

The Naming Assessment in Multicultural Europe (NAME): Development and Validation in a Multicultural Memory Clinic

S. Franzen^{1*} , E. van den Berg¹ , Y. Ayhan², D.D. Satoer³, Ö. Türkoğlu², G.E. Genç Akpulat², E.G. Visch-Brink³, E.A. Scheffers⁴, J. Kranenburg⁴, L.C. Jiskoot¹, J. van Hemmen¹ and J.M. Papma¹ 

¹Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

²Department of Psychiatry, Hacettepe University Faculty of Medicine, Ankara, Turkey

³Department of Neurosurgery, Erasmus MC University Medical Center, Rotterdam, The Netherlands

⁴Department of Internal Medicine, Haaglanden Medical Center, The Hague, The Netherlands

(RECEIVED June 24, 2021; FINAL REVISION December 2, 2021; ACCEPTED December 11, 2021)

Abstract

Objective: 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).

Method: 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 ($\rho = .35-.73$). 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.

Keywords: Language tests, Dementia, Alzheimer disease, Aphasia, Cultural diversity, Cross-cultural comparison, Naming

INTRODUCTION

Naming impairment is frequently reported across a variety of neurological diseases, such as in temporal lobe epilepsy (Hamberger, 2015), post-stroke (Engelter et al., 2006), in brain tumors (Satoer, Vincent, Smits, Dirven & Visch-Brink, 2013), and in various neurodegenerative diseases, such as Alzheimer’s disease dementia (AD) and frontotemporal dementia (FTD; Grossman et al., 2004). 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 (Gleichgerrcht, Fridriksson & Bonilha, 2015). 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 (Ivanova & Hallowell, 2013).

The Boston Naming Test (BNT; Kaplan, Goodglass & Weintraub, 1983) is the most widely available and used test to assess naming impairment in the USA, Canada, and Europe (Maruta, Guerreiro, de Mendonca, Hort & Scheltens, 2011; Rabin, Paolillo & Barr, 2016). 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

*Correspondence and reprint requests to: Sanne Franzen, Department of Neurology, Erasmus University Medical Center, NF-331, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. E-mail: s.franzen@erasmusmc.nl

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 (Baird, Ford & Podell, 2007; Boone, Victor, Wen, Razani & Ponton, 2007), even after controlling for age, general cognitive impairment, formal education, and reading level (Baird et al., 2007). Research suggests that the test stimuli themselves may be systematically biased against certain groups (Boone et al., 2007), and studies from Australia (Worrall, Yiu, Hickson & Barnett, 1995), New Zealand (Barker-Collo, 2007), French-speaking Canada (Roberts & Doucet, 2011), and Korea (Kim & Na, 1999), 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 (Kim et al., 2017). 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 (Allegri et al., 1997). 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 (Byrd et al., 2021). 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 (Gasquoine, 2009; Franzen on behalf of the European Consortium on Cross-Cultural Neuropsychology, 2021).

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 scores through increased vocabulary and exposure to certain items not otherwise encountered in daily life, but can also (indirectly) impact the test scores through differences in the processing of the stimuli. Reis, Petersson, Castro-Caldas and Ingvar (2001) 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 (Reis, Faisca, Ingvar & Petersson, 2006), 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 (Franzen et al., 2020; Ivanova & Hallowell, 2013). 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” (Gollan, Weissberger, Runnqvist, Montoya & Cera, 2012); 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; Ardila, 2007). The CLNT consists of a set of 40 items that have corresponding words in many languages according to the Swadesh list (Swadesh, 1952), 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 (Galvez-Lara et al., 2015). Ardila warned, however, that his test may have low sensitivity due to ceiling effects, which were observed in control participants across several countries (Abou-Mrad et al., 2017; Galvez-Lara et al., 2015). 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 (Vivas, Manoilloff, Linares, Zaionz & Montero, 2020), and the Test de Dénomination de Québec-60 images (Macoir, Beaudoin, Bluteau, Potvin & Wilson, 2018). 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 (Franzen, Papma, van den Berg & Nielsen, 2021). 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 (Franzen et al., 2021).

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.

