of scientific session and business meeting, and organization of student and postdoc networking event (2020–2022, 4 ECTS).
- Global ambassador of the Diversity and Disparities PIA (2019–2021).
- Co-chair of Diversity & Frontotemporal Dementia workgroup of ISTAART (2021, 1 ECTS).
- Member of the Diversity and Disparities Special Interest Groups 'Sex/Gender', 'LGBTQIA+', 'Bilingualism, Literacy, and Language' and LMIC workgroup (attending SIG calls 2019-present, 1 ECTS).
- Member of the Cultural Special Interest Group of the International Neuropsychological Society (INS).
- Co-lead workgroup 'Developing a Common Language and Glossary of Terms for Cultural/Cross-Cultural Neuropsychology' of the Cultural Special Interest Group INS (2021, 0.5 ECTS).
- Executive commission member for the Central Commission on Cultural Diversity and Psychology of the Dutch Association for Psychologists (2019–2021, 2 ECTS).
- Advisory committee member of 'Zorgstandaard dementie' [national care standards for dementia] (2019, 0.5 ECTS).
- Advisory panel member 'Taking Care of Caregivers' (4 meetings, 2018–2020, 0.5 ECTS).

Grants and awards

- Co-PI "Intercultural dementia diagnostic and care in the memory clinic" (ZonMw, 2017, €418.131).
- Co-applicant "Cross-cultural neuropsychological assessment of social cognition" (Université de Paris, 2020, 1 PhD student, ~€150.000).
- Principal investigator community program about healthy aging "Gezond Ouder Worden" (Gemeente Rotterdam, 2019, €39.751).
- Co-applicant community program about healthy aging "Gezond Ouder Worden" (Gemeente Rotterdam, 2017, €16.155).
- Co-investigator "Development of the Moving GENIE online to support Culturally and Linguistically Diverse family carers of persons living with dementia" (Australian Association of Gerontology, 2021, $29.771 [AUS]).
- Research and Education grant Erasmus Trustfonds "Promoting equality in health care: dementia diagnostics and care" in collaboration with the department of Neurology and Neuropsychology, Université de Rabat Mohammed V Medical Center (postponed due to COVID-19 pandemic, €4.500).
- Several travel grants (€3.375).
- First runner-up Avicenna award (Nederlandse Vereniging voor Psychiatrie – Afdeling Transculturele Psychiatrie, 2020).

Other activities

- Neuropsychologist 0.22 FTE (2017–2021, 60 ECTS)
- Coordinator multicultural memory clinic (2017–2021, 22 ECTS)
- Weekly research meetings Alzheimer Center (2017–2021, 4 ECTS)
- Weekly multidisciplinary meeting (2017–2021, 4 ECTS)
- Half-yearly regional meeting in neuropsychology (2017–2021, 0.5 ECTS)
- Rater BioGen EMERGE trial (1.5 ECTS)

Total ECTS

Activity	ECTS
Conferences	22.1
Courses	10.2
Teaching	24.2
Chairs, board memberships, professional affiliations	14.5
Other activities	92
Total	**163**

7.5 List of abbreviations

3MS	Modified Mini-Mental State Examination
AA	Alzheimer's Association
Aβ	Amyloid beta
AD	Alzheimer's disease
ADAS-Cog	Alzheimer's Disease Assessment Scale-Cognitive Subscale
ADL	Activities of daily living
ALT	Alanine transaminase
ANOVA	Analysis of variance
ANZCTR	Australian New Zealand Clinical Trial Registry
AST	Aspartate transaminase
AUC	Area under the curve
BADS	Behavioural Assessment of the Dysexecutive Syndrome
BCSB	Brief Cognitive Screening Battery
BDI	Beck Depression Inventory
BENCI	Computerized Battery for Neuropsychological Evaluation of Children
BMI	Body mass index
BNT	Boston Naming Test
CCD	Cross-Cultural Dementia Screening
CC-SIT	Cross-Cultural Smell Identification Test
CDR	Clinical Dementia Rating
CDT	Clock Drawing Test
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
CIND	Cognitive impairment – no dementia
CLNT	Cross-Linguistic Naming Test
CNS	Central nervous system
CNTB	European Cross-Cultural Neuropsychological Test Battery
CONSORT	Consolidated Standards of Reporting Trials
COWAT	Controlled Oral Word Association Test
CP	Constructional Praxis
CSF	Cerebrospinal fluid
CSI	Caregiver Strain Index
C-SSRS	Columbia Suicide Severity Rating Scale
CV	Curriculum vitae
CVF	Category Verbal Fluency
CVVLT	Chinese Version Verbal Learning Test
DBS	Deep brain stimulation
DLB	Dementia with Lewy Bodies
DS(B/F)	Digit Span (Backward/Forward)
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSMB	Data Safety Monitoring Board
DSST	Digit Symbol Substitution Test
ECCroN	European Consortium on Cross-Cultural Neuropsychology
EEA	European Economic Area
EFPA	European Federation of Psychologists' Associations
EU	European Union

EudraCT	European Union Drug Regulating Authorities Clinical Trial Database
FAB	Frontal Assessment Battery
FAQ	Functional Activities Questionnaire
FCSRT	Free and Cued Selective Reminding Test
FDA	Food and Drug Administration
(FDG-)PET	(Fluorodeoxyglucose-)positron emission tomography
FDR	False discovery rate
FDT	Five Digit Test
FESN	Federation of European Societies in Neuropsychology
FINGER	The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability
FOME	Fuld Object Memory Evaluation
FTD	Frontotemporal dementia
GAM	Generalized additive model
GDS	Geriatric Depression Scale
HIV	Human immunodeficiency virus
iADL	Instrumental activities of daily living
ICD	International Classification of Diseases and Related Health Problems
IQ	Intelligence quotient
IQCODE	Informant Questionnaire on Cognitive Decline
IQR	Interquartile range
ISLT	International Shopping List Test
ISTAART	International Society to Advance Alzheimer's Research & Treatment
KSRT	Korean Story Recall Test
L1	First language
L2	Second language
LICA	Literacy Independent Cognitive Assessment
LILACS	Latin American and Caribbean Health Sciences Literature
LM	Logical Memory (see WMS)
MCI	Mild cognitive impairment
MINT	Multilingual Naming Test
MMSE	Mini-Mental State Examination
MMSE-I	Mini-Mental State Examination for illiterate individuals
MoCA	Montreal Cognitive Assessment
MRI	Magnetic resonance imaging
mVAT	Modified Visual Association Test
NAME	Naming Assessment in Multicultural Europe
NCT	National Clinical Trial
NIA	National Institute on Aging
NIH	National Institutes of Health
NLCA	Non-Language based Cognitive Assessment
NVN	Nederlandse Vereniging voor Neuropsychologie
OMB	Office of Management and Budget
PD(D)	Parkinson's disease (dementia)
PIA	Professional Interest Area (ISTAART)
PMIS	Picture based Memory Impairment Screen
PREA	Pediatric Research Equity Act

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PWH	Person living with human immunodeficiency virus (HIV)
Q1	First quartile
Q3	Third quartile
RAVLT	Rey Auditory Verbal Learning Test
RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
RCT	Randomized controlled trials
RPT	Recall of Pictures Test
RUDAS	Rowland Universal Dementia Assessment Scale
SASH	Short Acculturation Scale for Hispanics
SCI	Subjective cognitive impairment
SD	Standard deviation
SDT	Stick Design Test
SE	Standard error
SES	Socioeconomic status
SMC	Subjective memory complaints
SVLT	Seoul Verbal Learning Test
TMA-93	Memory Associative Test of the district of Seine-Saint-Denis-93
TMT	Trail Making Test
TN-LIN	The Neuropsychological Investigations Laboratory Naming Test
TNI-93	Test des Neuf Images du 93
TNT	Texas Spanish Naming Test
TOMM	Test of memory malingering
TULIPA	Towards a Universal Language: Intervention & Psychodiagnostic Assessment
UCLA	University of California Los Angeles
UK	United Kingdom
ULN	Upper limit of normal
UNESCO	United Nations Educational, Scientific, and Cultural Organization
USA	United States of America
VaD	Vascular dementia
VAT	Visual Association Test
VR	Virtual Reality
VR	Visual Reproduction (see WMS)
WAIS(-R)	Wechsler Adult Intelligence Scale(-Revised)
WISC	Wechsler Intelligence Scale for Children
WHO	World Health Organization
WMS(-R)	Wechsler Memory Scale(-Revised)

References

1. McKenzie, K. J., & Crowcroft, N. S. (1994). Race, Ethnicity, Culture, and Science. *The British Medical Journal, 309*(6950), 286-287. https://doi.org/10.1136/bmj.309.6950.286
2. Bhopal, R. (2004). Glossary of Terms Relating to Ethnicity and Race: For Reflection and Debate. *Journal of Epidemiology & Community Health, 58*(6), 441-445.
3. Flanagin, A., Frey, T., Christiansen, S. L., & A. M. A. Manual of Style Committee. (2021). Updated Guidance on the Reporting of Race and Ethnicity in Medical and Science Journals. *JAMA: Journal of the American Medical Association, 326*(7), 621-627. https://doi.org/10.1001/jama.2021.13304
4. Peeren, E. (2018). Taal Valt Niet 'Op Te Ruimen'. In W. Modest & R. Lelijveld (Eds.), *Woorden Doen Ertoe: Een Incomplete Gids Voor Woordkeuze Binnen De Culturele Sector* (pp. 43-46). https://www.tropenmuseum.nl/sites/default/files/2018-06/WordsMatter_Nederlands.PDF
5. Ardila, A. (2020). Cross-Cultural Neuropsychology: History and Prospects. *RUDN Journal of Psychology and Pedagogics, 17*(1), 64-78. 10.22363/2313-1683-2020-17-1-64-78
6. Richardson, J. T. (2003). Howard Andrew Knox and the Origins of Performance Testing on Ellis Island, 1912-1916. *History of Psychology, 6*(2), 143-170.
7. Kotik-Friedgut, B. (2006). Development of the Lurian Approach: A Cultural Neurolinguistic Perspective. *Neuropsychology Review, 16*(1), 43-52. https://doi.org/10.1007/s11065-006-9003-9
8. Arunkumar, M., van Paridon, J., Ostarek, M., & Huettig, F. (2021). Do Illiterates Have Illusions? A Conceptual (Non)Replication of Luria (1976). *Journal of Cultural Cognitive Science, 5*(2), 143-158. https://doi.org/10.1007/s41809-021-00080-x
9. Cattell, R. (1949). *Culture Free Intelligence Test, Scale 1, Handbook.* Institute of Personality and Ability.
10. Vernon, P. E. (1969). *Intelligence and Cultural Environment*. Methuen.
11. Irvine, S. H., & Berry, J. W. (Eds.). (1988). *Human Abilities in Cultural Context*. Cambridge University Press.
12. Ardila, A. (2005). Cultural Values Underlying Psychometric Cognitive Testing. *Neuropsychology Review, 15*(4), 185-195. https://doi.org/10.1007/s11065-005-9180-y
13. 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
14. Greenfield, P. (1997). You Can't Take It with You: Why Ability Assessments Don't Cross Cultures. *American Psychologist, 52*(10), 1115–1124. https://doi.org/10.1037/0003-066X.52.10.1115
15. Sue, D. W., & Sue, D. (1990). *Counseling the Culturally Different: Theory and Practice* (2nd ed.). John Wiley & Sons.
16. Judd, T., Capetillo, D., Carrion-Baralt, J., Marmol, L. M., San Miguel-Montes, L., Navarrete, M. G., . . . Valdes, J. (2009). Professional Considerations for Improving the Neuropsychological Evaluation of Hispanics: A National Academy of Neuropsychology Education Paper. *Archives of Clinical Neuropsychology, 24*(2), 127-135. https://doi.org/10.1093/arclin/acp016

17. Ardila, A. (2007). Toward the Development of a Cross-Linguistic Naming Test. *Archives of Clinical Neuropsychology, 22*(3), 297-307. https://doi.org/10.1016/j.acn.2007.01.016
18. Barker-Collo, S. (2007). Boston Naming Test Performance of Older New Zealand Adults. *Aphasiology, 21*(12), 1171-1180. https://doi.org/10.1080/02687030600821600
19. 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 & Geriatric Cognitive Disorders Extra, 4*(1), 86-94. https://doi.org/10.1159/000360695
20. 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
21. Al-Jawahiri, F., & Nielsen, T. R. (2021). Effects of Acculturation on the Cross-Cultural Neuropsychological Test Battery (CNTB) in a Culturally and Linguistically Diverse Population in Denmark. *Archives of Clinical Neuropsychology, 36*(3), 381-393. https://doi.org/10.1093/arclin/acz083
22. Glymour, M. M., & Manly, J. J. (2008). Lifecourse Social Conditions and Racial and Ethnic Patterns of Cognitive Aging. *Neuropsychology Review, 18*(3), 223-254. https://doi.org/10.1007/s11065-008-9064-z
23. Seblova, D. (2020). Racial/Ethnic Differences in the Association of High School Quality with Later Life Cognitive Function. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 16*, e046179.
24. Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S., & Ingvar, M. (1998). The Illiterate Brain - Learning to Read and Write During Childhood Influences the Functional Organization of the Adult Brain. *Brain, 121*, 1053-1063. https://doi.org/10.1093/brain/121.6.1053
25. Carreiras, M., Seghier, M. L., Baquero, S., Estevez, A., Lozano, A., Devlin, J. T., & Price, C. J. (2009). An Anatomical Signature for Literacy. *Nature, 461*(7266), 983-U245. https://doi.org/10.1038/nature08461
26. Nitrini, R., Caramelli, P., Herrera Junior, E., Porto, C. S., Charchat-Fichman, H., Carthery, M. T., . . . Lima, E. P. (2004). Performance of Illiterate and Literate Nondemented Elderly Subjects in Two Tests of Long-Term Memory. *Journal of the International Neuropsychological Society, 10*(4), 634-638. https://doi.org/10.1017/S1355617704104062
27. Folia, V., & Kosmidis, M. H. (2003). Assessment of Memory Skills in Illiterates: Strategy Differences or Test Artifact? *The Clinical Neuropsychologist, 17*(2), 143-152. https://doi.org/10.1076/clin.17.2.143.16505
28. 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
29. 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
30. Nielsen, T. R., & Waldemar, G. (2016). Effects of Literacy on Semantic Verbal Fluency in an Immigrant Population. *Neuropsychology, Development, and Cognition. Section B,*

Aging, Neuropsychology and Cognition, 23(5), 578-590. https://doi.org/10.1080/13825585.2015.1132668

31. Caramelli, P., Carthery-Goulart, M. T., Porto, C. S., Charchat-Fichman, H., & Nitrini, R. (2007). Category Fluency as a Screening Test for Alzheimer Disease in Illiterate and Literate Patients. *Alzheimer Disease and Associated Disorders, 21*(1), 65-67. https://doi.org/10.1097/WAD.0b013e31802f244f
32. Nielsen, T. R., & Jorgensen, 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
33. 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.
34. Fernandez, A. L., & Marcopulos, B. A. (2008). A Comparison of Normative Data for the Trail Making Test from Several Countries: Equivalence of Norms and Considerations for Interpretation. *Scandinavian Journal of Psychology, 49*(3), 239-246. https://doi.org/10.1111/j.1467-9450.2008.00637.x
35. Zendedel, R., Schouten, B. C., van Weert, J. C. M., & van den Putte, B. (2018). 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
36. 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
37. 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, 106*, 159-167. https://doi.org/10.1016/j.socscimed.2014.01.019
38. Casas, R., Guzman-Velez, E., Cardona-Rodriguez, J., Rodriguez, N., Quinones, G., Izaguirre, B., & Tranel, D. (2012). Interpreter-Mediated Neuropsychological Testing of Monolingual Spanish Speakers. *The Clinical Neuropsychologist, 26*(1), 88-101. https://doi.org/10.1080/13854046.2011.640641
39. Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading Level Attenuates Differences in Neuropsychological Test Performance between African American and White Elders. *Journal of the International Neuropsychological Society, 8*(3), 341-348.
40. Medina, L. D., Torres, S., Gioia, A., Ochoa Lopez, A., Wang, J., & Cirino, P. T. (2021). Reporting of Demographic Variables in Neuropsychological Research: An Update of O'bryant et al.'s Trends in the Current Literature. *Journal of the International Neuropsychological Society, 27*(5), 497-507. https://doi.org/10.1017/S1355617720001083
41. Manly, J. J., & Echemendia, R. J. (2007). Race-Specific Norms: Using the Model of Hypertension to Understand Issues of Race, Culture, and Education in Neuropsychology. *Archives of Clinical Neuropsychology, 22*(3), 319-325. https://doi.org/10.1016/j.acn.2007.01.006
42. Maj, M., D'Elia, L., Satz, P., Janssen, R., Zaudig, M., Uchiyama, C., . . . World Health Organization - Division of Mental Health Global Programme on AIDS. (1993).

Evaluation of Two New Neuropsychological Tests Designed to Minimize Cultural Bias in the Assessment of HIV-1 Seropositive Persons: A Who Study. *Archives of Clinical Neuropsychology, 8*(2), 123-135.

