![CAMBRIDGE UNIVERSITY PRESS]

The role of daily spoken language on the performance of language tests: The Indonesian experience

Heni Gerda Pesau[1], Aria Saloka Immanuel[2], Augustina Sulastri[3] and Gilles van Luijtelaar[4]

[1]Atma Jaya University, Makassar, Sulawesi, Indonesia; [2]Udayana University, Bali, Indonesia; [3]Soegijapranata Catholic University, Semarang, Indonesia and [4]Donders Centre for Cognition, Radboud University, Nijmegen, the Netherlands

Abstract

The performance of cognitive tests is highly dependent on the proficiency of the language in which the tests are administered. Not all Indonesians speak Indonesian daily and many are bilingual. We investigate whether language(s) spoken affects the performance on three language tests in 840 participants ranging in age (16–80) and education (6–22 years). Analysis of covariance followed by Helmert contrasts showed a disadvantage on the Boston Naming Test (BNT) and on the most difficult items of the Token Test for those who do not speak Bahasa daily. Bilinguals did worse on the BNT. Education had a large positive effect on the language tests, age a smaller negative effect. This suggests that besides age and education, the factor of language spoken, either in public or at home, needs to be taken into account when a participant's test scores on the BNT and TT are interpreted and compared with normative data.

Introduction

The accurate assessment of cognitive functioning is crucial in job selection (Hausdorf, LeBlanc, & Chawla, 2003; Outtz, 2002) and prediction of academic success (Peng & Kievit, 2020; Tikhomirova, Malykh, & Malykh, 2020) in healthy subjects, as well as to evaluate cognitive deficits, predict functional outcomes, and monitor patient recovery in various brain injury patients (Casaletto & Heaton, 2017; Mansour & Lajiness-O'Neill, 2015). A variety of cognitive skills such as learning and verbal and visuo-spatial memory, various types of attention, processing speed, reasoning, judgment, mental flexibility, problem-solving, spatial, and language functions among others can be assessed (Kessels & Hendriks, 2016). Language tests are one of the most commonly used tests in clinical settings and elsewhere for the differential diagnosis of aphasia and dementia, as well as in evaluating patients with traumatic brain injury (Maseda et al., 2014).

Normative comparisons are a crucial element in any type of standardized assessment procedure (Huizenga, van Rentergem, Grasman, Muslimovic, & Schmand, 2016). Such assessment involves taking the performance of an individual and comparing that performance to reference groups of the same age, sex, race, and educational attainment. It has become evident that these demographic factors impact performance on the different neuropsychological tests and thus necessitate that the evaluation of the performance of a client or patient be based upon comparisons with like individuals. These normative comparisons allow for a determination of whether an individual is performing as would be expected given this person's demographic characteristics or if their performance is poorer than expected.

Most available normative reference data for neuropsychological tests aimed at the language domain such as the Token Test, Boston Naming Test and Verbal Fluency Test are derived from exclusively or predominantly WEIRD people (Western, Educated, Industrialized, Rich, and Democratic) – that is, Caucasian samples (Henrich, Heine, & Norenzayan, 2010) and monolingual societies. Only recently has the importance of non-WEIRD neuropsychological research been acknowledged (Pathak, Rijal, & Pathak, 2021). It is also true that there is a substantial likelihood that language tests are not culture free and/or that an earlier tested normative group is no longer representative due to changes in health and education level (Siciliano, Chiorri, Battini, Sant'Elia, Altieri, Trojano, & Santangelo, 2018). Moreover, one cannot simply assume that the language tests are valid measures for any population other than the one for which the tests were first developed and normed. It is obvious that lower test performances may be misinterpreted – in, for example, ethnic minority subjects; or for subjects who were not assessed in their first language (Smith, Ivnik, & Lucas, 2008). If outdated or normative scores from other countries or cultures are used, they may be falsely interpreted as indicative

of cognitive impairment and can lead to disproportionate misdiagnosis. Therefore, it is imperative that each country and culture needs recently collected representative normative data.