METHOD

Development and Pilot Studies of the Naming Assessment in Multicultural Europe

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 (2007), as well as items from various other sources, such as the dataset by Snodgrass and Vanderwart (1980). 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 (2013) 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 (Kuperman, Stadthagen-Gonzalez & Brysbaert, 2012; Luniewska et al., 2016; project Gutenberg), Dutch (Brysbaert, Stevens, De Deyne, Voorspoels & Storms, 2014; Keuleers, Brysbaert & New, 2010; Luniewska et al., 2016), Spanish (Alonso, Fernandez & Diez, 2015; Luniewska et al., 2016; opensubtitles.org), Polish (Luniewska et al., 2016; opensubtitles.org), and Turkish (Luniewska et al., 2016; opensubtitles.org). Age of acquisition data only was available for Portuguese (Cameirao & Vicente, 2010; Marques, Fonseca, Morais & Pinto, 2007), French (Ferrand et al., 2008), Italian (Luniewska et al., 2016), German (Birchenough, Davies & Connelly, 2017; Luniewska et al., 2016), Swedish (Luniewska et al., 2016), Russian (Luniewska et al., 2016) and Hebrew (Luniewska et al., 2016). Frequency data only was available for Arabic (Dukes, 2009; 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.

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.

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 \leq 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.

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 (2007) 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 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 patient 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.

Validation Study

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; Storey, Rowland, Basic, Conforti & Dickson, 2004) 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.

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, Goudsmit et al., 2017), modified Visual Association Test (mVAT, Franzen et al., 2019) and RUDAS (Goudsmit et al., 2018). 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 (Nielsen et al., 2018). Demographic data were collected at the neuropsychological assessment, with level of education scored according to the system of Verhage (1964), 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, Marin, Sabogal, Marin, Otero-Sabogal & Perez-Stable 1987) 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 (Caman et al., 2019).

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 (Jessen et al., 2020), mild cognitive impairment (Albert et al., 2011), and dementia subtypes (e.g. McKhann et al., 2011; Román et al., 1993), and the DSM-V for primary psychiatric disorders (American Psychiatric Association, 2013). Although neuropsychologists

Table 1. Percent correct per NAME item by group

Item		Controls (<i>n</i> = 39)	Other patients (<i>n</i> = 106)	AD/mixed (<i>n</i> = 30)	<i>p</i> -value AD/mixed vs. rest [†]
<i>Nature</i>					
Tree	(E)	100%	91.30%	76.70%	<.05
Sun	(E)	92.30%	86.60%	70.00%	<.05
Moon	(E)	87.20%	75.70%	46.70%	<.01
Sea	(E)	100%	90.60%	73.30%	<.01
Fire	(E)	100%	84.00%	33.30%	<.001
<i>Animals</i>					
Dog	(E)	100%	100%	100%	n.s.
Fish	(E)	100%	99.00%	90.00%	<.05
Bird	(E)	100%	95.30%	90.00%	n.s.
Ant*	(M)	94.90%	66.00%	36.70%	<.001
Snake	(H)	100%	95.20%	80.00%	<.01
Worm	(H)	84.60%	57.30%	16.70%	<.001
<i>Colors</i>					
Red	(E)	100%	97.20%	76.70%	n.s.
Green	(E)	100%	94.30%	73.30%	<.01
Black	(E)	100%	98.10%	86.70%	<.05
White	(M)	100%	96.20%	83.30%	<.05
<i>Verbs</i>					
Eat	(E)	97.40%	91.30%	56.70%	<.001
Drink	(E)	100%	99.00%	86.70%	<.01
Sit	(E)	94.90%	89.60%	80.00%	n.s.
Walk	(E)	94.90%	98.10%	80.00%	<.01
Sleep	(E)	100%	98.10%	90.00%	0.05
Laugh*	(E)	100%	96.20%	86.70%	<.05
Swim	(H)	100%	90.60%	63.30%	<.001
Drive*	(H)	100%	98.10%	86.70%	<.01
<i>Body and body parts</i>					
Hair*	(E)	94.90%	92.20%	83.30%	n.s.
Ear	(E)	100%	95.20%	83.30%	<.05
Eye	(E)	100%	97.10%	86.70%	<.05
Nose	(E)	100%	96.20%	76.70%	<.01
Tongue	(E)	92.30%	87.70%	70.00%	<.05
Foot	(E)	100%	96.10%	86.70%	<.05
Hand*	(E)	100%	97.20%	96.70%	n.s.
Bone	(H)	97.40%	85.80%	63.30%	<.01
Wing*	(H)	92.30%	75.50%	46.70%	<.01
Feather*	(H)	87.20%	76.00%	43.30%	<.001
<i>Objects*</i>					
Boat	(E)	100%	93.40%	66.70%	<.001
Book	(E)	100%	90.60%	83.30%	<.05
Table	(E)	100%	97.10%	86.70%	<.05
Chair	(E)	100%	99.00%	96.70%	n.s.
Pants	(E)	100%	95.30%	90.00%	n.s.
Bread	(E)	100%	95.10%	76.70%	<.01
Apple	(E)	100%	98.10%	96.70%	n.s.
Rope	(H)	100%	96.10%	86.70%	<.05
Bucket	(H)	97.40%	94.20%	73.30%	<.01
Candle	(H)	100%	97.20%	76.70%	<.001
Football	(H)	100%	98.10%	90.00%	0.05
Key	(H)	100%	99.10%	93.30%	n.s.
Axe	(H)	94.90%	84.50%	80.00%	n.s.
Cigarette	(H)	100%	96.20%	86.70%	<.05
Ring	(H)	82.10%	62.30%	36.70%	<.01
Envelope	(H)	97.40%	89.60%	80.00%	n.s.
Scissors	(H)	100%	99.10%	90.00%	<.05