43. Centraal Bureau voor de Statistiek. (2018). Bevolking. https://www.cbs.nl/nl-nl/achtergrond/2018/47/bevolking
44. 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*, 1-8. https://doi.org/10.1017/S0033291720000586
45. Alzheimer Nederland. (2014). Cijfers en Feiten over Dementie en Allochtonen. https://www.alzheimer-nederland.nl/sites/default/files/directupload/cijfers-feiten-dementie-allochtonen.pdf
46. Parlevliet, J. L., Uysal-Bozkir, O., Goudsmit, M., van Campen, J. P., Kok, R. M., Ter Riet, G., . . . de Rooij, S. E. (2016). Prevalence of Mild Cognitive Impairment and Dementia in Older Non-Western Immigrants in the Netherlands: A Cross-Sectional Study. *International Journal of Geriatric Psychiatry, 31*(9), 1040-1049. https://doi.org/10.1002/gps.4417
47. 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.
48. Uitewaal, P. J., Manna, D. R., Bruijnzeels, M. A., Hoes, A. W., & Thomas, S. (2004). Prevalence of Type 2 Diabetes Mellitus, Other Cardiovascular Risk Factors, and Cardiovascular Disease in Turkish and Moroccan Immigrants in North West Europe: A Systematic Review. *Preventive Medicine, 39*(6), 1068-1076. https://doi.org/10.1016/j.ypmed.2004.04.009
49. van Laer, S. D., Snijder, M. B., Agyemang, C., Peters, R. J., & van den Born, B. H. (2018). Ethnic Differences in Hypertension Prevalence and Contributing Determinants - the HELIUS Study. *European Journal of Preventive Cardiology, 25*(18), 1914-1922. https://doi.org/10.1177/2047487318803241
50. Aichberger, M. C., Schouler-Ocak, M., Mundt, A., Busch, M. A., Nickels, E., Heimann, H. M., . . . Rapp, M. A. (2010). Depression in Middle-Aged and Older First Generation Migrants in Europe: Results from the Survey of Health, Ageing and Retirement in Europe (SHARE). *European Psychiatry, 25*(8), 468-475. https://doi.org/10.1016/j.eurpsy.2009.11.009
51. Schellingerhout, R. (2004). *Gezondheid en Welzijn van Allochtone Ouderen*. Sociaal en Cultureel Planbureau.
52. Goudsmit, M., Parlevliet, J. L., van Campen, J., & Schmand, B. (2011). Dementiediagnostiek bij Oudere Migranten op de Geheugenpolikliniek: Obstakels en Oplossingen. *Tijdschrift voor Gerontologie en Geriatrie, 42*, 204-214.
53. Mukadam, N., Cooper, C., & Livingston, G. (2011). A Systematic Review of Ethnicity and Pathways to Care in Dementia. *International Journal of Geriatric Psychiatry, 26*(1), 12-20. https://doi.org/10.1002/gps.2484
54. van Wezel, N., Francke, A. L., Kayan Acun, E., Deville, W. L., van Grondelle, N. J., & Blom, M. M. (2018). Explanatory Models and Openness About Dementia in Migrant Communities: A Qualitative Study among Female Family Carers. *Dementia: The International Journal of Social Research and Practice, 17*(7), 840-857. https://doi.org/10.1177/1471301216655236

55. van Wezel, N., Francke, A. L., Kayan-Acun, E., Deville, W. L. J. M., van Grondelle, N. J., & Blom, M. M. (2016). Family Care for Immigrants with Dementia: The Perspectives of Female Family Carers Living in the Netherlands. *Dementia: The International Journal of Social Research and Practice, 15*(1), 69-84. https://doi.org/10.1177/1471301213517703
56. Borra, R., van Dijk, R., & Rohlof, H. (2012). *Cultuur, Classificatie en Diagnose: Cultuursensitief Werken met de DSM-IV*. Bohn Stafleu van Loghum.
57. 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
58. Goudsmit, M., Parlevliet, J. L., Van Campen, J. P. C. M., & Schmand, B. (2014). *De Cross-Culturele Dementiescreening (CCD)*. Bohn Stafleu van Loghum.
59. Babacan-Yıldız, G., Ur-Özçelik, E., Kolukısa, M., Işık, A. T., Gürsoy, E., Kocaman, G., & Çelebi, A. (2016). Validity and Reliability Studies of Modified Mini Mental State Examination (MMSE-E) for Turkish Illiterate Patients with Diagnosis of Alzheimer Disease. *Türk Psikiyatri Dergisi, 27*(1), 41-46.
60. 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.
61. Wang, Y.-P., & Gorenstein, C. (2013). Psychometric Properties of the Beck Depression Inventory-II: A Comprehensive Review. *Revista Brasileira de Psiquiatria, 35*(4), 416-431.
62. Gungen, C., Ertan, T., Eker, E., Yasar, R., & Engin, F. (2002). [Reliability and Validity of the Standardized Mini Mental State Examination in the Diagnosis of Mild Dementia in Turkish Population] Standardize Mini Mental Test'in Turk Toplumunda Hafif Demans Tanisinda Gecerlik Ve Guvenilirligi. *Türk Psikiyatri Dergisi, 13*(4), 273-281.
63. Güleç, H., Kavakçi, O., Yazici Güleç, M., & Küçükalioğlu, C. I. (2007). Psychometric Properties of the Turkish Version of the Frontal Assessment Battery in Patients with Schizophrenia [Sizofreni Hastalarinda Frontal Değerlendirme Bataryasi Türkçe Uyarlamasinin Psikometrik Özellikleri]. *Düsünen Adam, 20*, 151-157.
64. van Loo, E. H., Wiebrands, C., & van Laar, T. (2007). De 'Frontal Assessment Battery' (FAB) voor Screening Op Frontaalkwabpathologie bij Neurodegeneratieve Ziekten. *Tijdschrift voor Neurologie en Neurochirurgie, 108*, 115-120.
65. Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Lea & Fibiger.
66. Schmand, B., Houx, P., & de Koning, I. (2012). Normen Van Psychologische Tests Voor Gebruik in De Klinische Neuropsychologie. 1-31. https://www.psynip.nl/wp-content/uploads/2016/07/Handleiding-normen-Np-tests-2012.pdf
67. Lindeboom, J., & Schmand, B. (2003). *Visual Association Test*. PITS.
68. Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.
69. Kapci, E. G., Uslu, R., Turkcapar, H., & Karaoglan, A. (2008). Beck Depression Inventory II: Evaluation of the Psychometric Properties and Cut-Off Points in a Turkish Adult Population. *Depression and Anxiety, 25*(10), E104-110. https://doi.org/10.1002/da.20371

70. Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A New Clinical Scale for the Staging of Dementia. *The British Journal of Psychiatry, 140*, 566-572.
71. Flores, G. (2005). The Impact of Medical Interpreter Services on the Quality of Health Care: A Systematic Review. *Medical Care Research and Review, 62*(3), 255-299.
72. Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., . . . Alzheimer's Disease International. (2005). Global Prevalence of Dementia: A Delphi Consensus Study. *Lancet, 366*(9503), 2112-2117. https://doi.org/10.1016/S0140-6736(05)67889-0
73. Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The Global Prevalence of Dementia: A Systematic Review and Metaanalysis. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 9*(1), 63-75 e62. https://doi.org/10.1016/j.jalz.2012.11.007
74. World Health Organization. (2011). *Global Health and Aging.* https://www.who.int/ageing/publications/global_health.pdf
75. Nielsen, T. R., Vogel, A., Phung, T. K., Gade, A., & Waldemar, G. (2011). Over- and under-Diagnosis of Dementia in Ethnic Minorities: A Nationwide Register-Based Study. *International Journal of Geriatric Psychiatry, 26*(11), 1128-1135. https://doi.org/10.1002/gps.2650
76. Gurland, B., Wilder, D., Lantigua, R., Mayeux, R., Stern, Y., Chen, J., . . . Killeffer, E. (1997). Differences in Rates of Dementia between Ethnoracial Groups. In L. G. Martin & B. J. Soldo (Eds.), *Racial and Ethnic Differences in the Health of Older Americans* (pp. 233-269). National Academy Press.
77. Ardila, A., Rosselli, M., & Rosas, P. (1989). Neuropsychological Assessment in Illiterates: Visuospatial and Memory Abilities. *Brain and Cognition, 11*(2), 147-166.
78. Teng, E. L. (2002). Cultural and Educational Factors in the Diagnosis of Dementia. *Alzheimer Disease and Associated Disorders, 16 Suppl 2*, S77-79.
79. Julayanont, P., & Ruthirago, D. (2018). The Illiterate Brain and the Neuropsychological Assessment: From the Past Knowledge to the Future New Instruments. *Applied Neuropsychology-Adult, 25*(2), 174-187. https://doi.org/10.1080/23279095.2016.1250211
80. Paddick, S. M., Gray, W. K., McGuire, J., Richardson, J., Dotchin, C., & Walker, R. W. (2017). Cognitive Screening Tools for Identification of Dementia in Illiterate and Low-Educated Older Adults, a Systematic Review and Meta-Analysis. *International Psychogeriatrics, 29*(6), 897-929. https://doi.org/10.1017/S1041610216001976
81. Rosli, R., Tan, M. P., Gray, W. K., Subramanian, P., & Chin, A. V. (2016). Cognitive Assessment Tools in Asia: A Systematic Review. *International Psychogeriatrics, 28*(2), 189-210. https://doi.org/10.1017/S1041610215001635
82. Vasconcelos, L. G., Brucki, S. M. D., & Bueno, O. F. A. (2007). Cognitive and Functional Dementia Assessment Tools: Review of Brazilian Literature. *Dementia & Neuropsychologia, 1*(1), 18-23. https://doi.org/10.1590/S1980-57642008DN10100004
83. Jacova, C., Kertesz, A., Blair, M., Fisk, J. D., & Feldman, H. H. (2007). Neuropsychological Testing and Assessment for Dementia. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 3*(4), 299-317. https://doi.org/10.1016/j.jalz.2007.07.011
84. Puente, A. E., & Ardila, A. (2000). Neuropsychological Assessment of Hispanics. In E. Fletcher-Janzen, T. L. Strickland, & C. Reynolds (Eds.), *Handbook of Cross-Cultural Neuropsychology* (pp. 87-104). Kluwers Academic/Plenum Publishers.

85. UNESCO Institute for Statistics. (n.d.). *Primary Education, Duration (Years). All Countries and Economies.* Retrieved 22-10-2018 from data.uis.unesco.org
86. Boone, K. B., Victor, T. L., Wen, J., Razani, J., & Ponton, M. (2007). The Association between Neuropsychological Scores and Ethnicity, Language, and Acculturation Variables in a Large Patient Population. *Archives of Clinical Neuropsychology, 22*(3), 355-365. https://doi.org/10.1016/j.acn.2007.01.010
87. Carstairs, J. R., Myors, B., Shores, E. A., & Fogarty, G. (2006). Influence of Language Background on Tests of Cognitive Abilities: Australian Data. *Australian Psychologist, 41*(1), 48-54. https://doi.org/10.1080/00050060500391878
88. Kisser, J. E., Wendell, C. R., Spencer, R. J., & Waldstein, S. R. (2012). Neuropsychological Performance of Native Versus Non-Native English Speakers. *Archives of Clinical Neuropsychology, 27*(7), 749-755. https://doi.org/10.1093/arclin/acs082
89. Wong, T. (2011). Neuropsychology of Chinese Americans. In D. E. M. Fujii (Ed.), *Studies on Neuropsychology, Neurology, and Cognition. The Neuropsychology of Asian Americans* (pp. 29-46). Psychology Press.
90. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine, 25*(24), 3186-3191.
91. Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., . . . 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
92. Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., & Kleijnen, J. (2003). The Development of QUADAS: A Tool for the Quality Assessment of Studies of Diagnostic Accuracy Included in Systematic Reviews. *BMC Medical Research Methodology, 3*, 25. https://doi.org/10.1186/1471-2288-3-25
93. American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Author.
94. American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th text revised ed.). Author.
95. American Psychiatric Association. (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.). Author.
96. McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., . . . 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: The Journal of the Alzheimer's Association, 7*(3), 263-269. https://doi.org/10.1016/j.jalz.2011.03.005
97. Petersen, R. C. (2004). Mild Cognitive Impairment as a Diagnostic Entity. *Journal of Internal Medicine, 256*(3), 183-194. https://doi.org/10.1111/j.1365-2796.2004.01388.x
98. Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). Oxford University Press.
99. Chan, A., Tam, J., Murphy, C., Chiu, H., & Lam, L. (2002). Utility of Olfactory Identification Test for Diagnosing Chinese Patients with Alzheimer's Disease. *Journal of Clinical and Experimental Neuropsychology, 24*(2), 251-259. https://doi.org/10.1076/jcen.24.2.251.992

100. Kim, B. S., Lee, D. W., Bae, J. N., Kim, J. H., Kim, S., Kim, K. W., . . . Chang, S. M. (2017). Effects of Education on Differential Item Functioning on the 15-Item Modified Korean Version of the Boston Naming Test. *Psychiatry Investigation, 14*(2), 126-135. https://doi.org/10.4306/pi.2017.14.2.126
101. Lee, J. H., Lee, K. U., Lee, D. Y., Kim, K. W., Jhoo, J. H., Kim, J. H., . . . Woo, J. I. (2002). Development of the Korean Version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): Clinical and Neuropsychological Assessment Batteries. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 57*(1), P47-53.
102. Loewenstein, D. A., Arguelles, T., Barker, W. W., & Duara, R. (1993). A Comparative Analysis of Neuropsychological Test Performance of Spanish-Speaking and English-Speaking Patients with Alzheimer's Disease. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 48*(3), P142-149.
103. Shim, Y., Ryu, H. J., Lee, D. W., Lee, J. Y., Jeong, J. H., Choi, S. H., . . . Ryu, S. H. (2015). Literacy Independent Cognitive Assessment: Assessing Mild Cognitive Impairment in Older Adults with Low Literacy Skills. *Psychiatry Investigation, 12*(3), 341-348. https://doi.org/10.4306/pi.2015.12.3.341
104. de Paula, J. J., Schlottfeldt, C. G., Moreira, L., Cotta, M., Bicalho, M. A., Romano-Silva, M. A., . . . Malloy-Diniz, L. F. (2010). Psychometric Properties of a Brief Neuropsychological Protocol for Use in Geriatric Populations. *Revista de Psiquiatria Clínica, 37*(6), 251-255.
105. de Paula, J. J., Bertola, L., Avila, R. T., Moreira, L., Coutinho, G., de Moraes, E. N., . . . Malloy-Diniz, L. F. (2013). Clinical Applicability and Cutoff Values for an Unstructured Neuropsychological Assessment Protocol for Older Adults with Low Formal Education. *PLoS One, 8*(9), e73167. https://doi.org/10.1371/journal.pone.0073167
106. de Paula, J. J., Querino, E. H., Oliveira, T. D., Sedo, M., & Malloy-Diniz, L. F. (2015). Transcultural Issues on the Assessment of Executive Functions and Processing Speed in Older Adults with Low Formal Education: Usefulness of the Five Digits Test in the Assessment of Dementia. *Geriatrics & Gerontology International, 15*(3), 388-389. https://doi.org/10.1111/ggi.12364
107. Jacinto, A. F., Brucki, S. M. D., Porto, C. S., de Arruda Martins, M., de Albuquerque Citero, V., & Nitrini, R. (2014). Suggested Instruments for General Practitioners in Countries with Low Schooling to Screen for Cognitive Impairment in the Elderly. *International Psychogeriatrics, 26*(7), 1121-1125. https://doi.org/10.1017/S1041610214000325
108. Kim, H. J., Baek, M. J., & Kim, S. (2014). Alternative Type of the Trail Making Test in Nonnative English-Speakers: The Trail Making Test-Black & White. *PLoS One, 9*(2), e89078. https://doi.org/10.1371/journal.pone.0089078
109. Qiao, J., Wang, X., Lu, W., Cao, H., & Qin, X. (2016). Validation of Neuropsychological Tests to Screen for Dementia in Chinese Patients with Parkinson's Disease. *American Journal of Alzheimer's Disease and Other Dementias, 31*(4), 368-374. https://doi.org/10.1177/1533317515619478
110. Salmon, D. P., Jin, H., Zhang, M. Y., Grant, I., & Yu, E. (1995). Neuropsychological Assessment of Chinese Elderly in the Shanghai Dementia Survey. *The Clinical Neuropsychologist, 9*(2), 159-168. https://doi.org/10.1080/13854049508401598
111. Aprahamian, I., Martinelli, J. E., Neri, A. L., & Yassuda, M. S. (2010). The Accuracy of the Clock Drawing Test Compared to That of Standard Screening Tests for Alzheimer's Disease: Results from a Study of Brazilian Elderly with Heterogeneous Educational

Backgrounds. *International Psychogeriatrics, 22*(1), 64-71. https://doi.org/10.1017/S1041610209991141

112. Baiyewu, O., Unverzagt, F. W., Lane, K. A., Gureje, O., Ogunniyi, A., Musick, B., . . . Hendrie, H. C. (2005). The Stick Design Test: A New Measure of Visuoconstructional Ability. *Journal of the International Neuropsychological Society, 11*(5), 598-605. https://doi.org/10.1017/S135561770505071X
113. Chan, C. C., Yung, C. Y., & Pan, P. C. (2005). Screening of Dementia in Chinese Elderly Adults by the Clock Drawing Test and the Time and Change Test. *Hong Kong Medical Journal, 11*(1), 13-19.
114. Das, S. K., Bose, P., Biswas, A., Dutt, A., Banerjee, T. K., Hazra, A. M., . . . Roy, T. (2007). An Epidemiologic Study of Mild Cognitive Impairment in Kolkata, India. *Neurology, 68*(23), 2019-2026. https://doi.org/10.1212/01.wnl.0000264424.76759.e6
115. 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
116. Lam, L. C., Chiu, H. F., Ng, K. O., Chan, C., Chan, W. F., Li, S. W., & Wong, M. (1998). Clock-Face Drawing, Reading and Setting Tests in the Screening of Dementia in Chinese Elderly Adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 53*(6), P353-357.
117. Park, S. J., Lee, J. E., Lee, K. S., & Kim, J. S. (2018). Comparison of Odor Identification among Amnestic and Non-Amnestic Mild Cognitive Impairment, Subjective Cognitive Decline, and Early Alzheimer's Dementia. *Neurological Sciences, 39*(3), 557-564. https://doi.org/10.1007/s10072-018-3261-1
118. Sahadevan, S., Lim, J. P., Tan, N. J., & Chan, S. P. (2002). Psychometric Identification of Early Alzheimer Disease in an Elderly Chinese Population with Differing Educational Levels. *Alzheimer Disease and Associated Disorders, 16*(2), 65-72.
119. Storey, J. E., Rowland, J. T., Basic, D., & Conforti, D. A. (2002). Accuracy of the Clock Drawing Test for Detecting Dementia in a Multicultural Sample of Elderly Australian Patients. *International Psychogeriatrics, 14*(3), 259-271.
120. Yap, P. L., Ng, T. P., Niti, M., Yeo, D., & Henderson, L. (2007). Diagnostic Performance of Clock Drawing Test by CLOX in an Asian Chinese Population. *Dementia and Geriatric Cognitive Disorders, 24*(3), 193-200. https://doi.org/10.1159/000107080
121. Chiu, H. F., Chan, C. K., Lam, L. C., Ng, K. O., Li, S. W., Wong, M., & Chan, W. F. (1997). The Modified Fuld Verbal Fluency Test: A Validation Study in Hong Kong. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 52*(5), P247-250.
122. de Paula, J. J., Moreira, L., Nicolato, R., de Marco, L. A., Correa, H., Romano-Silva, M. A., . . . Malloy-Diniz, L. F. (2012). The Tower of London Test: Different Scoring Criteria for Diagnosing Alzheimer's Disease and Mild Cognitive Impairment. *Psychological Reports, 110*(2), 477-488. https://doi.org/10.2466/03.10.13.PR0.110.2.477-488
123. Mok, E. H., Lam, L. C., & Chiu, H. F. (2004). Category Verbal Fluency Test Performance in Chinese Elderly with Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders, 18*(2), 120-124. https://doi.org/10.1159/000079190
124. Radanovic, M., Carthery-Goulart, M. T., Charchat-Fichman, H., Herrera, E., Jr., Lima, E. E. P., Smid, J., . . . Nitrini, R. (2007). Analysis of Brief Language Tests in the Detection

of Cognitive Decline and Dementia. *Dementia & Neuropsychologia, 1*(1), 37-45. https://doi.org/10.1590/S1980-57642008DN10100007