Indonesia provides an especially challenging environment for the administration of standardized tests as it is enormously heterogeneous regarding daily spoken language and language spoken at home. Over 300 different native languages are spoken in Indonesia. Although Bahasa Indonesia is the official language used in education, mass media, public information, most business, and administration, this language is not the daily language for most Indonesians. Bahasa Indonesia is their first language only for less than 10 percent of the total population; for over 200 million Indonesians their second language. Therefore, it is important to investigate whether speaking Bahasa daily in public or at home does influence the performance of neuropsychological tests and in particular language tests. As so many Indonesians are bilingual, it is also important to know the influence this may have when linguistic processes are assessed.

It is widely acknowledged that language comprises receptive and expressive language functions. Receptive language is the capacity to read, to repeat words and sentences including the understanding of spoken language. Oral comprehension, defined as the ability to process and manipulate information received through speech, can be measured by the Token Test (TT). The TT consists of items such as "Touch the second circle and the first red square" with increasing levels of difficulty. It is affected in patients with Wernicke's aphasia (Jahagirdar, 2014). There are several studies that provide normative data for the TT (Strauss, Sherman & Spreen, 2006) for various Anglo-Saxon and Romance languages (Moreira, Schlottfeldt, de Paula, Daniel, Paiva, Cazita, Coutinho, Salgado, & Malloy-Diniz, 2011; Peña-Casanova, Quiñones-Úbeda, Gramunt-Fombuena, Aguilar, Casas, Molinuevo, Robles, Rodríguez, Barquero, Antúnez, Martínez-Parra, Frank-García, Fernández, Molano, Alfonso, Sol, & Blesa, 2009), but not for Asian languages such as Bahasa Indonesia. The TT is sensitive for left parietal temporal infarcts and moderately sensitive for Alzheimer disease (Paula, Bertola, Nicolato, Moraes, & Malloy-Diniz, 2012). Effects of demographic factors such as age and education on the TT are modest (Peña-Casanova et al., 2009).

The other main category is expressive language: besides spontaneous speech often observed by a clinician during an intake, the naming of objects belongs to this category and the Boston Naming Test (BNT) is often used. It is one of the two most commonly administered tests in Western oriented countries (Kiran, Cherney, Kagan, Haley, Antonucci, Schwartz, Holland, & Simmons-Mackie, 2018). Normative data have been published for various Western cultures and countries (Mitrushina, Boone, Razani, & D'Elia, 2005; Quiñones-Ubeda, Peña-Casanova, Böhm, Gramunt-Fombuena, & Comas, 2004) including Brazil (Leite, Miotto, Nitrini, & Yassuda, 2017). The BNT has been adapted for the Indonesian population (I-BNT) and preliminary normative scores have been proposed (Sulastri, Utami, Jongsma, Hendriks, & van Luijtelaar, 2019). The scores of the BNT are influenced by the level of education while age effects are predominantly found for people over 60 (Peña-Casanova et al., 2009). Similar results were obtained with the I-BNT (Wahyuningrum, van Luijtelaar, & Sulastri, 2021).

Verbal fluency is another aspect of expressive language: it involves coming up with words starting with an obligatory phonological marker (phonemic verbal fluency test) or with words from a semantic category (e.g., animals, vegetables). Phonemic verbal fluency deficits have been noticed in various patient categories and mainly but not exclusively in those with injury to the frontal lobe or fronto-parietal networks (Lim, Kim, Lee, Yoo, Kim, Kim, & Lee, 2019; Manca, Mitolo, Stabile, Bevilacqua, Sharrack, & Venneri, 2019; Pasquier, Lebert, Grymonprez, & Petit, 1995; Klumpp & Deldin, 2010). This is one of the reasons why phonetic verbal fluency can also be considered a marker of executive functioning (Bialystok, Craik, & Luk, 2012; Luo, Luk, & Bialystok, 2010).

The primary aim of our research is to establish whether the scores of these three language function tests are affected by whether or not people speak Bahasa daily in public, and whether or not they speak Bahasa at home. This distinction between language usage at home and in public was earlier used by Sari, van de Vijver, Chasiotis, and Bender (2018) in the Indonesian context. It can be assumed that there will be a disadvantage for those that do not speak Bahasa daily in public. The type of language spoken is relevant since if the negative effects of not speaking Bahasa daily in public or at home are established, then corrections for normative scores for the language function tests will be required.