(Continued)

Table 1. (Continued)

Item		Controls (<i>n</i> = 39)	Other patients (<i>n</i> = 106)	AD/mixed (<i>n</i> = 30)	<i>p</i> -value AD/mixed vs. rest [†]
Match	(H)	94.90%	82.10%	53.30%	0.001
Glasses	(H)	100%	100%	86.70%	<.01
<i>Occupations*</i>					
Doctor	(M)	100%	94.30%	53.30%	<.001
Teacher	(M)	100%	89.60%	70.00%	<.01
Policeman	(M)	100%	83.70%	60.00%	<.01
Baker	(H)	94.90%	72.60%	46.70%	<.01
Butcher	(H)	97.40%	78.30%	40.00%	<.001
Dentist	(H)	100%	85.60%	50.00%	<.001
Firefighter	(H)	97.40%	75.50%	16.70%	<.001
Chef	(H)	84.60%	61.20%	30.00%	<.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 (2007). 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.



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

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].

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, naming subtest of RPT) 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 (1964) 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.

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.

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

	Controls (<i>n</i> = 39)	Rotterdam cohort (<i>n</i> = 75)	Ankara Hacettepe cohort (<i>n</i> = 61)	
Age	61.8 (7.2)	64.8 (12.7)	74.0 (8.5)	<i>p</i> < .001
Education <i>n</i> (%)				
Zero years of education	2 (5.1%)	9 (12%)	15 (24.6%)	<i>p</i> = .43
>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 <i>n</i> (% male)	12 (30.8%)	36 (48.0%)	23 (37.7%)	<i>p</i> = .19
Years in the Netherlands	39 (11, <i>n</i> = 24)	38 (13)	–	–
RUDAS ²	27.7 (1.8)	22.2 (5.1, <i>n</i> = 62)	–	–
3MS ²	–	–	60.9 (22.9)	–
Diagnosis <i>n</i> (%)				
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	–	3 (15.0%)	2 (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 (UNESCO Institute for Statistics, n.d.) – 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.

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).

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,

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

	<i>n</i>	ρ	<i>p</i> -value
<i>Convergent validity</i>			
Animal fluency	154	0.73	<.001
Foods fluency	93	0.58	<.001
Naming subtest of Recall of Pictures Test	69	0.35	0.004
RUDAS	99	0.68	<.001
3MS	60	0.82	<.001
<i>Divergent validity</i>			
Modified Visual Association Test	61	0.54	<.001
CCD Objects test A	60	0.61	<.001
CCD Sun-Moon test A	62	–0.35	0.005
CCD Sun-Moon test B	59	–0.25	0.06

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

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).