125. Fernandez, A. L. (2013). Development of a Confrontation Naming Test for Spanish-Speakers: The Cordoba Naming Test. *The Clinical Neuropsychologist, 27*(7), 1179-1198. https://doi.org/10.1080/13854046.2013.822931
126. Kim, H., & Na, D. L. (1999). Normative Data on the Korean Version of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology, 21*(1), 127-133. https://doi.org/10.1076/jcen.21.1.127.942
127. Marquez de la Plata, C., Vicioso, B., Hynan, L., Evans, H. M., Diaz-Arrastia, R., Lacritz, L., & Cullum, C. M. (2008). Development of the Texas Spanish Naming Test: A Test for Spanish Speakers. *The Clinical Neuropsychologist, 22*(2), 288-304. https://doi.org/10.1080/13854040701250470
128. Marquez de la Plata, C., Arango-Lasprilla, J. C., Alegret, M., Moreno, A., Tarraga, L., Lara, M., . . . Cullum, C. M. (2009). Item Analysis of Three Spanish Naming Tests: A Cross-Cultural Investigation. *NeuroRehabilitation, 24*(1), 75-85. https://doi.org/10.3233/NRE-2009-0456
129. Baek, M. J., Kim, H. J., & Kim, S. (2012). Comparison between the Story Recall Test and the Word-List Learning Test in Korean Patients with Mild Cognitive Impairment and Early Stage of Alzheimer's Disease. *Journal of Clinical and Experimental Neuropsychology, 34*(4), 396-404. https://doi.org/10.1080/13803395.2011.645020
130. Chang, C. C., Kramer, J. H., Lin, K. N., Chang, W. N., Wang, Y. L., Huang, C. W., . . . Wang, P. N. (2010). Validating the Chinese Version of the Verbal Learning Test for Screening Alzheimer's Disease. *Journal of the International Neuropsychological Society, 16*(2), 244-251. https://doi.org/10.1017/S1355617709991184
131. Chung, J. C. (2009). Clinical Validity of Fuld Object Memory Evaluation to Screen for Dementia in a Chinese Society. *International Journal of Geriatric Psychiatry, 24*(2), 156-162. https://doi.org/10.1002/gps.2085
132. Grober, E., Ehrlich, A. R., Troche, Y., Hahn, S., & Lipton, R. B. (2014). Screening Older Latinos for Dementia in the Primary Care Setting. *Journal of the International Neuropsychological Society, 20*(8), 848-855. https://doi.org/10.1017/S1355617714000708
133. Loewenstein, D. A., Duara, R., Arguelles, T., & Arguelles, S. (1995). Use of the Fuld Object-Memory Evaluation in the Detection of Mild Dementia among Spanish and English-Speaking Groups. *The American Journal of Geriatric Psychiatry, 3*(4), 300-307. https://doi.org/10.1097/00019442-199503040-00004
134. Maillet, D., Narme, P., Amieva, H., Matharan, F., Bailon, O., Le Clesiau, H., & Belin, C. (2017). The TMA-93: A New Memory Test for Alzheimer's Disease in Illiterate and Less Educated People. *American Journal of Alzheimer's Disease and Other Dementias, 32*(8), 461-467. https://doi.org/10.1177/1533317517722630
135. Maillet, D., Matharan, F., Le Clesiau, H., Bailon, O., Peres, K., Amieva, H., & Belin, C. (2016). TNI-93: A New Memory Test for Dementia Detection in Illiterate and Low-Educated Patients. *Archives of Clinical Neuropsychology, 31*(8), 896-903. https://doi.org/10.1093/arclin/acw065
136. Rideaux, T., Beaudreau, S. A., Fernandez, S., & O'Hara, R. (2012). Utility of the Abbreviated Fuld Object Memory Evaluation and MMSE for Detection of Dementia and Cognitive Impairment Not Dementia in Diverse Ethnic Groups. *Journal of Alzheimer's disease: JAD, 31*(2), 371-386. https://doi.org/10.3233/JAD-2012-112180

137. Saka, E., Mihci, E., Topcuoglu, M. A., & Balkan, S. (2006). Enhanced Cued Recall Has a High Utility as a Screening Test in the Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment in Turkish People. *Archives of Clinical Neuropsychology, 21*(7), 745-751. https://doi.org/10.1016/j.acn.2006.08.007
138. Takada, L. T., Caramelli, P., Fichman, H. C., Porto, C. S., Bahia, V. S., Anghinah, R., . . . Nitrini, R. (2006). Comparison between Two Tests of Delayed Recall for the Diagnosis of Dementia. *Arquivos de Neuro-psiquiatria, 64*(1), 35-40.
139. Verghese, J., Noone, M. L., Johnson, B., Ambrose, A. F., Wang, C., Buschke, H., . . . Mathuranath, P. S. (2012). Picture-Based Memory Impairment Screen for Dementia. *Journal of the American Geriatrics Society, 60*(11), 2116-2120. https://doi.org/10.1111/j.1532-5415.2012.04191.x
140. Nielsen, T. R., Segers, K., Vanderaspoilden, V., Bekkhus-Wetterberg, P., Minthon, L., Pissiota, A., . . . 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
141. Unverzagt, F. W., Morgan, O. S., Thesiger, C. H., Eldemire, D. A., Luseko, J., Pokuri, S., . . . Hendrie, H. C. (1999). Clinical Utility of CERAD Neuropsychological Battery in Elderly Jamaicans. *Journal of the International Neuropsychological Society, 5*(3), 255-259.
142. Wu, J. B., Lyu, Z. H., Liu, X. J., Li, H. P., & Wang, Q. (2017). Development and Standardization of a New Cognitive Assessment Test Battery for Chinese Aphasic Patients: A Preliminary Study. *Chinese Medical Journal, 130*(19), 2283-2290. https://doi.org/10.4103/0366-6999.215326
143. Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., . . . Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and Neuropsychological Assessment of Alzheimer's Disease. *Neurology, 39*(9), 1159-1165.
144. Nielsen, T. R., Segers, K., Vanderaspoilden, V., Beinhoff, U., Minthon, L., Pissiota, A., . . . Waldemar, G. (2018). Validation of a European Cross-Cultural Neuropsychological Test Battery (CNTB) for Evaluation of Dementia. *International Journal of Geriatric Psychiatry*. https://doi.org/10.1002/gps.5002
145. Daugherty, J. C., Puente, A. E., Fasfous, A. F., Hidalgo-Ruzzante, N., & Perez-Garcia, 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
146. Espino, D. V., & Lewis, R. (1998). Dementia in Older Minority Populations. Issues of Prevalence, Diagnosis, and Treatment. *The American Journal of Geriatric Psychiatry, 6*(2 Suppl 1), S19-25.
147. Abou-Mrad, F., Tarabey, L., Zamrini, E., Pasquier, F., Chelune, G., Fadel, P., & Hayek, M. (2015). Sociolinguistic Reflection on Neuropsychological Assessment: An Insight into Selected Culturally Adapted Battery of Lebanese Arabic Cognitive Testing. *Neurological Sciences, 36*(10), 1813-1822. https://doi.org/10.1007/s10072-015-2257-3
148. Arguelles, T., Loewenstein, D., & Arguelles, S. (2001). The Impact of the Native Language of Alzheimer's Disease and Normal Elderly Individuals on Their Ability to Recall Digits. *Aging and Mental Health, 5*(4), 358-365. https://doi.org/10.1080/13607860120083314

149. Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1986). Digit Memory in Chinese and English: Evidence for a Temporally Limited Store. *Cognition, 23*(1), 1-20.
150. 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.
151. Ayabe-Kanamura, S., Saito, S., Distel, H., Martinez-Gomez, M., & Hudson, R. (1998). Differences and Similarities in the Perception of Everyday Odors. A Japanese-German Cross-Cultural Study. *Annals of the New York Academy of Sciences, 855*, 694-700.
152. Alves, J., Petrosyan, A., & Magalhaes, R. (2014). Olfactory Dysfunction in Dementia. *World Journal of Clinical Cases, 2*(11), 661-667. https://doi.org/10.12998/wjcc.v2.i11.661
153. Johns, E. K., Phillips, N. A., Belleville, S., Goupil, D., Babins, L., Kelner, N., . . . Chertkow, H. (2009). Executive Functions in Frontotemporal Dementia and Lewy Body Dementia. *Neuropsychology, 23*(6), 765-777. https://doi.org/10.1037/a0016792 19899835
154. Levy, G., Jacobs, D. M., Tang, M. X., Cote, L. J., Louis, E. D., Alfaro, B., . . . Marder, K. (2002). Memory and Executive Function Impairment Predict Dementia in Parkinson's Disease. *Movement Disorders: Official Journal of the Movement Disorder Society, 17*(6), 1221-1226. https://doi.org/10.1002/mds.10280
155. Armentano, C. G. D., Porto, C. S., Brucki, S. M. D., & Nitrini, R. (2009). Study on the Behavioural Assessment of the Dysexecutive Syndrome (BADS) Performance in Healthy Individuals, Mild Cognitive Impairment and Alzheimer's Disease: A Preliminary Study. *Dementia & Neuropsychologia, 3*(2), 101-107. https://doi.org/10.1590/S1980-57642009DN30200006
156. Armentano, C. G. D., Porto, C. S., Nitrini, R., & Brucki, S. M. D. (2013). Ecological Evaluation of Executive Functions in Mild Cognitive Impairment and Alzheimer Disease. *Alzheimer Disease & Associated Disorders, 27*(2), 95-101. https://doi.org/10.1097/WAD.0b013e3182654 0b4
157. Abou-Mrad, F., Chelune, G., Zamrini, E., Tarabey, L., Hayek, M., & Fadel, P. (2017). Screening for Dementia in Arabic: Normative Data from an Elderly Lebanese Sample. *The Clinical Neuropsychologist, 31*(sup1), 1-19. https://doi.org/10.1080/13854046.2017.1288270
158. Galvez-Lara, M., Moriana, J. A., Vilar-Lopez, R., Fasfous, A. F., Hidalgo-Ruzzante, N., & Perez-Garcia, M. (2015). Validation of the Cross-Linguistic Naming Test: A Naming Test for Different Cultures? A Preliminary Study in the Spanish Population. *Journal of Clinical and Experimental Neuropsychology, 37*(1), 102-112. https://doi.org/10.1080/13803395.2014.1003533
159. Thompson, T. A. C., Wilson, P. H., Snyder, P. J., Pietrzak, R. H., Darby, D., Maruff, P., & Buschke, H. (2011). Sensitivity and Test-Retest Reliability of the International Shopping List Test in Assessing Verbal Learning and Memory in Mild Alzheimer's Disease. *Archives of Clinical Neuropsychology, 26*(5), 412-424. https://doi.org/10.1093/arclin/acr039
160. Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological Validity of a Simplified Version of the Multiple Errands Shopping Test. *Journal of the International Neuropsychological Society, 9*(1), 31-44.
161. Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., & Riva, G. (2014). Virtual Multiple Errands Test (VMET): A Virtual Reality-Based Tool to Detect

Early Executive Functions Deficit in Parkinson's Disease. *Frontiers in Behavioral Neuroscience, 8*, 405. https://doi.org/10.3389/fnbeh.2014.00405

162. Besnard, J., Richard, P., Banville, F., Nolin, P., Aubin, G., Le Gall, D., . . . Allain, P. (2016). Virtual Reality and Neuropsychological Assessment: The Reliability of a Virtual Kitchen to Assess Daily-Life Activities in Victims of Traumatic Brain Injury. *Applied Neuropsychology-Adult, 23*(3), 223-235. https://doi.org/10.1080/23279095.2015.1048514
163. Jovanovski, D., Zakzanis, K., Ruttan, L., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Ecologically Valid Assessment of Executive Dysfunction Using a Novel Virtual Reality Task in Patients with Acquired Brain Injury. *Applied Neuropsychology-Adult, 19*(3), 207-220. https://doi.org/10.1080/09084282.2011.643956
164. Nabors, N. A., Evans, J. D., & Strickland, T. L. (2000). Neuropsychological Assessment and Intervention with African Americans. In E. Fletcher-Janzen, T. L. Strickland, & C. R. Reynolds (Eds.), *Handbook of Cross-Cultural Neuropsychology* (pp. 31-42). Kluwer Academic/Plenum.
165. Silverberg, N. D., Hanks, R. A., & Tompkins, S. C. (2013). Education Quality, Reading Recognition, and Racial Differences in the Neuropsychological Outcome from Traumatic Brain Injury. *Archives of Clinical Neuropsychology, 28*(5), 485-491. https://doi.org/10.1093/arclin/act023
166. Fasfous, A. F., Al-Joudi, H. F., Puente, A. E., & Perez-Garcia, M. (2017). Neuropsychological Measures in the Arab World: A Systematic Review. *Neuropsychology Review, 27*(2), 158-173. https://doi.org/10.1007/s11065-017-9347-3
167. International Test Commission. (2017). The ITC Guidelines for Translating and Adapting Tests. www.InTestCom.org
168. Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.). (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum Associates, Inc.
169. Iliescu, D. (2017). *Adapting Tests in Linguistic and Cultural Situations*. Cambridge University Press.
170. van Mol, C., & de Valk, H. (2016). Migration and Immigrants in Europe: A Historical and Demographic Perspective. In B. Garcés-Mascareñas & R. Penninx (Eds.), *Integration Processes and Policies in Europe. Imiscoe Research Series.* (pp. 31-55). Springer.
171. Rosenbaum, B., Kristensen, M., & Schmidt, J. (2008). [Dementia in Elderly Turkish Immigrants] Demenslidelser Hos Aeldre Tyrkiske Indvandrere. *Ugeskr Laeger, 170*(50), 4109-4113.
172. Monge-Maillo, B., & Lopez-Velez, R. (2012). Migration and Malaria in Europe. *Mediterranean Journal of Hematology & Infectious Diseases, 4*(1), e2012014. https://doi.org/10.4084/MJHID.2012.014
173. Lingscheid, T., Kurth, F., Clerinx, J., Marocco, S., Trevino, B., Schunk, M., . . . TropNet Schistosomiasis Investigator Group. (2017). Schistosomiasis in European Travelers and Migrants: Analysis of 14 Years Tropnet Surveillance Data. *American Journal of Tropical Medicine and Hygiene, 97*(2), 567-574. https://doi.org/10.4269/ajtmh.17-0034
174. Kristiansen, M., Razum, O., Tezcan-Güntekin, H., & Krasnik, A. (2016). Aging and Health among Migrants in a European Perspective. *Public Health Reviews, 37*, 20. https://doi.org/10.1186/s40985-016-0036-1

175. Parrón, T., Hernández, A. F., Pla, A., & Villanueva, E. (1996). Clinical and Biochemical Changes in Greenhouse Sprayers Chronically Exposed to Pesticides. *Human & ExperimentalToxicology, 15*(12), 957-963. https://doi.org/10.1177/096032719601501203
176. Blakemore, A., Kenning, C., Mirza, N., Daker-White, G., Panagioti, M., & Waheed, W. (2018). Dementia in UK South Asians: A Scoping Review of the Literature. *BMJ Open, 8*(4), e020290. https://doi.org/10.1136/bmjopen-2017-020290
177. Van Tubergen, F., & Kalmijn, M. (2005). Destination-Language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries. *American Journal of Sociology, 110*(5), 1412-1457. https://doi.org/10.1086/428931
178. 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.
179. Danmarks Statistik. (2018). Indvandrere I Danmark. 51-72. https://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=29445&sid=indv2018
180. 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.
181. Rosselli, M., Ardila, A., & Rosas, P. (1990). Neuropsychological Assessment in Illiterates. I. Language and Praxic Abilities. *Brain and Cognition, 12*(2), 281-296.
182. Crul, M., & Doomernik, J. (2003). The Turkish and Moroccan Second Generation in the Netherlands: Divergent Trends between and Polarization within the Two Groups. *The International Migration Review, 37*(4), 1039-1064.
183. Martin, P. M. (2014). Countries of Immigration - Germany. In J. F. Hollifield, P. L. Martin, & P. M. Orrenius (Eds.), *Controlling Immigration: A Global Perspective* (pp. 224-251). Stanford University Press.
184. Franzen, S., van den Berg, E., Goudsmit, M., Jurgens, C. K., van de Wiel, L., Kalkisim, Y., . . . 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, 26*(3), 331-351. https://doi.org/10.1017/S1355617719000894
185. Oumellal, A., El Alaoui Faris, M., & Benabdeljlil, M. (2018). The Trail Making Test in Morocco: Normative Data Stratified by Age and Level of Education. *Open Journal of Medical Psychology, 7*(1), 1-12.
186. Azdad, A., Benabdeljlil, M., Al Zemmouri, K., & El Alaoui Faris, M. (2019). Standardization and Validation of Montreal Cognitive Assessment (MoCA) in the Moroccan Population. *International Journal of Brain and Cognitive Sciences, 8*(1), 1–5.
187. Karakas, S., Erdogan Bakar, E., & Dogutepe Dinçer, E. (2013). *Bilnot Bataryası El Kitabı: Nöropsikolojik Testlerin Yetiskinler Için Arastırma Ve Gelistirme Çalısmaları: Bilnot-Yetiskin (Cilt I)*. Egitim Yayınevi.
188. American Psychological Association. (2003). Guidelines on Multicultural Education, Training, Research, Practice, and Organizational Change for Psychologists. *The American Psychologist, 58*(5), 377-402. https://doi.org/10.1037/0003-066X.58.5.377.
189. Hessen, E., Hokkanen, L., Ponsford, J., van Zandvoort, M. J., Watts, A., Evans, J. D., & Haaland, K. Y. (2018). Core Competencies in Clinical Neuropsychology Training across the World. *The Clinical Neuropsychologist, 32*(4), 642-656. https://doi.org/10.1080/13854046.2017.1413210 29214891

190. Smith, G. (2018). Education and Training in Clinical Neuropsychology: Recent Developments and Documents from the Clinical Neuropsychology Synarchy. *Archives of Clinical Neuropsychology, 34*(3), 418-431. https://doi.org/10.1093/arclin/acy075
191. Engerman, S. L., & Sokoloff, K. L. (2012). Five Hundred Years of European Colonization: Inequality and Paths of Development. In C. Lloyd, J. Metzer, & R. Sutch (Eds.), *Settler Economies in World History* (pp. 65-103). Brill. https://doi.org/https://doi.org/10.1163/9789004232655_005
192. Diez Guardia, N., & Pichelmann, K. (2006). *Labour Migration Patterns in Europe: Recent Trends, Future Challenges.* https://ec.europa.eu/economy_finance/publications/pages/publication644_en.pdf
193. Nielsen, T. R., Vogel, A., Riepe, M. W., de Mendonca, A., Rodriguez, G., Nobili, F., . . . 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
194. Hsu, C.-C., & Sandford, B. A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1-8.
195. Iqbal, S., & Pipon-Young, L. (2009). The Delphi Method. *The Psychologist, 22*(7), 598-601.
196. 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
197. Nielsen, T. R., Andersen, B. B., Gottrup, H., Lutzhoft, J. H., Hogh, P., & Waldemar, G. (2013). Validation of the Rowland Universal Dementia Assessment Scale for Multicultural Screening in Danish Memory Clinics. *Dementia and Geriatric Cognitive Disorders, 36*(5-6), 354-362. https://doi.org/10.1159/000354375
198. Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State". A Practical Method for Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research, 12*(3), 189-198.
199. Rait, G., Burns, A., Baldwin, R., Morley, M., Chew-Graham, C., & St Leger, A. S. (2000). Validating Screening Instruments for Cognitive Impairment in Older South Asians in the United Kingdom. *International Joural of Geriatric Psychiatry, 15*(1), 54-62. https://doi.org/10.1002/(sici)1099-1166(200001)15:1<54::aid-gps77>3.0.co;2-c
200. Rait, G., Morley, M., Burns, A., Baldwin, R., Chew-Graham, C., & St Leger, A. S. (2000). Screening for Cognitive Impairment in Older African-Caribbeans. *Psychological Medicine, 30*(4), 957-963. https://doi.org/10.1017/s0033291799002305
201. Franzen, S., van den Berg, E., Kalkisim, Y., van de Wiel, L., Harkes, M., van Bruchem-Visser, R. L., . . . 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
202. Nielsen, T. R., Segers, K., Vanderaspoilden, V., Beinhoff, U., Minthon, L., Pissiota, A., . . . Waldemar, G. (2019). Validation of a Brief Multicultural Cognitive Examination (MCE) for Evaluation of Dementia. *International Journal of Geriatric Psychiatry, 34*(7), 982-989. https://doi.org/10.1002/gps.5099