Although bilingualism is often considered an advantage in many cognitive tasks (Bialystok et al., 2012; Mindt, Arentoft, Kubo Germano, D'Aquila, Scheiner, Pizzirusso, Sandoval, & Gollan, 2008; Schroeder, Marian, Shook, & Bartolotti, 2016), the verbal production of picture naming such as in the BNT or in other similar tasks that require lexical access has been found to be negatively affected in the form of slower responses and/or more errors in bilinguals compared to monolinguals, even though the naming task was done in their dominant language (Costa, 2005; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Michael & Gollan, 2005). This slower word retrieval was independent of factors mediating language proficiency for bilinguals – such as order of acquisition and language dominance (Ivanova & Costa, 2008).

Baldo, Shimamura, Delis, Kramer, and Kaplan (2001) proposed that the phonemic VFT should be considered as an executive function task considering that neuroimaging studies have shown converging evidence for the involvement of executive control in this task – that is, mediated by the left frontal areas, and specifically the posterior opercular area of Broca's area (Paulesu, Goldacre, Scifo, Cappa, Gilardi, Castiglioni, Perani, & Fazio, 1997), an area that is also recruited in cognitive tasks without language production (Yeung, Nystrom, Aronson, & Cohen, 2006). Therefore, considering that phonemic word production involves executive functioning as well as word production, little or no effects of bilingualism can be expected on the VFT. Minimal effects of bilingualism on the most difficult items of the TT are expected: since language comprehension is not known to be affected by speaking a second or third language. In all, the second aim of this study is to establish the effects of bilingualism on the three language tests, among others whether bilinguals will show the expected reduction in the performance of the BNT (Kohnert, Hernandez, & Bates, 1998).

The third aim is to explore the effects of the demographics, such as age, education, and sex on the three language tests and to compare the size of these effects with those of language spoken in public and at home.

Methods

Participants

Participants were recruited by research assistants from six different universities, three on Java Island, one from Bali, East Kalimantan, and south Sulawesi, in all cases in urbanized parts of the four

islands. All participants completed the questionnaire regarding demographic variables such as age, sex, place of birth, years of education, marital status, ethnic group of both parents, and a health questionnaire, regarding the use of alcohol, drugs, or medication, current and past neurological of psychiatric issues, or other factors known to influence health status. The data of participants that reported head trauma, drug abuse, or other previous or current illnesses that might have influenced the performance on the tests were excluded. Respondents were also asked whether they spoke Bahasa Indonesia daily in public or not; and, in the cases where they spoke Bahasa daily in public, it was further asked if they spoke, in addition to Bahasa, either one or two other languages in public. Finally, it was asked whether they spoke Indonesian at home, or if they spoke any other language besides Bahasa Indonesia at home.

The total sample (n = 840) consisted of a rather large range in age (16–80 years, M = 35.5, SD = 15.25) with 61.7% females, and education level, varying from only elementary school (6 years) to more than 17 years (postgraduate with a maximum of 22 years). Participants were categorized into age decade groups as is often done (Fernández & Marcopulos, 2008): (i) age 20–29 years, (ii) age 30–39 years, (iii) age 40–49 years, (iv) age 50–59 years. The data of the group over 60 years were pooled, as well as those 16–19 years of age.

The years of education were categorized as well, and the five categories represented the Indonesian education system: cat 1, educated for less than seven years (i.e., Elementary School (ES)); cat 2, education between 7–9 years encompasses Junior High School (JHS); cat 3, education between 10–12 years (Senior High School (SHS) or equivalent); cat 4, education between 13–16 years (Undergraduate (UG) or equivalent), and cat 5, education more than 16 years (Graduate and Postgraduate).