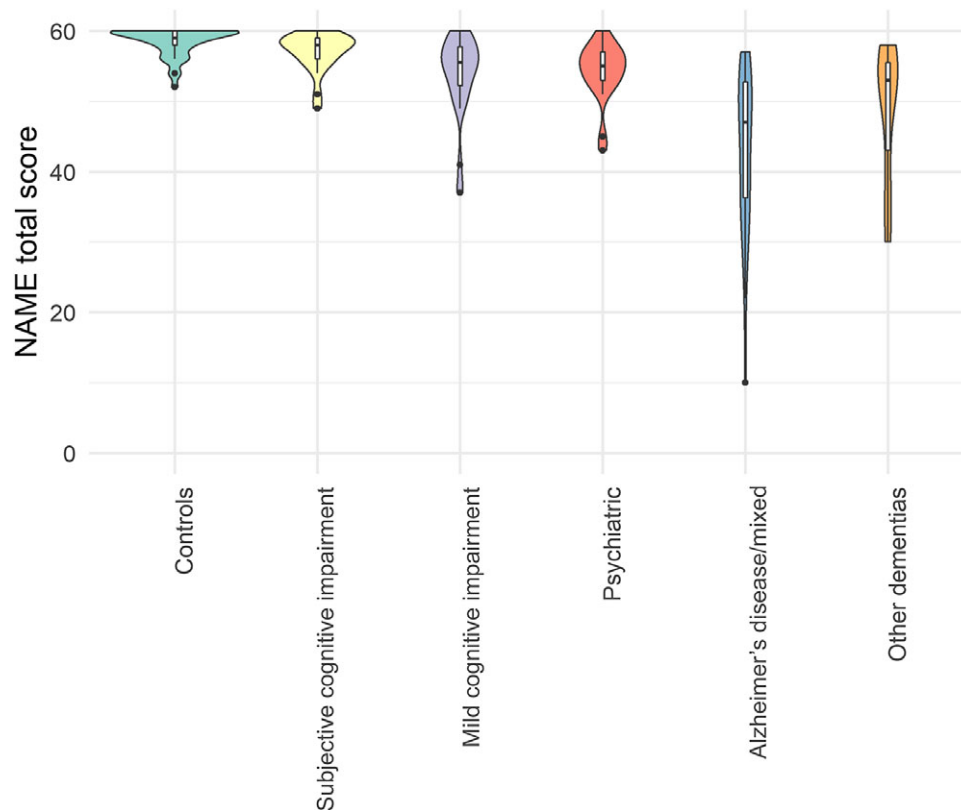


Fig. 2. Violin plot of the NAME scores by diagnosis type.

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%).

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 (Ardila, 2007) and the Recall of Pictures Test of the European Cross-Cultural Neuropsychological Test Battery (Nielsen et al., 2018) highlighted issues with sensitivity/ceiling effects and limited

diagnostic accuracy – the CLNT had a specificity of 94.6%, but sensitivity of only 58.3% (Galvez-Lara et al., 2015) 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 s) 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 (Howard et al., 2021). 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 (Nielsen, Andersen, Kastrup, Phung & Waldemar, 2011; Nielsen, Vogel, Phung, Gade & Waldemar, 2011).

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 (Zec, Markwell, Burkett & Larsen, 2005). 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 (Franzen et al., 2021). 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 (Franzen et al., 2019).

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 (Franzen on behalf of the European Consortium on Cross-Cultural Neuropsychology, 2021). 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 (Ivanova, Salmon & Gollan, 2014). 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. Hillis, Oh & Ken, 2004; Pisoni et al., 2018). 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. Eng, Vonk, Salzberger & Yoo, 2019).

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.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S135561772100148X>

ACKNOWLEDGMENTS

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.

FINANCIAL SUPPORT

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

CONFLICTS OF INTEREST

The authors have nothing to disclose.

REFERENCES

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. doi: [10.1080/13854046.2017.1288270](https://doi.org/10.1080/13854046.2017.1288270)