203. Narme, P., Maillet, D., Palisson, J., Le Clesiau, H., Moroni, C., & Belin, C. (2019). How to Assess Executive Functions in a Low-Educated and Multicultural Population Using a Switching Verbal Fluency Test (the TFA-93) in Neurodegenerative Diseases? *American Journal of Alzheimer's Disease and Other Dementias, 34*(7-8), 469-477. https://doi.org/10.1177/1533317519833844
204. Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x
205. 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
206. Nielsen, T. R., Segers, K., Vanderaspoilden, V., Bekkhus-Wetterberg, P., Bjorklof, G. H., Beinhoff, U., . . . Waldemar, G. (2019). Validation of the Rowland Universal Dementia Assessment Scale (RUDAS) in a Multicultural Sample across Five Western European Countries: Diagnostic Accuracy and Normative Data. *International Psychogeriatrics, 31*(2), 287-296. https://doi.org/10.1017/S1041610218000832
207. Lim, Y. Y., Prang, K. H., Cysique, L., Pietrzak, R. H., Snyder, P. J., & Maruff, P. (2009). A Method for Cross-Cultural Adaptation of a Verbal Memory Assessment. *Behavior Research Methods, 41*(4), 1190-1200. https://doi.org/10.3758/BRM.41.4.1190
208. Buhl, C., Stokholm, J., & Gade, A. (2013). Clinical Utility of Short Social Cognitive Tests in Early Differentiation of Behavioral Variant Frontotemporal Dementia from Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders Extra, 3*(1), 376-385. https://doi.org/10.1159/000355123
209. Merceron, K., & Prouteau, A. (2013). Évaluation de la Cognition Sociale en Langue Française chez L'adulte : Outils Disponibles et Recommandations de Bonne Pratique Clinique. *L'Évolution Psychiatrique, 72*, 53-70.
210. Merten, T. (2019, July 13). Foreign-Language Adaptations of Svts/Pvts and Intercultural Problems. Sixth European Conference on Symptom Validity Assessment, Amsterdam, the Netherlands.
211. Daugherty, J. C., Querido, L., Quiroz, N., Wang, D., Hidalgo-Ruzzante, N., Fernandes, S., . . . Valera, E. (2021). The Coin in Hand-Extended Version: Development and Validation of a Multicultural Performance Validity Test. *Assessment, 28*(1), 186-198. https://doi.org/10.1177/1073191119864652
212. 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
213. Nijdam-Jones, A., Rivera, D., Rosenfeld, B., & Arango-Lasprilla, J. C. (2019). The Effect of Literacy and Culture on Cognitive Effort Test Performance: An Examination of the Test of Memory Malingering in Colombia. *Journal of Clinical and Experimental Neuropsychology, 41*(10), 1015-1023. https://doi.org/10.1080/13803395.2019.1644294
214. 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

215. 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
216. Steele, C. M. (1997). A Threat in the Air. How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist, 52*(6), 613-629.
217. Carta, M. G., Bernal, M., Hardoy, M. C., Haro-Abad, J. M., & Report on the Mental Health in Europe Working Group. (2005). Migration and Mental Health in Europe. *Clinical Practice and Epidemiology in Mental Health, 1*, 13. https://doi.org/10.1186/1745-0179-1-13
218. Fazil, Q., Wallace, L. M., & Hussain, A. (2006). An Exploration of the Explanatory Models of Illness Amongst Pushtuun Families Living in the UK Who Are High Attenders in General Practice. *Diversity and Equality in Health and Care, 3*, 171-181.
219. de Freitas, D. F., Fernandes-Jesus, M., Ferreira, P. D., & Coimbra, S. (2018). Psychological Correlates of Perceived Ethnic Discrimination in Europe: A Metaanalysis. *Psychology of Violence, 8*(6), 712–725. https://doi.org/10.1037/vio0000215
220. Sempertegui, G. A., Knipscheer, J. W., Baliatsas, C., & Bekker, M. H. J. (2019). Symptom Manifestation and Treatment Effectiveness, -Obstacles and -Facilitators in Turkish and Moroccan Groups with Depression in European Countries: A Systematic Review. *Journal of Affective Disorders, 247*, 134-155. https://doi.org/10.1016/j.jad.2018.12.060
221. Elbulok-Charcape, M. M., Rabin, L. A., Spadaccini, A. T., & Barr, W. B. (2014). Trends in the Neuropsychological Assessment of Ethnic/Racial Minorities: A Survey of Clinical Neuropsychologists in the United States and Canada. *Cultural Diversity and Ethnic Minority Psychology, 20*(3), 353-361. https://doi.org/10.1037/a0035023
222. Rivera Mindt, M., Byrd, D., Saez, P., & Manly, J. (2010). Increasing Culturally Competent Neuropsychological Services for Ethnic Minority Populations: A Call to Action. *The Clinical Neuropsychologist, 24*(3), 429-453. https://doi.org/10.1080/13854040903058960
223. Hokkanen, L., Lettner, S., Barbosa, F., Constantinou, M., Harper, L., Kasten, E., . . . Hessen, E. (2019). Training Models and Status of Clinical Neuropsychologists in Europe: Results of a Survey on 30 Countries. *The Clinical Neuropsychologist, 33*(1), 32-56. https://doi.org/10.1080/13854046.2018.1484169
224. Hadziabdic, E., Heikkila, K., Albin, B., & Hjelm, K. (2011). Problems and Consequences in the Use of Professional Interpreters: Qualitative Analysis of Incidents from Primary Healthcare. *Nursing Inquiry, 18*(3), 253-261. https://doi.org/10.1111/j.1440-1800.2011.00542.x
225. Tribe, R., & Thompson, K. (2017). Working with Interpreters: Guidelines for Psychologists. 1-34. https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Working%20with%20interpreters%20-%20guidelines%20for%20psychologists.pdf
226. Turoff, M. (2002). The Policy Delphi. In H. Linstone & M. Turoff (Eds.), *The Delphi Method* (pp. 80-96). https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf
227. Takizawa, C., Thompson, P. L., van Walsem, A., Faure, C., & Maier, W. C. (2015). Epidemiological and Economic Burden of Alzheimer's Disease: A Systematic Literature Review of Data across Europe and the United States of America. *Journal of Alzheimer's Disease, 43*(4), 1271-1284.
228. Crul, M., & Vermeulen, H. (2003). The Second Generation in Europe. *The International Migration Review, 37*(4), 965-986.

229. 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*, 815-830. https://doi.org/10.1093/arclin/acaa083
230. Zendedel, R., Schouten, B. C., van Weert, J. C. M., & van den Putte, B. (2018). Informal Interpreting in General Practice: The Migrant Patient's Voice. *Ethnicity & Health, 23*(2), 158-173. https://doi.org/10.1080/13557858.2016.1246939
231. Aghvinian, M., Santoro, A. F., Gouse, H., Joska, J. A., Linda, T., Thomas, K. G. F., & Robbins, R. N. (2020). Taking the Test: A Qualitative Analysis of Cultural and Contextual Factors Impacting Neuropsychological Assessment of Xhosa-Speaking South Africans. *Archives of Clinical Neuropsychology*. https://doi.org/10.1093/arclin/acaa115
232. Franzen, S., on behalf of the European Consortium on Cross-Cultural Neuropsychology (ECCroN). (2021). Cross-Cultural Neuropsychological Assessment in Europe: Position Statement of the European Consortium on Cross-Cultural Neuropsychology (ECCroN). *The Clinical Neuropsychologist*, Advance online publication. https://doi.org/10.1080/13854046.2021.1981456
233. 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
234. Ponsford, J. (2017). International Growth of Neuropsychology. *Neuropsychology, 31*(8), 921-933. https://doi.org/10.1037/neu0000415
235. 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
236. Arnett, P. A. (Ed.). (2013). *Secondary Influences on Neuropsychological Test Performance: Research Findings and Practical Applications* (1 ed.). Oxford University Press.
237. National Center for Education Statistics. (2003). Adult Literacy Supplemental Assessment. https://nces.ed.gov/naal/pdf/AdultLiteracyFactSheet.pdf
238. 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.
239. Corsi, P. M. (1972). Human Memory and the Medial Temporal Region of the Brain. *Dissertation Abstracts International, 34*, 819B.
240. 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
241. Franzen, S., van den Berg, E., Ayhan, Y., Satoer, D. D., Türkoğlu, Ö., Genç Akpulat, E., . . . Papma, J. M. (2022). The Naming Assessment in Multicultural Europe (NAME): Development and Validation in a Multicultural Memory Clinic. *Journal of the International Neuropsychological Society,* Advance online publication. https://doi.org/10.1017/S135561772100148X
242. 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.

243. Goudsmit, M., van Campen, J., Franzen, S., van den Berg, E., Schilt, T., & Schmand, B. (2020). Dementia Detection with a Combination of Informant-Based and Performance-Based Measures in Low-Educated and Illiterate Elderly Migrants. *The Clinical Neuropsychologist*, 1-19. https://doi.org/10.1080/13854046.2020.1711967
244. 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. In *Health Status of Older Migrants in the Netherlands [Unpublished Dissertation].*
245. 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.
246. Gilmore-Bykovskyi, A. L., Jin, Y., Gleason, C., Flowers-Benton, S., Block, L. M., Dilworth-Anderson, P., . . . Zuelsdorff, M. (2019). Recruitment and Retention of Underrepresented Populations in Alzheimer's Disease Research: A Systematic Review. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 5*, 751-770. https://doi.org/10.1016/j.trci.2019.09.018
247. 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.
248. Jessen, F., Amariglio, R. E., Buckley, R. F., van der Flier, W. M., Han, Y., Molinuevo, J. L., . . . Wagner, M. (2020). The Characterisation of Subjective Cognitive Decline. *Lancet Neurology, 19*(3), 271-278. https://doi.org/10.1016/S1474-4422(19)30368-0
249. Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . 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: The Journal of the Alzheimer's Association, 7*(3), 270-279. https://doi.org/10.1016/j.jalz.2011.03.008 21514249
250. Roman, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., . . . Scheinberg, P. (1993). Vascular Dementia: Diagnostic Criteria for Research Studies. Report of the NINDS-AIREN International Workshop. *Neurology, 43*(2), 250-260.
251. American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596
252. Ortega, L. V., Aprahamian, I., Martinelli, J. E., Cecchini, M. A., Cacao, 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
253. 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
254. 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 18609332

255. 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
256. 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, 4*(6), 531-538.
257. 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
258. Spreij, L. A., Gosselt, I. K., Visser-Meily, J. M. A., & Nijboer, T. C. W. (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
259. Gruters, A. A. A., Christie, H. L., Ramakers, I. H. G. B., Verhey, F. R. J., Kessels, R. P. C., & de Vugt, M. E. (2020). Neuropsychological Assessment and Diagnostic Disclosure at a Memory Clinic: A Qualitative Study of the Experiences of Patients and Their Family Members. *Clinical Neuropsychologist*. https://doi.org/10.1080/13854046.2020.1749936
260. 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. *Clinical Neuropsychologist, 24*(5), 759-778. https://doi.org/10.1080/13854041003712951
261. Quesque, F., Coutrot, A., Cruz de Souza, L., Baez, S., Cardona, J. F., Neely-Prado, A., . . . Bertoux, M. (2020). Culture Shapes Our Understanding of Others' Thoughts and Emotions: An Investigation across 12 Countries. https://hal.archives-ouvertes.fr/hal-02995977
262. Lecours, A. R., Mehler, J., Parente, M. A., Caldeira, A., Cary, L., Castro, M. J., . . . Soares Junqueira, A. M. (1987). Illiteracy and Brain Damage - 1. Aphasia Testing in Culturally Contrasted Populations (Control Subjects). *Neuropsychologia, 25*(1B), 231-245.
263. Reis, A., Guerreiro, M., & Castro-Caldas, A. (1994). Influence of Educational Level of Non Brain-Damaged Subjects on Visual Naming Capacities. *Journal of Clinical and Experimental Neuropsychology, 16*(6), 939-942. https://doi.org/10.1080/01688639408402705
264. Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual Association Test to Detect Early Dementia of the Alzheimer Type. *Journal of Neurology Neurosurgery and Psychiatry, 73*(2), 126-133. https://doi.org/10.1136/jnnp.73.2.126
265. Kessels, R. P., Nys, G. M., Brands, A. M., & van Zandvoort, M. J. (2004). [The Location Learning Test as a Measure of Spatial Memory: Applicability of a Modified Administration Procedure and Normative Data] De Location Learning Test als Maat voor het Ruimtelijk Geheugen: Bruikbaarheid van een Nieuwe Afnameprocedure en Normgegevens. *Tijdschrift voor Gerontologie en Geriatrie, 35*(4), 147-152.
266. Grober, E., Buschke, H., Crystal, H., Bang, S., & Dresner, R. (1988). Screening for Dementia by Memory Testing. *Neurology, 38*(6), 900-903. https://doi.org/10.1212/Wnl.38.6.900
267. Bucks, R. S., & Willison, J. R. (1997). Development and Validation of the Location Learning Test (LLT): A Test of Visuo-Spatial Learning Designed for Use with Older

Adults and in Dementia. *Clinical Neuropsychologist, 11*(3), 273-286. https://doi.org/10.1080/13854049708400456

268. Bramao, I., Reis, A., Petersson, K. M., & Faisca, L. (2011). The Role of Color Information on Object Recognition: A Review and Meta-Analysis. *Acta Psychologica, 138*(1), 244-253. https://doi.org/10.1016/j.actpsy.2011.06.010
269. Ardila, A. (1998). A Note of Caution: Normative Neuropsychological Test Performance: Effects of Age, Education, Gender and Ethnicity: A Comment on Saykin et al. (1995). *Applied Neuropsychology, 5*(1), 51-53. https://doi.org/10.1207/s15324826an0501_8
270. Nielsen, T. R., Andersen, B. B., Kastrup, M., Phung, T. K., & Waldemar, G. (2011). Quality of Dementia Diagnostic Evaluation for Ethnic Minority Patients: A Nationwide Study. *Dementia and Geriatric Cognitive Disorders, 31*(5), 388-396. https://doi.org/10.1159/000327362
271. Hamberger, M. J. (2015). Object Naming in Epilepsy and Epilepsy Surgery. *Epilepsy & Behavior, 46*, 27-33. https://doi.org/10.1016/j.yebeh.2014.12.019
272. Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., . . . Lyrer, P. A. (2006). Epidemiology of Aphasia Attributable to First Ischemic Stroke: Incidence, Severity, Fluency, Etiology, and Thrombolysis. *Stroke, 37*(6), 1379-1384. https://doi.org/10.1161/01.STR.0000221815.64093.8c
273. Satoer, D., Vincent, A., Smits, M., Dirven, C., & Visch-Brink, E. (2013). Spontaneous Speech of Patients with Gliomas in Eloquent Areas before and Early after Surgery. *Acta Neurochirurgica, 155*(4), 685-692. https://doi.org/10.1007/s00701-013-1638-8
274. Grossman, M., McMillan, C., Moore, P., Ding, L., Glosser, G., Work, M., & Gee, J. (2004). What's in a Name: Voxel-Based Morphometric Analyses of MRI and Naming Difficulty in Alzheimer's Disease, Frontotemporal Dementia and Corticobasal Degeneration. *Brain, 127*(Pt 3), 628-649. https://doi.org/10.1093/brain/awh075
275. Gleichgerrcht, E., Fridriksson, J., & Bonilha, L. (2015). Neuroanatomical Foundations of Naming Impairments across Different Neurologic Conditions. *Neurology, 85*(3), 284-292. https://doi.org/10.1212/WNL.0000000000001765
276. 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
277. Maruta, C., Guerreiro, M., de Mendonca, A., Hort, J., & Scheltens, P. (2011). The Use of Neuropsychological Tests across Europe: The Need for a Consensus in the Use of Assessment Tools for Dementia. *European Journal of Neurology, 18*(2), 279-285. https://doi.org/10.1111/j.1468-1331.2010.03134.x
278. Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in Test-Usage Practices of Clinical Neuropsychologists in the United States and Canada over a 10-Year Period: A Follow-up Survey of INS and NAN Members. *Archives of Clinical Neuropsychology, 31*(3), 206-230. https://doi.org/10.1093/arclin/acw007
279. Baird, A. D., Ford, M., & Podell, K. (2007). Ethnic Differences in Functional and Neuropsychological Test Performance in Older Adults. *Archives of Clinical Neuropsychology, 22*(3), 309-318. https://doi.org/10.1016/j.acn.2007.01.005
280. Worrall, L. E., Yiu, E. M. L., Hickson, L. M. H., & Barnett, H. M. (1995). Normative Data for the Boston Naming Test for Australian Elderly. *Aphasiology, 9*(6), 541-551. https://doi.org/10.1080/02687039508248713

281. Roberts, P. M., & Doucet, N. (2011). Performance of French-Speaking Quebec Adults on the Boston Naming Test. *Canadian Journal of Speech-Language Pathology and Audiology, 35*(3), 254–267.
282. Allegri, R. F., Mangone, C. A., Villavicencio, A. F., Rymberg, S., Taragano, F. E., & Baumann, D. (1997). Spanish Boston Naming Test Norms. *Clinical Neuropsychologist, 11*(4), 416-420. https://doi.org/10.1080/13854049708400471
283. Byrd, D. A., Rivera Mindt, M. M., Clark, U. S., Clarke, Y., Thames, A. D., Gammada, E. Z., & Manly, J. J. (2021). Creating an Antiracist Psychology by Addressing Professional Complicity in Psychological Assessment. *Psychological Assessment, 33*(3), 279–285. https://doi.org/10.1037/pas0000993
284. Gasquoine, P. G. (2009). Race-Norming of Neuropsychological Tests. *Neuropsychology Review, 19*(2), 250-262. https://doi.org/10.1007/s11065-009-9090-5
285. Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-Ratings of Spoken Language Dominance: A Multi-Lingual Naming Test (MINT) and Preliminary Norms for Young and Aging Spanish-English Bilinguals. *Bilingualism (Cambridge, England), 15*(3), 594-615. https://doi.org/10.1017/S1366728911000332
286. Swadesh, M. (1952). Lexicostatistic Dating of Prehistoric Ethnic Contacts. *Proceedings of the American Philosophical Society, 96*, 152–163.
287. Vivas, L., Manoiloff, L., Linares, N., Zaionz, A. F., & Montero, L. (2020). Argentinean Psycholinguistic Picture Naming Test in Color. *Applied Neuropsychology-Adult*. https://doi.org/10.1080/23279095.2020.1780238
288. Macoir, J., Beaudoin, C., Bluteau, J., Potvin, O., & Wilson, M. A. (2018). TDQ-60 - a Color Picture-Naming Test for Adults and Elderly People: Validation and Normalization Data. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 25*(5), 753-766. https://doi.org/10.1080/13825585.2017.1372355
289. Snodgrass, J. G., & Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174-215. https://doi.org/10.1037//0278-7393.6.2.174
290. Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-Acquisition Ratings for 30,000 English Words. *Behavior Research Methods, 44*(4), 978-990. https://doi.org/10.3758/s13428-012-0210-4
291. Luniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Andelkovic, D., . . . Unal-Logacev, O. (2016). Ratings of Age of Acquisition of 299 Words across 25 Languages: Is There a Cross-Linguistic Order of Words? *Behavior Research Methods, 48*(3), 1154-1177.
292. Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of Age of Acquisition and Concreteness for 30,000 Dutch Words. *Acta Psychologica, 150*, 80-84. https://doi.org/10.1016/j.actpsy.2014.04.010
293. Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A New Measure for Dutch Word Frequency Based on Film Subtitles. *Behavior Research Methods, 42*(3), 643-650. https://doi.org/10.3758/BRM.42.3.643
294. Alonso, M. A., Fernandez, A., & Diez, E. (2015). Subjective Age-of-Acquisition Norms for 7,039 Spanish Words. *Behavior Research Methods, 47*(1), 268-274. https://doi.org/10.3758/s13428-014-0454-2