Table 1 gives an overview of the demographics of the participants. Most of them had at least an undergraduate type of education (53.1%), while 35.7% had completed SHS. A small percentage finished only JHS (6.9%) or ES (4.3%). The age group best represented in our sample were those between 20 to 29 years (35.5%), the three groups between 30 to 59 were equally represented (from 15 to 16.8%), while the youngest and eldest group were less well represented. Culturally our sample represents in large part the urban Javanese population (57%) and the urban parts of Bali, East Kalimantan, and South Sulawesi (each 14%). The sample contained more females than males.

Almost all participants (94.4%) indicated that they were speaking daily Bahasa in public, about half of them were speaking only Bahasa, the other half Bahasa plus one of two other languages. These percentages were rather different for the languages spoken at home: here only 64.2% indicated that they were speaking Bahasa at home, and more than one third of the sample spoke another language at home.

Four groups (G1 to G4) were formed regarding the in-public daily spoken language(s) and three groups (A1 to A3) regarding the language spoken at home (see Table 1). We used what the participants spoke in public (no Bahasa (G1), only Bahasa (G2), Bahasa plus one other language (G3) or Bahasa plus two or more other languages (G4)) and at home (no Bahasa (A1), only Bahasa (A2), Bahasa plus another language (A3)) as independent factors in the various statistical analyses.

Stimuli and Materials

The data used for the present study were taken from recently collected data of three neuropsychological language tests adapted for

Table 1. Demographic characteristics (N = 840)

	M (SD)	Group	n	Percent (%)
Age	35.45 (15.25)	16–19	87	10.4
		20–29	298	35.5
		30–39	141	16.8
		40–49	126	15.0
		50–59	128	15.2
		> 60	60	7.1
Education (in years)	13.74 (3.00)	0–6	36	4.3
		7–9	58	6.9
		10–12	300	35.7
		13–16	399	47.5
		> 17	47	5.6
Sex		Male	322	38.3
		Female	518	61.7
Daily language spoken in public		Other than Indonesian language	47	5.6
		Indonesian	394	46.9
		Indonesian plus one other language	318	37.9
		Indonesian plus two or more languages	81	9.6
Language spoken at Home		Not speaking Bahasa Indonesia	301	35.8
		Bahasa Indonesia	445	53.0
		Bahasa Indonesian plus another language	94	11.2

Note: This table demonstrates age, education, and sex of the participants as well as languages spoken daily, in public and at home in numbers and in percentages.

Indonesia. The letters S, K, T for the phonemic version of the VFT were chosen, as proposed by Hendrawan and Hatta (2010) and applied by Pesau and van Luijtelaar (2021); the for-Indonesia adapted version of the I-BNT (Sulastri et al., 2019), and a from-English translated version of the TT. Data will be available after the normative data of these tests have been published.

Procedure

The tests were administered in Bahasa Indonesia, the official language of Indonesia that is used nationwide in public media, administration, and business. The participants did receive seventy-five thousand rupiahs (equal to five US dollars) after finishing the series of tests. The current research was conducted under the Helsinki Declaration, ethical clearance was provided by the ethics committee of Soegijapranata University under number 001B/B.7.5/FP.KEP/IV/2018). All subjects gave written informed consent.

Research Questions and Data Analyses

We had three major research questions. One: Does the language spoken daily, either in public or at home, affect the performance on the language tests? Two: Are there differences between monolinguals and bilinguals in the performance on the three language tests? Three: Were there effects and what were the effect sizes of sex, age and education on the three language tests?

Differences between the four (language usage in public) and three (language spoken at home) groups regarding their level of education (five levels), age (seven categories) as done previously (Wahyuningrum et al., 2021) and sex were analyzed with two ANOVAs with groups as between subject factor. This was done to establish if these demographic factors should be additionally included in the analyses of the group differences on the three language tests. The analyses of whether the "language groups" G1 - G4 and A1 - A3 differ regarding their performance of verbal tests should not be contaminated by other between group differences such as education, age and sex. Therefore, the analysis of the demographic factors as dependent variables was necessary to establish objectively whether these demographic factors should be included in the ANCOVA or not. Therefore, the ANCOVA was used to establish the effect of language usage daily in public (with groups as between-subjects factor), while controlling for the significant demographic factors age, education, and sex by using them as co-factors. Since sex was without any significant effect, the ANCOVA's were redone without sex as co-factor.