- Albert M.S., DeKosky S.T., Dickson D., Dubois B., Feldman H.H., Fox N.C. . . . et al. (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. doi: [10.1016/j.jalz.2011.03.008](https://doi.org/10.1016/j.jalz.2011.03.008)
- Allegri R.F., Fernandez Villavicencio A., Taragano F.E., Rymberg S., Mangone C.A., & Baumann D. (1997). Spanish boston naming test norms. *The Clinical Neuropsychologist*, 11(4), 416–420. doi: [10.1080/13854049708400471](https://doi.org/10.1080/13854049708400471)
- 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. doi: [10.3758/s13428-014-0454-2](https://doi.org/10.3758/s13428-014-0454-2)
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Ardila A. (2007). Toward the development of a cross-linguistic naming test. *Archives of Clinical Neuropsychology*, 22(3), 297–307. doi: [10.1016/j.acn.2007.01.016](https://doi.org/10.1016/j.acn.2007.01.016)
- 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. doi: [10.1016/j.acn.2007.01.005](https://doi.org/10.1016/j.acn.2007.01.005)
- Barker-Collo S. (2007). Boston naming test performance of older New Zealand adults. *Aphasiology*, 21(12), 1171–1180. doi: [10.1080/02687030600821600](https://doi.org/10.1080/02687030600821600)
- 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. doi: [10.3758/s13428-016-0718-0](https://doi.org/10.3758/s13428-016-0718-0)
- 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. doi: [10.1016/j.acn.2007.01.010](https://doi.org/10.1016/j.acn.2007.01.010)
- Brybaert 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. doi: [10.1016/j.actpsy.2014.04.010](https://doi.org/10.1016/j.actpsy.2014.04.010)
- 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. doi: [10.1037/pas0000993](https://doi.org/10.1037/pas0000993)
- 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. doi: [10.1159/000500939](https://doi.org/10.1159/000500939)
- 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. doi: [10.3758/BRM.42.2.474](https://doi.org/10.3758/BRM.42.2.474)
- Dukes K. (2009). The Quranic Arabic Corpus. Retrieved from <https://corpus.quran.com/>
- 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. doi: [10.1177/1747021818765997](https://doi.org/10.1177/1747021818765997)
- 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. doi: [10.1161/01.STR.0000221815.64093.8c](https://doi.org/10.1161/01.STR.0000221815.64093.8c)
- 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.
- 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*, [10.1080/13854046.2021.1981456](https://doi.org/10.1080/13854046.2021.1981456), Advance online publication.
- Franzen S., Papma J.M., van den Berg E., & Nielsen T.R. (2021). Cross-cultural neuropsychological assessment in the European Union: A Delphi expert study. *Archives of Clinical Neuropsychology*, 36(5), 815–830. doi: [10.1093/arclin/aca083](https://doi.org/10.1093/arclin/aca083)
- 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. doi: [10.1017/S1355617719000894](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., & 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. doi: [10.1159/000501151](https://doi.org/10.1159/000501151)
- Galvez-Lara M., Moriana J.A., Vilar-Lopez R., Fafous 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. doi: [10.1080/13803395.2014.1003533](https://doi.org/10.1080/13803395.2014.1003533)
- Gasquoine P.G. (2009). Race-norming of neuropsychological tests. *Neuropsychology Review*, 19, 250–262. doi: [10.1007/s11065-009-9090-5](https://doi.org/10.1007/s11065-009-9090-5)
- Gleichgerrecht E., Fridriksson J., & Bonilha L. (2015). Neuroanatomical foundations of naming impairments across different neurologic conditions. *Neurology*, 85(3), 284–292. doi: [10.1212/WNL.0000000000001765](https://doi.org/10.1212/WNL.0000000000001765)
- 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. doi: [10.1017/S1366728911000332](https://doi.org/10.1017/S1366728911000332)
- 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. doi: [10.1080/13803395.2016.1209464](https://doi.org/10.1080/13803395.2016.1209464)
- 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. doi: [10.1159/000490174](https://doi.org/10.1159/000490174)
- Grossman M., McMillan C., Moore P., Ding L., Glosner 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. doi: [10.1093/brain/awh075](https://doi.org/10.1093/brain/awh075)
- Hamberger M.J. (2015). Object naming in epilepsy and epilepsy surgery. *Epilepsy & Behavior*, 46, 27–33. doi: [10.1016/j.yebeh.2014.12.019](https://doi.org/10.1016/j.yebeh.2014.12.019)
- 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. doi: [10.1002/ana.10812](https://doi.org/10.1002/ana.10812)
- 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. doi: [10.1212/WNL.0000000000011699](https://doi.org/10.1212/WNL.0000000000011699)