295. Cameirao, M. L., & Vicente, S. G. (2010). Age-of-Acquisition Norms for a Set of 1,749 Portuguese Words. *Behavior Research Methods, 42*(2), 474-480. https://doi.org/10.3758/BRM.42.2.474
296. Marques, J. F., Fonseca, F. L., Morais, A. S., & Pinto, I. A. (2007). Estimated Age of Acquisition Norms for 834 Portuguese Nouns and Their Relation with Other Psycholinguistic Variables. *Behavior Research Methods, 39*(3), 439-444. https://doi.org/10.3758/bf03193013
297. Ferrand, L., Bonin, P., Meot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age-of-Acquisition and Subjective Frequency Estimates for All Generally Known Monosyllabic French Words and Their Relation with Other Psycholinguistic Variables. *Behavior Research Methods, 40*(4), 1049-1054.
298. Birchenough, J. M. H., Davies, R., & Connelly, V. (2017). Rated Age-of-Acquisition Norms for over 3,200 German Words. *Behavior Research Methods, 49*(2), 484-501. https://doi.org/10.3758/s13428-016-0718-0
299. Dukes, K. (2009). *The Quranic Arabic Corpus.* In https://corpus.quran.com/
300. Verhage, F. (1964). *Intelligentie en Leeftijd: Onderzoek bij Nederlanders van Twaalf tot Zevenenzeventig Jaar. Proefschrift.* van Gorcum.
301. Caman, O. K., Karahan, S., Unal, F., Bilir, N., Saka, E., Bariskin, E., & Ayhan, Y. (2019). Adaptation of the Modified Mini-Mental State Examination (3MS) and Determination of Its Normative Values in Turkey. *Dementia and Geriatric Cognitive Disorders, 47*(4-6), 315-322. https://doi.org/10.1159/000500939
302. Howard, E., Irwin, D. J., Rascovsky, K., Nevler, N., Shellikeri, S., Tropea, T. F., . . . Cousins, K. A. Q. (2021). Cognitive Profile and Markers of Alzheimer Disease-Type Pathology in Patients with Lewy Body Dementias. *Neurology, 96*(14), e1855-e1864. https://doi.org/10.1212/WNL.0000000000011699
303. Zec, R. F., Markwell, S. J., Burkett, N. R., & Larsen, D. L. (2005). A Longitudinal Study of Confrontation Naming in the "Normal" Elderly. *Journal of the International Neuropsychological Society, 11*(6), 716-726. https://doi.org/10.1017/S1355617705050897
304. Ivanova, I., Salmon, D. P., & Gollan, T. H. (2014). Which Language Declines More? Longitudinal Versus Cross-Sectional Decline of Picture Naming in Bilinguals with Alzheimer's Disease. *Journal of the International Neuropsychological Society, 20*(5), 534-546. https://doi.org/10.1017/S1355617714000228 24725624
305. Hillis, A. E., Oh, S., & Ken, L. (2004). Deterioration of Naming Nouns Versus Verbs in Primary Progressive Aphasia. *Annals of Neurology, 55*(2), 268-275. https://doi.org/10.1002/ana.10812
306. Pisoni, A., Mattavelli, G., Casarotti, A., Comi, A., Riva, M., Bello, L., & Papagno, C. (2018). Object-Action Dissociation: A Voxel-Based Lesion-Symptom Mapping Study on 102 Patients after Glioma Removal. *Neuroimage: Clinical, 18*, 986-995. https://doi.org/10.1016/j.nicl.2018.03.022
307. Eng, N., Vonk, J. M. J., Salzberger, M., & Yoo, N. (2019). A Cross-Linguistic Comparison of Category and Letter Fluency: Mandarin and English. *Quarterly Journal of Experimental Psychology, 72*(3), 651-660. https://doi.org/10.1177/1747021818765997
308. Nielsen, T. R., Nielsen, D. S., & Waldemar, G. (2020). Barriers to Post-Diagnostic Care and Support in Minority Ethnic Communities: A Survey of Danish Primary Care Dementia Coordinators. *Dementia: The International Journal of Social Research and Practice, 19*(8), 2702-2713. https://doi.org/10.1177/1471301219853945

309. NIVEL. (2018). Persoonlijke en Maatschappelijke Gevolgen van de Ziekte voor Mensen met Dementie en hun Mantelzorgers. In *Een Samenhangend Beeld van Dementie en Dementiezorg* (pp. 53-64).
310. Berdai Chaouni, S., & De Donder, L. (2019). Invisible Realities: Caring for Older Moroccan Migrants with Dementia in Belgium. *Dementia: The International Journal of Social Research and Practice, 18*(7-8), 3113-3129. https://doi.org/10.1177/1471301218768923
311. Hossain, M. Z., & Khan, H. T. A. (2020). Barriers to Access and Ways to Improve Dementia Services for a Minority Ethnic Group in England. *Journal of Evaluation in Clinical Practice, 26*(6), 1629-1637. https://doi.org/10.1111/jep.13361
312. Al-Janabi, H., Frew, E., Brouwer, W., Rappange, D., & Van Exel, J. (2010). The Inclusion of Positive Aspects of Caring in the Caregiver Strain Index: Tests of Feasibility and Validity. *International Journal of Nursing Studies, 47*(8), 984-993. https://doi.org/10.1016/j.ijnurstu.2009.12.015
313. Robinson, B. C. (1983). Validation of a Caregiver Strain Index. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 38*(3), 344-348. https://doi.org/10.1093/geronj/38.3.344
314. Campbell, P., Wright, J., Oyebode, J., Job, D., Crome, P., Bentham, P., . . . Lendon, C. (2008). Determinants of Burden in Those Who Care for Someone with Dementia. *International Journal of Geriatric Psychiatry, 23*(10), 1078-1085. https://doi.org/10.1002/gps.2071
315. Li, R., Cooper, C., Bradley, J., Shulman, A., & Livingston, G. (2012). Coping Strategies and Psychological Morbidity in Family Carers of People with Dementia: A Systematic Review and Meta-Analysis. *Journal of Affective Disorders, 139*(1), 1-11. https://doi.org/10.1016/j.jad.2011.05.055
316. Dauphinot, V., Delphin-Combe, F., Mouchoux, C., Dorey, A., Bathsavanis, A., Makaroff, Z., . . . Krolak-Salmon, P. (2015). Risk Factors of Caregiver Burden among Patients with Alzheimer's Disease or Related Disorders: A Cross-Sectional Study. *Journal of Alzheimer's disease: JAD, 44*(3), 907-916. https://doi.org/10.3233/JAD-142337
317. Hirono, N., Kobayashi, H., & Mori, E. (1998). [Caregiver Burden in Dementia: Evaluation with a Japanese Version of the Zarit Caregiver Burden Interview]. *No To Shinkei, 50*(6), 561-567.
318. Terum, T. M., Andersen, J. R., Rongve, A., Aarsland, D., Svendsboe, E. J., & Testad, I. (2017). The Relationship of Specific Items on the Neuropsychiatric Inventory to Caregiver Burden in Dementia: A Systematic Review. *International Journal of Geriatric Psychiatry, 32*(7), 703-717. https://doi.org/10.1002/gps.4704
319. Tsai, C. F., Hwang, W. S., Lee, J. J., Wang, W. F., Huang, L. C., Huang, L. K., . . . Fuh, J. L. (2021). Predictors of Caregiver Burden in Aged Caregivers of Demented Older Patients. *BMC Geriatrics, 21*(1), 59. https://doi.org/10.1186/s12877-021-02007-1
320. Katz, S., Downs, T. D., Cash, H. R., & Grotz, R. G. (1970). Progress in the Development of the Index of ADL. *The Gerontologist, 10*(1), 20-30.
321. Lawton, M. P., & Brody, E. M. (1969). Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *The Gerontologist, 9*(3), 179-186.
322. Kruithof, W. J., Post, M. W., & Visser-Meily, J. M. (2015). Measuring Negative and Positive Caregiving Experiences: A Psychometric Analysis of the Caregiver Strain Index Expanded. *Clinical Rehabilitation, 29*(12), 1224-1233. https://doi.org/10.1177/0269215515570378

323. Tzeng, N.-S., Chang, C.-W., Hsu, J.-Y., Chou, Y.-C., Chang, H.-A., & Kao, Y.-C. (2015). Caregiver Burden for Patients with Dementia with or without Hiring Foreign Health Aides: A Cross-Sectional Study in a Northern Taiwan Memory Clinic. *Journal of Medical Sciences, 35*(6), 239-247. https://doi.org/10.4103/1011-4564.172999
324. Pendergrass, A., Malnis, C., Graf, U., Engel, S., & Graessel, E. (2018). Screening for Caregivers at Risk: Extended Validation of the Short Version of the Burden Scale for Family Caregivers (BSFC-S) with a Valid Classification System for Caregivers Caring for an Older Person at Home. *Bmc Health Services Research, 18*. https://doi.org/10.1186/s12913-018-3047-4
325. Zahir, A., Staffaroni, A. M., Wickham, R. E., Quinn, C. M., Sapozhnikova, A., Seidman, J., & Chiong, W. (2021). Caregiver "Objective Attitude" toward Patients with Neurodegenerative Disease: Consequences for Caregiver Strain and Relationship Closeness. *Aging and Mental Health, 25*(9), 1709-1715. https://doi.org/10.1080/13607863.2020.1771541
326. Hernandez-Padilla, J. M., Ruiz-Fernandez, M. D., Granero-Molina, J., Ortiz-Amo, R., Rodriguez, M. M. L., & Fernandez-Sola, C. (2021). Perceived Health, Caregiver Overload and Perceived Social Support in Family Caregivers of Patients with Alzheimer's: Gender Differences. *Health & Social Care in the Community, 29*(4), 1001-1009. https://doi.org/10.1111/hsc.13134
327. Ahmad, M., van den Broeke, J., Saharso, S., & Tonkens, E. (2020). Persons with a Migration Background Caring for a Family Member with Dementia: Challenges to Shared Care. *Gerontologist, 60*(2), 340-349. https://doi.org/10.1093/geront/gnz161
328. Etters, L., Goodall, D., & Harrison, B. E. (2008). Caregiver Burden among Dementia Patient Caregivers: A Review of the Literature. *Journal of the American Academy of Nurse Practitioners, 20*(8), 423-428. https://doi.org/10.1111/j.1745-7599.2008.00342.x
329. Annerstedt, L., Elmstahl, S., Ingvad, B., & Samuelsson, S. M. (2000). Family Caregiving in Dementia - an Analysis of the Caregiver's Burden and the "Breaking-Point" When Home Care Becomes Inadequate. *Scandinavian Journal of Public Health, 28*(1), 23-31. https://doi.org/10.1177/14034948000280010 6
330. Smith, K. J., George, C., & Ferreira, N. (2018). Factors Emerging from the "Zarit Burden Interview" and Predictive Variables in a UK Sample of Caregivers for People with Dementia. *International Psychogeriatrics, 30*, 1671–1678.
331. Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive Assessment of Psychopathology in Dementia. *Neurology, 44*(12), 2308-2314. https://doi.org/10.1212/wnl.44.12.2308
332. Franzen, S., van den Berg, E., & Papma, J. M. (2018). Neuropsychologische Diagnostiek bij niet-Westerse Migranten [Neuropsychological Assessment of Non-Western Immigrants]. *Tijdschrift voor Neuropsychologie, 13*(1), 59-67.
333. Alzheimercentrum Erasmus MC. (2019). Heeft Yasmina Dementie of niet? Naar De Geheugenpoli in Het Ziekenhuis. https://www.erasmusmc.nl/-/media/erasmusmc/pdf/1-themaspecifiek/alzheimercentrum/dwd-heeft-yasmina-dementie-of-niet-v3.pdf?la=nl-nl
334. Rohlof, H., Loevy, N., Sassen, L., & Helmich, S. (2002). The Cultural Interview. In R. Borra, R. van Dijk, & H. Rohlof (Eds.), *Cultuur, Classificatie En Diagnose: Cultuursensitief Werken Met De Dms-Iv* (pp. 251-260). Bohn, Stafleu van Loghum.

335. Groen, S., & van Dijk, R. (2020). De Culturele Formulering: Een Contextuele Benadering van Psychische Klachten. In J. de Jong & R. van Dijk (Eds.), *Handboek Culturele Psychiatrie En Psychotherapie*. Boom/de Tijdstroom.
336. Pharos. (2020). De Terugvraagmethode. https://www.pharos.nl/infosheets/laaggeletterdheid-en-beperkte-gezondheidsvaardigheden-de-terugvraagmethode/
337. Generieke Module Diversiteit. (2020). https://www.ggzstandaarden.nl/uploads/pdf/project/project_751c37f0-15af-4c4a-9cf9-705bdd72ba2f_diversiteit__authorized-at_27-09-2018.pdf
338. United Nations High Commissioner for Refugees. (2001). Asylum Applications in Industrialized Countries: 1980-1999. https://www.unhcr.org/3c3eb40f4.pdf
339. European Commission/EACEA/Eurydice. (2017). *Key Data on Teaching Languages at School in Europe – 2017 Edition Eurydice Report*. Publications Office of the European Union.
340. Spanish Const. art. 3. (1978).
341. Farkas, L. (2017). *Data Collection in the Field of Ethnicity: Analysis and Comparative Review of Equality Data Collection Practices in the European Union*. Publications Office of the European Union.
342. Ganguli, M., Chandra, V., Gilby, J. E., Ratcliff, G., Sharma, S. D., Pandav, R., . . . Belle, S. (1996). Cognitive Test Performance in a Community-Based Nondemented Elderly Sample in Rural India: The Indo-U.S. Cross-National Dementia Epidemiology Study. *International Psychogeriatrics, 8*(4), 507-524. https://doi.org/10.1017/s1041610296002852
343. Agranovich, A. V., & Puente, A. E. (2007). Do Russian and American Normal Adults Perform Similarly on Neuropsychological Tests? Preliminary Findings on the Relationship between Culture and Test Performance. *Archives of Clinical Neuropsychology, 22*(3), 273-282. https://doi.org/10.1016/j.acn.2007.01.003
344. Lozano-Ruiz, A., Fasfous, A. F., Ibanez-Casas, I., Cruz-Quintana, F., Perez-Garcia, M., & Pérez-Marfil, M. N. (2021). Cultural Bias in Intelligence Assessment Using a Culture-Free Test in Moroccan Children. *Archives of Clinical Neuropsychology*. https://doi.org/10.1093/arclin/acab005
345. Ibanez-Casas, I., Leonard, B. E., Pérez-García, M., & Puente, A. E. (in press). Development of a Computerized Battery for Cross-Cultural Neuropsychological Assessment: The EMBRACED Project. *Bethlehem University Journal*.
346. Fasfous, A. F., Peralta-Ramirez, M. I., Perez-Marfil, M. N., Cruz-Quintana, F., Catena-Martinez, A., & Perez-Garcia, M. (2015). Reliability and Validity of the Arabic Version of the Computerized Battery for Neuropsychological Evaluation of Children (BENCI). *Child Neuropsychology, 21*(2), 210-224. https://doi.org/10.1080/09297049.2014.896330
347. Nielsen, T. R., Vogel, A., & Waldemar, G. (2012). Comparison of Performance on Three Neuropsychological Tests in Healthy Turkish Immigrants and Danish Elderly. *International Psychogeriatrics, 24*(9), 1515-1521. https://doi.org/10.1017/S1041610212000440
348. Watermeyer, T., & Calia, C. (2019). Neuropsychological Assessment in Preclinical and Prodromal Alzheimer Disease: A Global Perspective. *Journal of Global Health, 9*(1), 010317. https://doi.org/10.7189/jogh.09.010317
349. Scior, K., Gray, J., Halsey, R., & Roth, A. (2007). Selection for Clinical Psychology Training: Is There Evidence of Any Bias Towards Applicants from Ethnic Minorities. *Clinical Psychology Forum, 175*, 7-11.

350. Health and Social Care Information Centre. (2013). NHS Workforce, Summary of Staff in the NHS: Results from September 2012 Census. https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics-overview/nhs-workforce-summary-of-staff-in-the-nhs-results-from-september-2012-census
351. Birk Jensen, I., & Kolman, L. (2018). Meaning of the EFPA Task Force Cultural and Ethnic Diversity for Psychologists in Europe. In A. Thomas (Ed.), *Cultural and Ethnic Diversity: How European Psychologists Can Meet the Challenges* (pp. 5-12). Hogrefe.
352. van Oosteren, C., Booi, H., Broekhuizen, J., Cohen, L., de Jong, I., Lindeman, E., . . . van Vliet, M. (2013). Structurele Ontwikkeling Amsterdam. In *Een Stad voor Iedereen* (pp. 13-30). Bureau Onderzoek en Statistiek Gemeente Amsterdam.
353. CIBG. (2021). Basisberoep en Specialisme - Specialismen Per 1 Juni 2021. In. Ministerie van Volksgezondheid, Welzijn en Sport. https://www.bigregister.nl/over-het-big-register/cijfers/basisberoep-en-specialisme
354. Byrd, D., Razani, J., Suarez, P., Lafosse, J. M., Manly, J., & Attix, D. K. (2010). Diversity Summit 2008: Challenges in the Recruitment and Retention of Ethnic Minorities in Neuropsychology. *The Clinical Neuropsychologist, 24*(8), 1279-1291. https://doi.org/10.1080/13854046.2010.521769
355. Araujo, N. B., Nielsen, T. R., Barca, M. L., Engedal, K., Marinho, V., Deslandes, A. C., . . . Laks, J. (2020). Brazilian Version of the European Cross-Cultural Neuropsychological Test Battery (CNTB-BR): Diagnostic Accuracy across Schooling Levels. *Brazilian Journal of Psychiatry, 42*(3), 286-294. https://doi.org/10.1590/1516-4446-2019-0539
356. British Psychological Society. (2017). BPS Practice Guidelines. https://www.bps.org.uk/news-and-policy/practice-guidelines
357. Dansk Psykolog Forening. (2001). Retningslinjer for Neuropsykologiske Undersøgelser – Specielt Med Hensyn Til Test. https://www.dp.dk/neuropsykologiske-undersogelser/
358. Nederlands Instituut voor Psychologen. (2015). *Beroepscode voor Psychologen 2015*. Retrieved 14-06-2021 from https://www.psynip.nl/en/dutch-association-psychologists/code-of-ethics/code-ethics-2015/respect-general/
359. Alzheimer's Association. (2015). 2015 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 11*, 324-384. https://doi.org/10.1016/j.jalz.2015.02.003
360. Barnes, L. L., & Bennett, D. A. (2014). Alzheimer's Disease in African Americans: Risk Factors and Challenges for the Future. *Health Affairs, 33*(4), 580-586.
361. Canevelli, M., Bruno, G., Grande, G., Quarata, F., Raganato, R., Remiddi, F., . . . Cesari, M. (2019). Race Reporting and Disparities in Clinical Trials on Alzheimer's Disease: A Systematic Review. *Neuroscience & Biobehavioral Reviews, 101*, 122-128. https://doi.org/10.1016/j.neubiorev.2019.03.020
362. Faison, W. E., Schultz, S. K., Aerssens, J., Alvidrez, J., Anand, R., Farrer, L. A., . . . Mintzer, J. E. (2007). Potential Ethnic Modifiers in the Assessment and Treatment of Alzheimer's Disease: Challenges for the Future. *International Psychogeriatrics, 19*(3), 539-558. https://doi.org/10.1017/S104161020700511X
363. Shin, J., & Doraiswamy, P. M. (2016). Underrepresentation of African-Americans in Alzheimer's Trials: A Call for Affirmative Action. *Frontiers in Aging Neuroscience, 8*, 123.
364. Bjornsson, T. D., Wagner, J. A., Donahue, S. R., Harper, D., Karim, A., Khouri, M. S., . . . Loew, C. (2003). A Review and Assessment of Potential Sources of Ethnic Differences in Drug Responsiveness. *Journal of Clinical Pharmacology, 43*, 943-967.