Helmert contrasts were used as post-hoc tests; they directly answered research question 1 and 2. They compared the group that did not speak Bahasa daily in public (G1) with the three groups that did speak Bahasa daily in public (G2, G3, G4) (research question 1); next, comparisons were made between the three groups speaking Bahasa in public: the group that spoke only daily Bahasa (G2, monolinguals) was compared with those that spoke Bahasa plus at least one other language (bilinguals, G3 and G4) (research question 2) , and finally between those that spoke Bahasa in public plus one other language (G3) with the group that spoke, in addition to Bahasa, two other languages in public (G4). In order to prevent type I errors, the p values that were used as representing a significant difference were set at p < .01 for the ANCOVAs.

A similar statistical analysis approach was followed for the factor "language spoken at home". In short, first the demographics variables were analyzed to check whether the three groups differed regarding age, sex, and education, followed by an ANCOVA to control for the effects of demographics. Also, since here sex was without any significant effect, the ANCOVAs were redone without sex as co-factor. The Helmert contrast compared the non-Bahasa speaking group (A1) with the two Bahasa speaking groups (A2 + A3) and the only-Bahasa speaking group (A2) with the group that speaks Bahasa as well as at least one other language in addition to Bahasa (A3). MANOVAs were used to get a general impression regarding the effect size (amount of explained variance) of the two factors (daily language spoken in public and language spoken at home) in relation to the demographic factors of age, education, and sex on the performance of the three verbal tests. The interpretation of effect sizes as expressed by η2 was according to Richardson (2011).

Results

Daily language spoken in public: sample characteristics

The one factor ANOVA regarding whether the four "daily language spoken in public" groups differed in age, education and sex showed significant medium-sized age (F = 15.48, df 3,84, p < .001, $\acute{η}^2$ = .05) and education effects (F = 11.85, df 3.84, p < .001, $\acute{η}2$= .04) and a small sex effect (F = 2.62, df 3.84, p < .001, $\acute{η}^2$= .01).

Subsequent post-hoc tests showed that the participants who did not speak Bahasa daily in public (G1) were significantly less educated and older compared to those who did (G2-G4). Regarding the participants who spoke Bahasa in public it was found that those who only used Bahasa as their daily public language (G2) were significantly younger than those who used Bahasa in addition to another language (G3, G4). Therefore, it is imperative to control for age and education in the comparisons between the four groups regarding the languages spoken daily in public. Finally, it was found that those who use three languages daily were better educated than those who use two daily languages.

Language spoken at home: sample characteristics

The distribution of the subjects regarding the language spoken at home already showed that more than one third (n = 301, 36%) did not speak Bahasa at home. The ANOVA analyzed whether the three groups differed regarding education, age, and sex. A moderate group effect for education was found (F = 30.38, df 2.84, p < .001, $\acute{η}^2$=.07): the post-hoc test confirmed that the education level for those that do not speak Bahasa at home (A1) was less than from the two Bahasa speaking groups (A2, A3). There was also an age effect (F = 15.41, df 2.84, p < .001, $\acute{η}2$ = .04): the post-hoc tests confirmed that the non-Bahasa speaking group (A1) was older than both Bahasa speaking groups (A2, A3) and that the Bahasa speaking group (A2) was younger than the Bahasa plus another language group (A3). Therefore, it is imperative to control for age and education in the comparisons between the three groups regarding the languages spoken at home. In total, the combined demographics of the two language factors showed that 44 of the 840 probands indicated that they speak Bahasa neither at home, nor in public.

The effect sizes of daily spoken language(s) on language tests compared

MANOVAs comparing the effect sizes of daily spoken language in public and language spoken at home were used to get a general impression regarding the effect size of both factors on the performance of the verbal tests as ensemble: the results were F = 3.49, df 54.25, p < .001, $\acute{η}$=.07 for daily language spoken and F = 3.45, df 36.16, p < .001, $\acute{η}$=.07 for languages spoken at home. This demonstrates that the overall contribution of both factors on the performances on all three language tests taken together is significant, moderate sized and equal.