- 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. doi: [10.1017/S1355617714000228](https://doi.org/10.1017/S1355617714000228)
- 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. doi: [10.1080/02687038.2013.805728](https://doi.org/10.1080/02687038.2013.805728)
- Jessen F., Amariglio R.E., Buckley R.F., van der Flier W.M., Han Y., Molinuevo J.L. . . . et al. (2020). The characterisation of subjective cognitive decline. *The Lancet Neurology*, 19(3), 271–278. doi: [10.1016/S1474-4422\(19\)30368-0](https://doi.org/10.1016/S1474-4422(19)30368-0)
- Kaplan E., Goodglass H., & Weintraub S. (1983). *The Boston Naming Test*. Philadelphia, PA: Lea & Fibiger.
- 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. doi: [10.3758/BRM.42.3.643](https://doi.org/10.3758/BRM.42.3.643)
- 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. doi: [10.4306/pi.2017.14.2.126](https://doi.org/10.4306/pi.2017.14.2.126)
- 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. doi: [10.1076/jcen.21.1.127.942](https://doi.org/10.1076/jcen.21.1.127.942)
- Kuperman V., Stadthagen-Gonzalez H., & Brysbaert M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990. doi: [10.3758/s13428-012-0210-4](https://doi.org/10.3758/s13428-012-0210-4)
- 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.
- 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. doi: [10.1080/13825585.2017.1372355](https://doi.org/10.1080/13825585.2017.1372355)
- 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. doi: [10.1177/07399863870092005](https://doi.org/10.1177/07399863870092005)
- 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. doi: [10.3758/bf03193013](https://doi.org/10.3758/bf03193013)
- 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. doi: [10.1111/j.1468-1331.2010.03134.x](https://doi.org/10.1111/j.1468-1331.2010.03134.x)
- McKhann G.M., Knopman D.S., Chertkow H., Hyman B.T., Jack C.R. Jr, Kawas C.H. . . . *et al.* (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. doi: [10.1016/j.jalz.2011.03.005](https://doi.org/10.1016/j.jalz.2011.03.005)
- 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. doi: [10.1159/000327362](https://doi.org/10.1159/000327362)
- Nielsen T.R., Segers K., Vanderaspolden V., Beinhoff U., Minthon L., Pissioti A., & Waldemar G. (2018). Validation of a European Cross-Cultural Neuropsychological Test Battery (CNTB) for evaluation of dementia. *International Journal of Geriatric Psychiatry*, [10.1002/gps.5002](https://doi.org/10.1002/gps.5002)
- 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. doi: [10.1002/gps.2650](https://doi.org/10.1002/gps.2650)
- 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. doi: [10.1016/j.nicl.2018.03.022](https://doi.org/10.1016/j.nicl.2018.03.022)
- 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. doi: [10.1093/arclin/acw007](https://doi.org/10.1093/arclin/acw007)
- 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. doi: [10.1016/j.bandc.2005.09.012](https://doi.org/10.1016/j.bandc.2005.09.012)
- Reis A., Petersson K.M., Castro-Caldas A., & Ingvar M. (2001). Formal schooling influences two- but not three-dimensional naming skills. *Brain and Cognition*, 47(3), 397–411. doi: [10.1006/brcg.2001.1316](https://doi.org/10.1006/brcg.2001.1316)
- 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.
- Román G.C., Tatemichi T.K., Erkinjuntti T., Cummings J.L., Masdeu J.C., Garcia J.H. . . . *et al.* (1993). Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*, 43(2), 250–260. doi: [10.1212/wnl.43.2.250](https://doi.org/10.1212/wnl.43.2.250)
- 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. doi: [10.1007/s00701-013-1638-8](https://doi.org/10.1007/s00701-013-1638-8)
- 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. doi: [10.1037//0278-7393.6.2.174](https://doi.org/10.1037//0278-7393.6.2.174)
- 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.
- Swadesh M. (1952). Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society*, 96, 152–163.
- UNESCO Institute for Statistics (n.d.). Primary education, duration (years). All countries and economies. Retrieved from data.uis.unesco.org
- Verhage F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zeventenveertig jaar*. Proefschrift. Assen: van Gorcum.
- Vivas L., Manoilloff L., Linares N., Zaionz A.F., & Montero L. (2020). Argentinean Psycholinguistic Picture Naming Test in color. *Applied Neuropsychology-Adult*, [10.1080/23279095.2020.1780238](https://doi.org/10.1080/23279095.2020.1780238)
- 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. doi: [10.1080/02687039508248713](https://doi.org/10.1080/02687039508248713)
- 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. doi: [10.1017/S1355617705050897](https://doi.org/10.1017/S1355617705050897)