365. Goldstein, D. B., Tate, S. K., & Sisodiya, S. M. (2003). Pharmacogenetics Goes Genomic. *Nature Reviews Genetics, 4*(12), 937-947.
366. Garza, M. A., Quinn, S. C., Li, Y., Assini-Meytin, L., Casper, E. T., Fryer, C. S., . . . Thomas, S. B. (2017). The Influence of Race and Ethnicity on Becoming a Human Subject: Factors Associated with Participation in Research. *Contemporary Clinical Trials Communications, 7*, 57-63. https://doi.org/10.1016/j.conctc.2017.05.009
367. Wendler, D., Kington, R., Madans, J., Van Wye, G., Christ-Schmidt, H., Pratt, L. A., . . . Emanuel, E. (2006). Are Racial and Ethnic Minorities Less Willing to Participate in Health Research? *PLoS Med, 3*(2), e19. https://doi.org/10.1371/journal.pmed.0030019
368. Cooper, C., Ketley, D., & Livingston, G. (2014). Systematic Review and Meta-Analysis to Estimate Potential Recruitment to Dementia Intervention Studies. *International Journal of Geriatric Psychiatry, 29*(5), 515-525. https://doi.org/10.1002/gps.4034
369. Schneider, L. S., Olin, J. T., Lyness, S. A., & Chui, H. C. (1997). Eligibility of Alzheimer's Disease Clinic Patients for Clinical Trials. *Journal of the American Geriatrics Society, 45*(8), 923-928. https://doi.org/10.1111/j.1532-5415.1997.tb02960.x
370. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *The British Medical Journal, 339*, b2535.
371. Food and Drug Administration. (2020). *Early Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industry*.
372. Galimberti, D., & Scarpini, E. (2011). Disease-Modifying Treatments for Alzheimer's Disease. *Therapeutic Advances in Neurological Disorders, 4*(4), 203-216.
373. National Center for Health Statistics. (2019). Tables of Summary Health Statistics for U.S. Adults: 2018 National Health Interview Survey. http://www.cdc.gov/nchs/nhis/SHS/tables.htm
374. Australian Institute of Health and Welfare. (2015). *The Health and Welfare of Australia's Aboriginal and Torres Strait Islander Peoples. Cat. No. Ihw 147*. AIHW. https://www.aihw.gov.au/getmedia/584073f7-041e-4818-9419-39f5a060b1aa/18175.pdf.aspx?inline=true
375. Pratt, L. A., & Brody, D. J. (2014). *Depression in the U.S. Household Population, 2009–2012. NCHS Data Brief, No 172.* National Center for Health Statistics.
376. Substance Abuse and Mental Health Services Administration. (2014). *Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings, NSDUH Series H-48, HHS Publication No. (SMA) 14-4863.*
377. Centers for Disease Control and Prevention. (2020). *HIV Surveillance Report, 2018 (Updated)*. http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html
378. Center for Disease Control and Prevention. (2019). *Health, United States, 2018*. National Center for Health Statistics.
379. Randolph, C. (2009). *RBANS Update: Repeatable Battery for the Assessment of Neuropsychological Status: Manual*. Pearson.
380. Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A New Rating Scale for Alzheimer's Disease. *The American Journal of Psychiatry, 141*(11), 1356–1364. https://doi.org/10.1176/ajp.141.11.1356
381. Buschke, H. (2002). Free and Cued Selective Reminding Test. *New England Journal of Medicine, 347*(22), 1761-1768.
382. Wechsler, D. (1987). *Manual for the Wechsler Memory Scale-Revised*. The Psychological Corporation.

383. Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23*, 56–62.
384. Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., . . . Mann, J. J. (2011). The Columbia-Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings from Three Multisite Studies with Adolescents and Adults. *The American Journal of Psychiatry, 168*(12), 1266-1277. https://doi.org/10.1176/appi.ajp.2011.10111704
385. Whitfield, K. E., Allaire, J. C., Belue, R., & Edwards, C. L. (2008). Are Comparisons the Answer to Understanding Behavioral Aspects of Aging in Racial and Ethnic Groups? *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 63*(5), P301-308. https://doi.org/10.1093/geronb/63.5.p301
386. Kim, E. S., Bruinooge, S. S., Roberts, S., Ison, G., Lin, N. U., Gore, L., . . . Schilsky, R. L. (2017). Broadening Eligibility Criteria to Make Clinical Trials More Representative: American Society of Clinical Oncology and Friends of Cancer Research Joint Research Statement. *Journal of Clinical Oncology, 35*(33), 3737-3744.
387. Levey, A. S., Bosch, J. P., Lewis, J. B., Greene, T., Rogers, N., & Roth, D. (1999). A More Accurate Method to Estimate Glomerular Filtration Rate from Serum Creatinine: A New Prediction Equation. Modification of Diet in Renal Disease Study Group. *Annals of Internal Medicine, 130*(6), 461-470. https://doi.org/10.7326/0003-4819-130-6-199903160-00002
388. Lim, E., Miyamura, J., & Chen, J. J. (2015). Racial/Ethnic-Specific Reference Intervals for Common Laboratory Tests: A Comparison among Asians, Blacks, Hispanics, and White. *Hawai'i Journal of Medicine & Public Health : A Journal of Asia Pacific Medicine & Public Health, 74*(9), 302-310.
389. Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *New England Journal of Medicine, 383*(9), 874-882. https://doi.org/10.1056/NEJMms2004740
390. Hussain-Gambles, M., Atkin, K., & Leese, B. (2004). Why Ethnic Minority Groups Are under-Represented in Clinical Trials: A Review of the Literature. *Health and Social Care in the Community, 12*(5), 382–388.
391. Cooper, L. A., Roter, D. L., Carson, K. A., Beach, M. C., Sabin, J. A., Greenwald, A. G., & Inui, T. S. (2012). The Associations of Clinicians' Implicit Attitudes About Race with Medical Visit Communication and Patient Ratings of Interpersonal Care. *American Public Health Association, 102*(5), 979-987. https://doi.org/10.2105/AJPH.2011.300558
392. Llanque, S. M., & Enriquez, M. (2012). Interventions for Hispanic Caregivers of Patients with Dementia: A Review of the Literature. *American Journal of Alzheimer's Disease and Other Dementias, 27*(1), 23-32. https://doi.org/10.1177/1533317512439794
393. Largent, E. A., Karlawish, J., & Grill, J. D. (2018). Study Partners: Essential Collaborators in Discovering Treatments for Alzheimer's Disease. *Alzheimer's Research & Therapy, 10*(1), 101. https://doi.org/10.1186/s13195-018-0425-4
394. Grill, J. D., Monsell, S., & Karlawish, J. (2012). Are Patients Whose Study Partners Are Spouses More Likely to Be Eligible for Alzheimer's Disease Clinical Trials? *Dementia and Geriatric Cognitive Disorders, 33*(5), 334-340. https://doi.org/10.1159/000339361
395. UNESCO. (n.d.). *Education for All 2000-2015: Achievements and Challenges*. Retrieved 22-10-2018 from http://unesdoc.unesco.org/images/0023/002322/232205e.pdf

396. Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., & Dunleavy, E. (2007). *Literacy in Everyday Life: Results from the 2003 National Assessment of Adult Literacy (NCES 2007–480)*. National Center for Education Statistics.
397. Food and Drug Administration. (1998). *A Guide to Informed Consent: Guidance for Institutional Review Board and Clinical Investigators.* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guide-informed-consent
398. Fanaroff, A. C., Li, S., Webb, L. E., Miller, V., Navar, A. M., Peterson, E. D., & Wang, T. Y. (2018). An Observational Study of the Association of Video- Versus Text-Based Informed Consent with Multicenter Trial Enrollment Lessons from the Palm Study (Patient and Provider Assessment of Lipid Management). *Circulation-Cardiovascular Quality and Outcomes, 11*(4). https://doi.org/10.1161/CIRCOUTCOMES.118.004675
399. Parikh, P. M., Prabhash, K., Govind, K. B., Digumarti, R., Pandit, S., Banerjee, I., . . . Gupta, S. (2014). Standard Operating Procedure for Audio Visual Recording of Informed Consent: An Initiative to Facilitate Regulatory Compliance. *Indian Journal of Cancer, 51*(2), 113-116. https://doi.org/10.4103/0019-509X.138158
400. Steis, M. R., & Schrauf, R. W. (2009). A Review of Translations and Adaptations of the Mini-Mental State Examination in Languages Other Than English and Spanish. *Research in Gerontological Nursing, 2*(3), 214-224. https://doi.org/10.3928/19404921-20090421-06
401. Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A Comprehensive Review. *Journal of the American Geriatrics Society, 40*(9), 922-935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x
402. Murden, R. A., McRae, T. D., Kaner, S., & Bucknam, M. E. (1991). Mini-Mental State Exam Scores Vary with Education in Blacks and Whites. *Journal of the American Geriatrics Society, 39*(2), 149-155. https://doi.org/10.1111/j.1532-5415.1991.tb01617.x
403. Zwart, L. A. R., Goudsmit, M., van Campen, J. P. C. M., Rijkers, C. J. M., & Wind, A. W. (2015). Using the MMSE as a Cognitive Screener among Turkish and Moroccan Migrants. *Tijdschrift voor Gerontologie en Geriatrie, 46*(1), 28-36. https://doi.org/10.1007/s12439-014-0105-1
404. Lim, W. S., Chong, M. S., & Sahadevan, S. (2007). Utility of the Clinical Dementia Rating in Asian Populations. *Clinical Medicine & Research, 5*(1), 61-70. https://doi.org/10.3121/cmr.2007.693
405. Food and Drug Administration. (2019). *Enhancing the Diversity of Clinical Trial Populations - Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry.* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial
406. Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *Journal of Pharmacology & Pharmacotherapeutics, 1*(2), 100-107. https://doi.org/10.4103/0976-500X.72352
407. Clark, L. T., Watkins, L., Pina, I. L., Elmer, M., Akinboboye, O., Gorham, M., . . . Regnante, J. M. (2019). Increasing Diversity in Clinical Trials: Overcoming Critical Barriers. *Current Problems in Cardiology, 44*(5), 148-172. https://doi.org/10.1016/j.cpcardiol.2018.11.002
408. Gauthier, M. A., & Clarke, W. P. (1999). Gaining and Sustaining Minority Participation in Longitudinal Research Projects. *Alzheimer Disease and Associated Disorders, 13*, S29-33.

409. Shavers, V. L., Lynch, C. F., & Burmeister, L. F. (2002). Racial Differences in Factors That Influence the Willingness to Participate in Medical Research Studies. *Annals of Epidemiology, 12*(4), 248-256.
410. Pediatric Research Equity Act, Pub. L. No. 108–155, 117 Stat. 1936, (2003).
411. Zahodne, L. B., Sharifian, N., Kraal, A. Z., Zaheed, A. B., Sol, K., Morris, E. P., . . . Brickman, A. M. (2021). Socioeconomic and Psychosocial Mechanisms Underlying Racial/Ethnic Disparities in Cognition among Older Adults. *Neuropsychology, 35*(3), 265-275. https://doi.org/10.1037/neu0000720
412. Llibre-Guerra, J. J., Li, Y., Allen, I. E., Llibre-Guerra, J. C., Rodriguez Salgado, A. M., Penalver, A. I., . . . Llibre-Rodriguez, J. J. (2021). Race, Genetic Admixture and Cognitive Performance in the Cuban Population. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*. https://doi.org/10.1093/gerona/glab063
413. Resende, E. D. F., Guerra, J. J. L., & Miller, B. L. (2019). Health and Socioeconomic Inequities as Contributors to Brain Health. *Jama Neurology, 76*(6), 633-634. https://doi.org/10.1001/jamaneurol.2019.0362
414. Canevelli, M., Bruno, G., Vico, C., Zaccaria, V., Lacorte, E., Iavicoli, I., . . . Cesari, M. (2018). Socioeconomic Disparities in Clinical Trials on Alzheimer's Disease: A Systematic Review. *European Journal of Neurology, 25*(4), 626-e643.
415. Mukadam, N., Cooper, C., & Livingston, G. (2013). Improving Access to Dementia Services for People from Minority Ethnic Groups. *Current Opinion in Psychiatry, 26*(4), 409-414. https://doi.org/10.1097/YCO.0b013e32835ee668
416. National Research Council Panel on Race, Ethnicity, and Health in Later Life. (2004). *Understanding Racial and Ethnic Differences in Health in Late Life: A Research Agenda*. (R. A. Bulatao & N. B. Anderson, Eds. 2010/07/30 ed.). National Academies Press (US). https://doi.org/10.17226/11036
417. Khunti, K., Kumar, S., & Brodie, J. (2009). *Diabetes UK and South Asian Health Foundation Recommendations on Diabetes Research Priorities for British South Asians*. Diabetes UK.
418. Chaturvedi, N. (2003). Ethnic Differences in Cardiovascular Disease. *Heart, 89*(6), 681-686. https://doi.org/10.1136/heart.89.6.681
419. Adjei, D. N., Stronks, K., Adu, D., Snijder, M. B., Modesti, P. A., Peters, R. J. G., . . . Agyemang, C. (2017). Relationship between Educational and Occupational Levels, and Chronic Kidney Disease in a Multi-Ethnic Sample - the HELIUS Study. *PLoS One, 12*(11), e0186460. https://doi.org/10.1371/journal.pone.0186460
420. Fischbacher, C. M., Bhopal, R., Rutter, M. K., Unwin, N. C., Marshall, S. M., White, M., & Alberti, K. G. (2003). Microalbuminuria Is More Frequent in South Asian Than in European Origin Populations: A Comparative Study in Newcastle, UK. *Diabetic Medicine, 20*(1), 31-36. https://doi.org/10.1046/j.1464-5491.2003.00822.x
421. Kivipelto, M., Mangialasche, F., Snyder, H. M., Allegri, R., Andrieu, S., Arai, H., . . . Carrillo, M. C. (2020). World-Wide FINGERS Network: A Global Approach to Risk Reduction and Prevention of Dementia. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 16*(7), 1078-1094. https://doi.org/10.1002/alz.12123
422. Ting, S. K. S., Foo, H., Chia, P. S., Hameed, S., Ng, K. P., Ng, A., & Kandiah, N. (2018). Dyslexic Characteristics of Chinese-Speaking Semantic Variant of Primary Progressive Aphasia. *The Journal of Neuropsychiatry and Clinical Neurosciences, 30*(1), 31-37. https://doi.org/10.1176/appi.neuropsych.17040081

423. National Institute of Health. (2001). NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research.
424. Alzheimer Europe. (2019). *Overcoming Ethical Challenges Affecting the Involvement of People with Dementia in Research: Recognising Diversity and Promoting Inclusive Research*.
425. Schneeberger, A., Hendrix, S., Mandler, M., Ellison, N., Burger, V., Brunner, M., . . . Dubois, B. (2015). Results from a Phase II Study to Assess the Clinical and Immunological Activity of AFFITOPE® AD02 in Patients with Early Alzheimer's Disease. *The Journal of Prevention of Alzheimer's Disease, 2*(2), 103-114. https://doi.org/10.14283/jpad.2015.63
426. Farlow, M. R., Andreasen, N., Riviere, M. E., Vostiar, I., Vitaliti, A., Sovago, J., . . . Graf, A. (2015). Long-Term Treatment with Active Abeta Immunotherapy with CAD106 in Mild Alzheimer's Disease. *Alzheimer's Research & Therapy, 7*(1), 23. https://doi.org/10.1186/s13195-015-0108-3
427. Vandenberghe, R., Riviere, M. E., Caputo, A., Sovago, J., Maguire, R. P., Farlow, M., . . . Graf, A. (2017). Active Abeta Immunotherapy CAD106 in Alzheimer's Disease: A Phase 2b Study. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 3*(1), 10-22. https://doi.org/10.1016/j.trci.2016.12.003
428. Gilman, S., Koller, M., Black, R. S., Jenkins, L., Griffith, S. G., Fox, N. C., . . . AN1792(QS-21)-201 Study Team. (2005). Clinical Effects of Abeta Immunization (AN1792) in Patients with AD in an Interrupted Trial. *Neurology, 64*(9), 1553-1562. https://doi.org/10.1212/01.WNL.0000159740.16984.3C
429. Coric, V., van Dyck, C. H., Salloway, S., Andreasen, N., Brody, M., Richter, R. W., . . . Berman, R. M. (2012). Safety and Tolerability of the Gamma-Secretase Inhibitor Avagacestat in a Phase 2 Study of Mild to Moderate Alzheimer Disease. *Archives of Neurology, 69*(11), 1430-1440. https://doi.org/10.1001/archneurol.2012.2194
430. Coric, V., Salloway, S., van Dyck, C. H., Dubois, B., Andreasen, N., Brody, M., . . . Berman, R. M. (2015). Targeting Prodromal Alzheimer Disease with Avagacestat: A Randomized Clinical Trial. *Jama Neurology, 72*(11), 1324-1333. https://doi.org/10.1001/jamaneurol.2015.0607
431. Galasko, D., Bell, J., Mancuso, J. Y., Kupiec, J. W., Sabbagh, M. N., van Dyck, C., . . . for the Alzheimer's Disease Cooperative Study. (2014). Clinical Trial of an Inhibitor of RAGE-Abeta Interactions in Alzheimer Disease. *Neurology, 82*(17), 1536-1542. https://doi.org/10.1212/WNL.0000000000000364
432. Rinne, J. O., Brooks, D. J., Rossor, M. N., Fox, N. C., Bullock, R., Klunk, W. E., . . . Grundman, M. (2010). 11C-PIB PET Assessment of Change in Fibrillar Amyloid-Beta Load in Patients with Alzheimer's Disease Treated with Bapineuzumab: A Phase 2, Double-Blind, Placebo-Controlled, Ascending-Dose Study. *Lancet Neurology, 9*(4), 363-372. https://doi.org/10.1016/S1474-4422(10)70043-0
433. Salloway, S., Sperling, R., Gilman, S., Fox, N. C., Blennow, K., Raskind, M., . . . Bapineuzumab 201 Clinical Trial Investigators. (2009). A Phase 2 Multiple Ascending Dose Trial of Bapineuzumab in Mild to Moderate Alzheimer Disease. *Neurology, 73*(24), 2061-2070. https://doi.org/10.1212/WNL.0b013e3181c67808
434. Salloway, S., Sperling, R., Fox, N. C., Blennow, K., Klunk, W., Raskind, M., . . . Bapineuzumab Clinical Trial Investigators. (2014). Two Phase 3 Trials of Bapineuzumab in Mild-to-Moderate Alzheimer's Disease. *New England Journal of Medicine, 370*(4), 322-333. https://doi.org/10.1056/NEJMoa1304839

435. Vandenberghe, R., Rinne, J. O., Boada, M., Katayama, S., Scheltens, P., Vellas, B., . . . Bapineuzumab Clinical Study Investigators. (2016). Bapineuzumab for Mild to Moderate Alzheimer's Disease in Two Global, Randomized, Phase 3 Trials. *Alzheimer's Research & Therapy, 8*(1), 18. https://doi.org/10.1186/s13195-016-0189-7
436. Brody, M., Liu, E., Di, J., Lu, M., Margolin, R. A., Werth, J. L., . . . Novak, G. (2016). A Phase II, Randomized, Double-Blind, Placebo-Controlled Study of Safety, Pharmacokinetics, and Biomarker Results of Subcutaneous Bapineuzumab in Patients with Mild to Moderate Alzheimer's Disease. *Journal of Alzheimer's disease: JAD, 54*(4), 1509-1519. https://doi.org/10.3233/JAD-160369
437. Nelson, T. J., Sun, M. K., Lim, C., Sen, A., Khan, T., Chirila, F. V., & Alkon, D. L. (2017). Bryostatin Effects on Cognitive Function and Pkcvarepsilon in Alzheimer's Disease Phase IIa and Expanded Access Trials. *Journal of Alzheimer's disease: JAD, 58*(2), 521-535. https://doi.org/10.3233/JAD-170161
438. Farlow, M. R., Thompson, R. E., Wei, L. J., Tuchman, A. J., Grenier, E., Crockford, D., . . . Alkon, D. L. (2019). A Randomized, Double-Blind, Placebo-Controlled, Phase II Study Assessing Safety, Tolerability, and Efficacy of Bryostatin in the Treatment of Moderately Severe to Severe Alzheimer's Disease. *Journal of Alzheimer's disease: JAD, 67*(2), 555-570. https://doi.org/10.3233/JAD-180759
439. Bilikiewicz, A., & Gaus, W. (2004). Colostrinin (a Naturally Occurring, Proline-Rich, Polypeptide Mixture) in the Treatment of Alzheimer's Disease. *Journal of Alzheimers Disease, 6*(1), 17-26.
440. Salloway, S., Honigberg, L. A., Cho, W., Ward, M., Friesenhahn, M., Brunstein, F., . . . Paul, R. (2018). Amyloid Positron Emission Tomography and Cerebrospinal Fluid Results from a Crenezumab Anti-Amyloid-Beta Antibody Double-Blind, Placebo-Controlled, Randomized Phase II Study in Mild-to-Moderate Alzheimer's Disease (BLAZE). *Alzheimer's Research & Therapy, 10*(1), 96. https://doi.org/10.1186/s13195-018-0424-5
441. Cummings, J. L., Cohen, S., van Dyck, C. H., Brody, M., Curtis, C., Cho, W., . . . Paul, R. (2018). ABBY: A Phase 2 Randomized Trial of Crenezumab in Mild to Moderate Alzheimer Disease. *Neurology, 90*(21), e1889-e1897. https://doi.org/10.1212/WNL.0000000000005550
442. Salloway, S., Sperling, R., Keren, R., Porsteinsson, A. P., van Dyck, C. H., Tariot, P. N., . . . ELND AD Investigators. (2011). A Phase 2 Randomized Trial of ELND005, Scyllo-Inositol, in Mild to Moderate Alzheimer Disease. *Neurology, 77*(13), 1253-1262.
443. Ostrowitzki, S., Lasser, R. A., Dorflinger, E., Scheltens, P., Barkhof, F., Nikolcheva, T., . . . SCarlet RoAD Investigators. (2017). A Phase III Randomized Trial of Gantenerumab in Prodromal Alzheimer's Disease. *Alzheimer's Research & Therapy, 9*(1), 95. https://doi.org/10.1186/s13195-017-0318-y
444. Relkin, N. R., Thomas, R. G., Rissman, R. A., Brewer, J. B., Rafii, M. S., van Dyck, C. H., . . . for the Alzheimer's Disease Cooperative Study. (2017). A Phase 3 Trial of IV Immunoglobulin for Alzheimer Disease. *Neurology, 88*(18), 1768-1775. https://doi.org/10.1212/WNL.0000000000003904
445. Wischik, C. M., Staff, R. T., Wischik, D. J., Bentham, P., Murray, A. D., Storey, J. M., . . . Harrington, C. R. (2015). Tau Aggregation Inhibitor Therapy: An Exploratory Phase 2 Study in Mild or Moderate Alzheimer's Disease. *Journal of Alzheimer's disease: JAD, 44*(2), 705-720. https://doi.org/10.3233/JAD-142874

446. Gauthier, S., Feldman, H. H., Schneider, L. S., Wilcock, G. K., Frisoni, G. B., Hardlund, J. H., . . . Wischik, C. M. (2016). Efficacy and Safety of Tau-Aggregation Inhibitor Therapy in Patients with Mild or Moderate Alzheimer's Disease: A Randomised, Controlled, Double-Blind, Parallel-Arm, Phase 3 Trial. *Lancet, 388*(10062), 2873-2884. https://doi.org/10.1016/S0140-6736(16)31275-2
447. Wilcock, G. K., Gauthier, S., Frisoni, G. B., Jia, J., Hardlund, J. H., Moebius, H. J., . . . Wischik, C. M. (2018). Potential of Low Dose Leuco-Methylthioninium Bis(Hydromethanesulphonate) (LMTM) Monotherapy for Treatment of Mild Alzheimer's Disease: Cohort Analysis as Modified Primary Outcome in a Phase III Clinical Trial. *Journal of Alzheimer's disease: JAD, 61*(1), 435-457.
448. Lannfelt, L., Blennow, K., Zetterberg, H., Batsman, S., Ames, D., Harrison, J., . . . PBT2-201-EURO study group. (2008). Safety, Efficacy, and Biomarker Findings of PBT2 in Targeting Abeta as a Modifying Therapy for Alzheimer's Disease: A Phase IIa, Double-Blind, Randomised, Placebo-Controlled Trial. *Lancet Neurology, 7*(9), 779-786. https://doi.org/10.1016/S1474-4422(08)70167-4
449. Villemagne, V. L., Rowe, C. C., Barnham, K. J., Cherny, R., Woodward, M., Bozinosvski, S., . . . Masters, C. L. (2017). A Randomized, Exploratory Molecular Imaging Study Targeting Amyloid Beta with a Novel 8-OH Quinoline in Alzheimer's Disease: The PBT2-204 IMAGINE Study. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 3*(4), 622-635. https://doi.org/10.1016/j.trci.2017.10.001
450. Landen, J. W., Cohen, S., Billing, C. B., Jr., Cronenberger, C., Styren, S., Burstein, A. H., . . . Binneman, B. (2017). Multiple-Dose Ponezumab for Mild-to-Moderate Alzheimer's Disease: Safety and Efficacy. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 3*(3), 339-347. https://doi.org/10.1016/j.trci.2017.04.003
451. Landen, J. W., Andreasen, N., Cronenberger, C. L., Schwartz, P. F., Borjesson-Hanson, A., Ostlund, H., . . . Bednar, M. M. (2017). Ponezumab in Mild-to-Moderate Alzheimer's Disease: Randomized Phase II PET-PIB Study. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 3*(3), 393-401. https://doi.org/10.1016/j.trci.2017.05.003
452. Gold, M., Alderton, C., Zvartau-Hind, M., Egginton, S., Saunders, A. M., Irizarry, M., . . . Sawchak, S. (2010). Rosiglitazone Monotherapy in Mild-to-Moderate Alzheimer's Disease: Results from a Randomized, Double-Blind, Placebo-Controlled Phase III Study. *Dementia and Geriatric Cognitive Disorders, 30*(2), 131-146. https://doi.org/10.1159/000318845
453. Fleisher, A. S., Raman, R., Siemers, E. R., Becerra, L., Clark, C. M., Dean, R. A., . . . Thal, L. J. (2008). Phase 2 Safety Trial Targeting Amyloid Beta Production with a Gamma-Secretase Inhibitor in Alzheimer Disease. *Archives of Neurology, 65*(8), 1031-1038. https://doi.org/10.1001/archneur.65.8.1031
454. Henley, D. B., Sundell, K. L., Sethuraman, G., Dowsett, S. A., & May, P. C. (2014). Safety Profile of Semagacestat, a Gamma-Secretase Inhibitor: IDENTITY Trial Findings. *Current Medical Research and Opinion, 30*(10), 2021-2032. https://doi.org/10.1185/03007995.2014.939167
455. Malpas, C. B., Vivash, L., Genc, S., Saling, M. M., Desmond, P., Steward, C., . . . O'Brien, T. J. (2016). A Phase IIa Randomized Control Trial of VELO15 (Sodium Selenate) in Mild-Moderate Alzheimer's Disease. *Journal of Alzheimer's disease: JAD, 54*(1), 223-232. https://doi.org/10.3233/JAD-160544

456. Farlow, M., Arnold, S. E., van Dyck, C. H., Aisen, P. S., Snider, B. J., Porsteinsson, A. P., . . . Siemers, E. R. (2012). Safety and Biomarker Effects of Solanezumab in Patients with Alzheimer's Disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 8*(4), 261-271. https://doi.org/10.1016/j.jalz.2011.09.224
457. Doody, R. S., Thomas, R. G., Farlow, M., Iwatsubo, T., Vellas, B., Joffe, S., . . . Solanezumab Study Group. (2014). Phase 3 Trials of Solanezumab for Mild-to-Moderate Alzheimer's Disease. *New England Journal of Medicine, 370*(4), 311-321. https://doi.org/10.1056/NEJMoa1312889
458. Honig, L. S., Vellas, B., Woodward, M., Boada, M., Bullock, R., Borrie, M., . . . Siemers, E. (2018). Trial of Solanezumab for Mild Dementia Due to Alzheimer's Disease. *New England Journal of Medicine, 378*(4), 321-330. https://doi.org/10.1056/NEJMoa1705971
459. Green, R. C., Schneider, L. S., Amato, D. A., Beelen, A. P., Wilcock, G., Swabb, E. A., . . . Tarenflurbil Phase 3 Study Group. (2009). Effect of Tarenflurbil on Cognitive Decline and Activities of Daily Living in Patients with Mild Alzheimer Disease: A Randomized Controlled Trial. *JAMA: Journal of the American Medical Association, 302*(23), 2557-2564. https://doi.org/10.1001/jama.2009.1866
460. Wilcock, G. K., Black, S. E., Hendrix, S. B., Zavitz, K. H., Swabb, E. A., Laughlin, M. A., & Tarenflurbil Phase II Study Investigators. (2008). Efficacy and Safety of Tarenflurbil in Mild to Moderate Alzheimer's Disease: A Randomised Phase II Trial. *Lancet Neurology, 7*(6), 483-493. https://doi.org/10.1016/S1474-4422(08)70090-5
461. Decourt, B., Drumm-Gurnee, D., Wilson, J., Jacobson, S., Belden, C., Sirrel, S., . . . Sabbagh, M. N. (2017). Poor Safety and Tolerability Hamper Reaching a Potentially Therapeutic Dose in the Use of Thalidomide for Alzheimer's Disease: Results from a Double-Blind, Placebo-Controlled Trial. *Current Alzheimer Research, 14*(4), 403-411. https://doi.org/10.2174/1567205014666170117141330
462. Lovestone, S., Boada, M., Dubois, B., Hull, M., Rinne, J. O., Huppertz, H. J., . . . ARGO Investigators. (2015). A Phase II Trial of Tideglusib in Alzheimer's Disease. *Journal of Alzheimer's disease: JAD, 45*(1), 75-88. https://doi.org/10.3233/JAD-141959
463. Aisen, P. S., Saumier, D., Briand, R., Laurin, J., Gervais, F., Tremblay, P., & Garceau, D. (2006). A Phase II Study Targeting Amyloid-Beta with 3APS in Mild-to-Moderate Alzheimer Disease. *Neurology, 67*(10), 1757-1763. https://doi.org/10.1212/01.wnl.0000244346.08950.64
464. Saumier, D., Duong, A., Haine, D., Garceau, D., & Sampalis, J. (2009). Domain-Specific Cognitive Effects of Tramiprosate in Patients with Mild to Moderate Alzheimer's Disease: ADAS-Cog Subscale Results from the ALPHASE Study. *The Journal of Nutrition, Health & Aging, 13*(9), 808-812. https://doi.org/10.1007/s12603-009-0217-4
465. Gauthier, S., Aisen, P. S., Ferris, S. H., Saumier, D., Duong, A., Haine, D., . . . Sampalis, J. (2009). Effect of Tramiprosate in Patients with Mild-to-Moderate Alzheimer's Disease: Exploratory Analyses of the MRI Sub-Group of the ALPHASE Study. *The Journal of Nutrition, Health & Aging, 13*(6), 550-557. https://doi.org/10.1007/s12603-009-0106-x
466. Pasquier, F., Sadowsky, C., Holstein, A., Leterme Gle, P., Peng, Y., Jackson, N., . . . Team, A. C. C. S. (2016). Two Phase 2 Multiple Ascending-Dose Studies of Vanutide Cridificar (ACC-001) and QS-21 Adjuvant in Mild-to-Moderate Alzheimer's Disease. *Journal of Alzheimer's disease: JAD, 51*(4), 1131-1143. https://doi.org/10.3233/JAD-150376

467. Arai, H., Suzuki, H., & Yoshiyama, T. (2015). Vanutide Cridificar and the QS-21 Adjuvant in Japanese Subjects with Mild to Moderate Alzheimer's Disease: Results from Two Phase 2 Studies. *Current Alzheimer Research, 12*(3), 242-254. https://doi.org/10.2174/1567205012666150302154121
468. van Dyck, C. H., Sadowsky, C., Le Prince Leterme, G., Booth, K., Peng, Y., Marek, K., . . . Ryan, J. M. (2016). Vanutide Cridificar (ACC-001) and QS-21 Adjuvant in Individuals with Early Alzheimer's Disease: Amyloid Imaging Positron Emission Tomography and Safety Results from a Phase 2 Study. *The Journal of Prevention of Alzheimer's Disease, 3*(2), 75-84. https://doi.org/10.14283/jpad.2016.91
469. Ketter, N., Liu, E., Di, J., Honig, L. S., Lu, M., Novak, G., . . . Brashear, H. R. (2016). A Randomized, Double-Blind, Phase 2 Study of the Effects of the Vaccine Vanutide Cridificar with QS-21 Adjuvant on Immunogenicity, Safety and Amyloid Imaging in Patients with Mild to Moderate Alzheimer's Disease. *The Journal of Prevention of Alzheimer's Disease, 3*(4), 192-201. https://doi.org/10.14283/jpad.2016.118
470. Scheltens, P., Hallikainen, M., Grimmer, T., Duning, T., Gouw, A. A., Teunissen, C. E., . . . Prins, N. D. (2018). Safety, Tolerability and Efficacy of the Glutaminyl Cyclase Inhibitor PQ912 in Alzheimer's Disease: Results of a Randomized, Double-Blind, Placebo-Controlled Phase 2a Study. *Alzheimer's Research & Therapy, 10*(1), 107. https://doi.org/10.1186/s13195-018-0431-6
471. Egan, M. F., Kost, J., Tariot, P. N., Aisen, P. S., Cummings, J. L., Vellas, B., . . . Michelson, D. (2018). Randomized Trial of Verubecestat for Mild-to-Moderate Alzheimer's Disease. *New England Journal of Medicine, 378*(18), 1691-1703. https://doi.org/10.1056/NEJMoa1706441
472. Egan, M. F., Kost, J., Voss, T., Mukai, Y., Aisen, P. S., Cummings, J. L., . . . Michelson, D. (2019). Randomized Trial of Verubecestat for Prodromal Alzheimer's Disease. *New England Journal of Medicine, 380*(15), 1408-1420. https://doi.org/10.1056/NEJMoa1812840
473. Uysal-Bozkir, O., Parlevliet, J. L., & de Rooij, S. E. (2013). Insufficient Cross-Cultural Adaptations and Psychometric Properties for Many Translated Health Assessment Scales: A Systematic Review. *Journal of Clinical Epidemiology, 66*(6), 608-618. https://doi.org/10.1016/j.jclinepi.2012.12.004
474. Gutchess, A. H., Yoon, C., Luo, T., Feinberg, F., Hedden, T., Jing, Q., . . . Park, D. C. (2006). Categorical Organization in Free Recall across Culture and Age. *Gerontology, 52*(5), 314-323. https://doi.org/10.1159/000094613
475. Taiebine, M., & El Alaoui Faris, M. (2019). Neurolinguistic and Acoustic Study of Logopenic Primary Progressive Aphasia in Arabic. *Acta Neuropsychologica, 17*(4), 469-485.
476. Tee, B. L., Deleon, J., Chen Li Ying, L. K., Miller, B. L., R, Y. L., Europa, E., . . . Gorno-Tempini, M. L. (2021). Tonal and Orthographic Analysis in a Cantonese-Speaking Individual with Nonfluent/Agrammatic Variant Primary Progressive Aphasia. *Neurocase*, 1-10. https://doi.org/10.1080/13554794.2021.1925302
477. Nell, V. (2000). *Cross-Cultural Neuropsychological Assessment: Theory and Practice*. Lawrence Erlbaum Associates Publishers.
478. Lezak, M. D. (1982). The Problem of Assessing Executive Functions. *International Journal of Psychology, 17*(2-3), 281-297. https://doi.org/10.1080/00207598208247445
479. Zartman, A. L., Hilsabeck, R. C., Guarnaccia, C. A., & Houtz, A. (2013). The Pillbox Test: An Ecological Measure of Executive Functioning and Estimate of Medication

Management Abilities. *Archives of Clinical Neuropsychology, 28*(4), 307-319. https://doi.org/10.1093/arclin/act014

480. Ziemnik, R. E., & Suchy, Y. (2019). Ecological Validity of Performance-Based Measures of Executive Functions: Is Face Validity Necessary for Prediction of Daily Functioning? *Psychological Assessment, 31*(11), 1307-1318. https://doi.org/10.1037/pas0000751
481. Nepal, G. M., Shrestha, A., & Acharya, R. (2019). Translation and Cross-Cultural Adaptation of the Nepali Version the Rowland Universal Dementia Assessment Scale (RUDAS). *Journal of Patient-Reported Outcomes, 3*, 38. https://doi.org/10.1186/s41687-019-0132-3
482. Sturzeneker Trés, E., da Costa Miranda, D., Brucki, S. M., Oliveira, M. O., Amore Cecchini, M., Sanches Yassuda, M., & Nitrini, R. (2017). The Stick Design Test (SDT): Can a Visuoconstruction Test Help Discriminate Healthy Controls (HC), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD)? *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*(7S), P1139-P1140.
483. Chatterjee, S. (2020). *Eyes, Brains and Culture: A Systematic Review of the Use of Eye-Tracking, EEG, and FMRI in Measuring Cultural Differences in Object and Scene Perception*. University of Glasgow.
484. Nisbett, R. E., & Masuda, T. (2003). Culture and Point of View. *Proceedings of the National Academy of Sciences of the United States of America, 100*(19), 11163-11170. https://doi.org/10.1073/pnas.1934527100
485. Chatterjee, S. (2020). *Differences in Object Perception: A Comparison of Indian and British Participants on Scene and Silhouetted Object Perception Tasks. Unpublished Dissertation.* University of Glasgow.
486. Koelkebeck, K., Uwatoko, T., Tanaka, J., & Kret, M. E. (2017). How Culture Shapes Social Cognition Deficits in Mental Disorders: A Review. *Social Neuroscience, 12*(2), 102-112. https://doi.org/10.1080/17470919.2016.1155482
487. Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth About False Belief. *Child Development, 72*(3), 655-684. https://doi.org/10.1111/1467-8624.00304
488. Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" Test: Systematic Review of Psychometric Properties and a Validation Study in Italy. *Cognitive Neuropsychiatry, 18*(4), 326-354. https://doi.org/10.1080/13546805.2012.721728
489. Custodio, N., Montesinos, R., Cruzado, L., Herrera-Perez, E., Failoc-Rojas, V. E., Pintado-Caipa, M., . . . Diaz, M. M. (2021). Social Cognition and Behavioral Assessments Improve the Diagnosis of Behavioral Variant of Frontotemporal Dementia in Older Peruvians with Low Educational Levels. *Frontiers in Neurology, 12*, 704109. https://doi.org/10.3389/fneur.2021.704109
490. Elfenbein, H. A., & Ambady, N. (2002). On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis. *Psychological Bulletin, 128*(2), 203-235. https://doi.org/10.1037/0033-2909.128.2.203
491. Perez-Zapata, D., Slaughter, V., & Henry, J. D. (2016). Cultural Effects on Mindreading. *Cognition, 146*, 410-414. https://doi.org/10.1016/j.cognition.2015.10.018
492. van den Berg, E., Poos, J. M., Jiskoot, L. C., Montagne, B., Kessels, R. P. C., Franzen, S., . . . Papma, J. M. (2021). Impaired Knowledge of Social Norms in Dementia and Psychiatric Disorders: Validation of the Social Norms Questionnaire-Dutch Version (SNQ-NL). *Assessment*, 10731911211008234. https://doi.org/10.1177/10731911211008234

493. Raven, J. C. (1941). Standardisation of Progressive Matrices. *British Journal of Medical Psychology, 19*, 137-150.
494. Owen, K. (1992). The Suitability of Raven Standard Progressive Matrices for Various Groups in South-Africa. *Personality and Individual Differences, 13*(2), 149-159. https://doi.org/10.1016/0191-8869(92)90037-P
495. Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (2009). Variation in Raven's Progressive Matrices Scores across Time and Place. *Learning and Individual Differences, 19*(3), 330-338. https://doi.org/10.1016/j.lindif.2008.10.006
496. Borra, R., van Dijk, R., & Verboom, R. (2016). *Cultuur en Psychodiagnostiek* (2nd ed.). Bohn, Stafleu & van Loghum.
497. Hoogsteder, M., & Borges Dias, E. (2016). Tests en Testgebruik in een Interculturele Context: Een Verkennend Overzicht. In R. Borra, R. van Dijk, & R. Verboom (Eds.), *Cultuur en Psychodiagnostiek* (2nd ed., pp. 47-80). Bohn, Stafleu & van Loghum.
498. van Wezel, N., van der Heide, I., Deville, W. L. J. M., Duran, G., Hoopman, R., Blom, M. M., . . . Francke, A. L. (2021). The Turkish Version of the SPPIC Validated among Informal Caregivers with a Turkish Immigrant Background. *BMC Geriatrics, 21*(1). https://doi.org/10.1186/s12877-021-02161-6
499. Pot, A. M., van Dyck, R., & Deeg, D. J. (1995). [Perceived Stress Caused by Informal Caregiving. Construction of a Scale]. *Tijdschrift voor Gerontologie en Geriatrie, 26*(5), 214-219.
500. van Wieringen, J. C. M., & Van Grondelle, N. J. (2014). Migrantenmantelzorgers: Onzichtbaar, Onmisbaar... Overbelast. *Bijblijven, 30*, 32-39.
501. van der Woude, A. (2021). Stereotiepe Benadering van Migrantenouderen Doet Geen Recht aan Diversiteit in de Praktijk. https://www.movisie.nl/artikel/stereotype-benadering-migrantenouderen-doet-geen-recht-aan-diversiteit-praktijk
502. Hokkanen, L., Barbosa, F., Ponchel, A., Constantinou, M., Kosmidis, M. H., Varako, N., . . . Hessen, E. (2020). Clinical Neuropsychology as a Specialist Profession in European Health Care: Developing a Benchmark for Training Standards and Competencies Using the Europsy Model? *Frontiers in Psychology, 11*, 559134. https://doi.org/10.3389/fpsyg.2020.559134
503. Ibáñez, A., Slachevsky, A., & Serrano, C. (2020). *Manual De Buenas Practicas Para El Diagnóstico De Demencias*. http://lac-cd.org/2020/06/17/manual-de-buenas-practicas-para-el-diagnostico-de-la-demancia/
504. Mooldijk, S. S., Licher, S., & Wolters, F. J. (2021). Characterizing Demographic, Racial, and Geographic Diversity in Dementia Research: A Systematic Review. *Jama Neurology*. https://doi.org/10.1001/jamaneurol.2021.2943
505. Raman, R., Quiroz, Y. T., Langford, O., Choi, J., Ritchie, M., Baumgartner, M., . . . Grill, J. D. (2021). Disparities by Race and Ethnicity among Adults Recruited for a Preclinical Alzheimer Disease Trial. *Jama Network Open, 4*(7). https://doi.org/10.1001/jamanetworkopen.2021.14364
506. Goudsmit, M., van de Vorst, I., van Campen, J., Parlevliet, J., & Schmand, B. (2021). Clinical Characteristics and Presenting Symptoms of Dementia - a Case-Control Study of Older Ethnic Minority Patients in a Dutch Urban Memory Clinic. *Aging and Mental Health*, 1-8. https://doi.org/10.1080/13607863.2021.1963416
507. Lenhard, A., Lenhard, W., & Gary, S. (2019). Continuous Norming of Psychometric Tests: A Simulation Study of Parametric and Semi-Parametric Approaches. *PLoS One, 14*(9), e0222279. https://doi.org/10.1371/journal.pone.0222279

508. American Psychological Association. *APA Dictionary of Psychology*. https://dictionary.apa.org/acculturation
509. Celenk, O., & van de Vijver, F. (2011). Assessment of Acculturation: Issues and Overview of Measures. *Online Readings in Psychology and Culture, 8*(1). https://doi.org/10.9707/2307-0919.1105
510. Lamar, M., Barnes, L. L., Leurgans, S. E., Fleischman, D. A., Farfel, J. M., Bennett, D. A., & Marquez, D. X. (2021). Acculturation in Context: The Relationship between Acculturation and Socioenvironmental Factors with Level of and Change in Cognition in Older Latinos. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 76*(4), E129-E139. https://doi.org/10.1093/geronb/gbaa156
511. Jang, Y., Small, B. J., & Haley, W. E. (2001). Cross-Cultural Comparability of the Geriatric Depression Scale: Comparison between Older Koreans and Older Americans. *Aging and Mental Health, 5*(1), 31-37. https://doi.org/10.1080/13607860020020618
512. Manson, S. M. (1995). Culture and Major Depression - Current Challenges in the Diagnosis of Mood Disorders. *Psychiatric Clinics of North America, 18*(3), 487-501. https://doi.org/10.1016/S0193-953x(18)30036-4
513. Debruyne, H., Van Buggenhout, M., Le Bastard, N., Aries, M., Audenaert, K., De Deyn, P. P., & Engelborghs, S. (2009). Is the Geriatric Depression Scale a Reliable Screening Tool for Depressive Symptoms in Elderly Patients with Cognitive Impairment? *International Journal of Geriatric Psychiatry, 24*(6), 556-562. https://doi.org/10.1002/gps.2154
514. Nichter, M. (2010). Idioms of Distress Revisited. *Culture, Medicine, and Psychiatry, 34*(2), 401-416. https://doi.org/10.1007/s11013-010-9179-6
515. Verboom, R. (2002). Psychodiagnostisch Onderzoek bij Migranten. In E. van Meekeren, R. May, & A. Limburg-Okken (Eds.), *Culturen binnen de Psychiatriemuren. Geestelijke Gezondheidszorg in een Multiculturele Samenleving* (pp. 94-108). Boom.
516. Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelicic, M., & Otgaar, H. (2019). When Patients Overreport Symptoms: More Than Just Malingering. *Current Directions in Psychological Science, 28*(3), 321-326. https://doi.org/10.1177/0963721419837681
517. Gianattasio, K. Z., Bennett, E. E., Wei, J., Mehrotra, M. L., Mosley, T., Gottesman, R. F., . . . Alzheimer's Disease Neuroimaging Initiative. (2021). Generalizability of Findings from a Clinical Sample to a Community-Based Sample: A Comparison of ADNI and ARIC. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 17*(8), 1265-1276. https://doi.org/10.1002/alz.12293
518. Gleason, C. E., Norton, D., Zuelsdorff, M., Benton, S. F., Wyman, M. F., Nystrom, N., . . . Asthana, S. (2019). Association between Enrollment Factors and Incident Cognitive Impairment in Blacks and Whites: Data from the Alzheimer's Disease Center. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 15*(12), 1533-1545. https://doi.org/10.1016/j.jalz.2019.07.015
519. Fokkema, T. (2011). Interculturalisatie van de Ouderenzorg: Een Terugblik en Roep om Meer Onderzoek. *Tijdschrift voor Gerontologie en Geriatrie, 42*, 2-6.
520. Adamec, Z., & Drápela, K. (2015). Generalized Additive Models as an Alternative Approach to the Modelling of the Tree Height-Diameter Relationship. *Journal of Forest Science, 61*, 235-243.
521. Vissenberg, R., Uysal, O., Goudsmit, M., van Campen, J., & Buurman-van Es, B. (2018). Barriers in Providing Primary Care for Immigrant Patients with Dementia:

GPs' Perspectives. *BJGP Open, 2*(4), bjgpopen18X101610. https://doi.org/10.3399/bjgpopen18X101610

522. Hogeschool Rotterdam. (n.d.). Hoe Herken Ik Vergeetachtigheid en Dementie? https://www.hogeschoolrotterdam.nl/contentassets/e7005ef70e2e446a989eb09622f981bf/signalenkaart_rotterdam_print-folderversie.pdf
523. Denktas, S., Koopmans, G., Birnie, E., Foets, M., & Bonsel, G. (2009). Ethnic Background and Differences in Health Care Use: A National Cross-Sectional Study of Native Dutch and Immigrant Elderly in the Netherlands. *International Journal for Equity in Health, 8*. https://doi.org/10.1186/1475-9276-8-35
524. Segers, K., Benoit, F., Colson, C., Kovac, V., Nury, D., & Vanderaspoilden, V. (2013). Pioneers in Migration, Pioneering in Dementia: First Generation Immigrants in a European Metropolitan Memory Clinic. *Acta Neurologica Belgica, 113*(4), 435-440. https://doi.org/10.1007/s13760-013-0245-z
525. Sikkes, S. A., de Lange-de Klerk, E. S., Pijnenburg, Y. A., Scheltens, P., & Uitdehaag, B. M. (2009). A Systematic Review of Instrumental Activities of Daily Living Scales in Dementia: Room for Improvement. *Journal of Neurology, Neurosurgery and Psychiatry, 80*(1), 7-12. https://doi.org/10.1136/jnnp.2008.155838
526. Dubbelman, M. A., Verrijp, M., Facal, D., Sanchez-Benavides, G., Brown, L. J. E., van der Flier, W. M., . . . Sikkes, S. A. M. (2020). The Influence of Diversity on the Measurement of Functional Impairment: An International Validation of the Amsterdam IADL Questionnaire in Eight Countries. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 12*(1), e12021. https://doi.org/10.1002/dad2.12021
527. Stringer, G., Leroi, I., Sikkes, S. A. M., Montaldi, D., & Brown, L. J. E. (2021). Enhancing 'Meaningfulness' of Functional Assessments: UK Adaptation of the Amsterdam IADL Questionnaire. *International Psychogeriatrics, 33*(1), 39-50. https://doi.org/10.1017/S1041610219001881
528. Fillenbaum, G. G., Chandra, V., Ganguli, M., Pandav, R., Gilby, J. E., Seaberg, E. C., . . . Nath, L. M. (1999). Development of an Activities of Daily Living Scale to Screen for Dementia in an Illiterate Rural Older Population in India. *Age and Ageing, 28*(2), 161-168. https://doi.org/10.1093/ageing/28.2.161
529. Kohrt, B. A., Rasmussen, A., Kaiser, B. N., Haroz, E. E., Maharjan, S. M., Mutamba, B. B., . . . Hinton, D. E. (2014). Cultural Concepts of Distress and Psychiatric Disorders: Literature Review and Research Recommendations for Global Mental Health Epidemiology. *International Journal of Epidemiology, 43*(2), 365-406. https://doi.org/10.1093/ije/dyt227
530. Laroi, F., Luhrmann, T. M., Bell, V., Christian, W. A., Jr., Deshpande, S., Fernyhough, C., . . . Woods, A. (2014). Culture and Hallucinations: Overview and Future Directions. *Schizophrenia Bulletin, 40 Suppl 4*, S213-220. https://doi.org/10.1093/schbul/sbu012
531. Nielsen, T. R., Phung, T. K., Chaaya, M., Mackinnon, A., & Waldemar, G. (2016). Combining the Rowland Universal Dementia Assessment Scale and the Informant Questionnaire on Cognitive Decline in the Elderly to Improve Detection of Dementia in an Arabic-Speaking Population. *Dementia and Geriatric Cognitive Disorders, 41*(1-2), 46-54. https://doi.org/10.1159/000441649
532. Statucka, M., Cherian, K., Fasano, A., Munhoz, R. P., & Cohn, M. (2021). Multiculturalism: A Challenge for Cognitive Screeners in Parkinson's Disease. *Movement Disorders Clinical Practice, 8*(5), 733-742. https://doi.org/10.1002/mdc3.13240

533. Onyike, C. U., Shinagawa, S., & Ellajosyula, R. (2021). Frontotemporal Dementia: A Cross-Cultural Perspective. *Advances in Experimental Medicine and Biology, 1281*, 141-150. https://doi.org/10.1007/978-3-030-51140-1_10
534. van den Berg, E., Jiskoot, L. C., Grosveld, M. J. H., van Swieten, J. C., & Papma, J. M. (2017). Qualitative Assessment of Verbal Fluency Performance in Frontotemporal Dementia. *Dementia and Geriatric Cognitive Disorders, 44*(1-2), 35-44. https://doi.org/10.1159/000477538
535. Sempertegui, G. A., Knipscheer, J. W., & Bekker, M. H. J. (2018). Development and Evaluation of Diversity-Oriented Competence Training for the Treatment of Depressive Disorders. *Transcultural Psychiatry, 55*(1), 31-54. https://doi.org/10.1177/1363461517725224
536. Nederlands Instituut voor Psychologen - Sectie Neuropsychologie. (2019). Cross-Culturele Diagnostiek Bij Lichte Cognitieve Stoornissen (MCI) en Dementie. In *Monodisciplinaire Richtlijn Neuropsychologisch Onderzoek bij Lichte Cognitieve Stoornissen (MCI) en Dementie*.
537. Zorgstandaard Dementie. (2020). https://www.vilans.nl/vilans/media/documents/producten/zorgstandaard-dementie.pdf
538. Dieleman-Bij de Vaate, A. J. M., Eizenga, W. H., Lunter-Driever, P. G. M., Moll van Charante, E. P., Perry, M., Schep-Akkerman, A., . . . van der Weele, G. M. (2020). NHG-Standaard Dementie (M21). https://richtlijnen.nhg.org/files/pdf/103_Dementie_april-2020.pdf
539. Leyerzapf, H., & Abma, T. (2012). Naar een Kleurrijk UMC. Ervaringen van Arts-Assisten en Opleiders op Medische Afdelingen. Eindrapport Onderzoek naar Doorstroom van Artsen met 'Allochtone' Achtergrond naar Specialisatie. VU medisch centrum.
540. Knipscheer, J. W., & Kleber, R. J. (2001). Help-Seeking Attitudes and Utilization Patterns for Mental Health Problems of Surinamese Migrants in the Netherlands. *Journal of Counseling Psychology, 48*(1), 28-38. https://doi.org/10.1037//0022-0167.48.1.28
541. Knipscheer, J. W., & Kleber, R. J. (2004). A Need for Ethnic Similarity in the Therapist-Patient Interaction? Mediterranean Migrants in Dutch Mental-Health Care. *Journal of Clinical Psychology, 60*(6), 543-554. https://doi.org/10.1002/jclp.20008
542. Conkova, N., & Lindenberg, J. (2018). Gezondheid en Welbevinden van Oudere Migranten in Nederland: Een Narratieve Literatuurstudie. *Tijdschrift voor Gerontologie en Geriatrie, 49*, 223-231.
543. Brody, J. L., Dalen, J., Annett, R. D., Scherer, D. G., & Turner, C. W. (2012). Conceptualizing the Role of Research Literacy in Advancing Societal Health. *Journal of Health Psychology, 17*(5), 724-730. https://doi.org/10.1177/1359105311425273
544. Gilmore-Bykovskyi, A., Croff, R., Glover, C. M., Jackson, J. D., Resendez, J., Perez, A., . . . Manly, J. J. (2021). Traversing the Aging Research and Health Equity Divide: Toward Intersectional Frameworks of Research Justice and Participation. *Gerontologist*. https://doi.org/10.1093/geront/gnab107
545. Conkova, N., & Lindenberg, J. (2020). The Experience of Aging and Perceptions of "Aging Well" among Older Migrants in the Netherlands. *Gerontologist, 60*(2), 270-278. https://doi.org/10.1093/geront/gnz125

The culturally, educationally, and linguistically diverse population of Europe is aging rapidly, resulting in an increase in the number of patients with a diverse background visiting memory clinics. Neuropsychologists are faced with diagnostic challenges, such as barriers in language and culture, as well as a lack of suitable tests and norms.

This PhD dissertation highlights the lack of appropriate neuropsychological tests in cognitive domains other than memory, particularly tests of language, executive functioning, and social cognition. Furthermore, it sheds light on how European countries approach the assessment of diverse populations. The focus subsequently shifts to solutions to the challenges in cross-cultural neuropsychological assessment. One chapter describes the development and validation of the TULIPA test battery, an instrument showing promising feasibility in a diverse memory clinic setting. This dissertation also highlights some of the newly developed neuropsychological tests that form part of this test battery, such as the Naming Assessment in Multicultural Europe (NAME) and modified Visual Association Test (mVAT). Both of these instruments break with tradition through their use of colored photographs instead of the black-and-white line drawings that are known to be less suitable for low educated populations.

Last, this dissertation addresses next steps in clinical practice and research. It presents the standpoints and goals of the European Consortium on Cross-Cultural Neuropsychology. Furthermore, it provides practical guidelines for clinicians on cross-cultural neuropsychological assessment. Last, it examines how eligibility criteria may contribute to the underrepresentation of diverse populations in clinical trials.