Another MANOVA was used to compare the effect size for the factor daily 'language spoken in public' compared with those of age, education, and sex on the three language tests taken together. The factor 'daily in-public spoken language' was significant with a moderate effect size (F = 2.84, df 54.25, p < .001, $\acute{η}^2$= .06), the effect of education was rather large (F = 19.89, df 18.82, p

$< .001$, $\acute{\eta}^2 = .31$), and the effect of age was moderate ($F = 3.47$, df 18.82, $p < .001$, $\acute{\eta}2 = .07$). The factor sex was not significant. The same analyses were repeated for the comparisons between the effect size of 'language spoken at home' and the three demographic factors. The factor language spoken at home was significant ($F = 1.63$, df 34.14, $p < .01$, $\acute{\eta}2 = .04$), as well as the factor education ($F = 2.70$, df 68.28, $p < .001$, $\acute{\eta}2 = .06$), and age ($F = 1.41$, df 85.35, $p < .01$, $\acute{\eta}2 = = .03$), while sex was not significant.

The effect of daily spoken language(s) in public on the three language tests

We next analyzed whether the factor 'daily in-public spoken language' affected the performance measures of all three language tests separately while controlling for the differences in education and age. The ANCOVA for the factor daily language spoken showed a significant effect for almost all variables of the I-BNT and the most difficult items of the TT. The results are presented in Table 2. The picture emerges that there are quite a few significant group effects, implying that, while controlling for education and age, the language spoken in public affected the scores of two of the three language tests, but in a different manner.

The first research question asked whether there is a disadvantage when tested in Bahasa Indonesia for individuals that do not speak Bahasa in public. The results of the ANCOVA showed group effects for subscale F of the TT ($F = 8.82$, df 3.83, $p < .001$; $\acute{\eta}2 = .03$), and for the Total score of the TT ($F = 7.47$, df 3.83, $p < .001$; $\acute{\eta}2 = .03$), while the outcomes of the first of Helmert's contrast (G1 vs combined groups G2, G3 and G4) confirmed the disadvantage for those that do not speak Bahasa Indonesia daily in public ($p < .01$) for scale F and the total score of the TT. There were no such effects on the two language production tests.

The second research question was answered by the second Helmert contrast in which G2 was compared with G3 and G4. That is, on the I-BNT, is there an advantage for those who are monolinguistic in public versus those who speak one or two other languages in public in addition to Bahasa? This ANCOVA also showed significant group effects for almost all variables of the I-BNT: the F-values for the number of spontaneous correct items were ($F = 4.68$, df 3.83, $p < .001$ $\acute{\eta}2 = .02$), for the number of correct responses after an a-phonemic cue ($F = 9.51$, df 3.83, $p < .001$, $\acute{\eta}2 = .03$), for the number of correct responses after the phonemic cue ($F = 3.60$, df 3.83, $p < .01$, $\acute{\eta}2 = .01$), for the number of total correct responses ($F = 4.49$, df 3.83, $p < .01$, $\acute{\eta}2 = .02$) and for the total time to complete the BNT ($F = 16.19$, df 3.83, $p < .001$; $\acute{\eta}2 = .06$). The Helmert contrast confirmed the advantage of monolingualism for the higher number of spontaneous correct items, for less use of phonemic and a-phonemic cues, as well as a much quicker time to complete the BNT (p's $<.05$). Interestingly, none of the TT and VFT scores were affected. The third contrast, the comparison between the bilingual and trilingual speaking participants, showed two more effects for the BNT: an advantage for the trilinguals compared to the bilinguals for the total number of correct items and for a higher number of a-phonemic cues. The third test, the VFT, did not show differences between any of the groups.

The ANCOVA also confirmed age and education effects, in addition to the mentioned effects of 'language spoken in public'. The results are presented in Table 3. Age and education effects were found for all three language tests. More specifically: all I-BNT variables except the number of correct responses after

phonemic cues showed the age-related differences, the F-scale and Total score for the TT, and all scores for the VFT. All I-BNT variables also showed effects of education, except the number of correct responses after a-phonemic cues, all TT variables, except scale A, and all four variables of the VFT. In fact, more and larger effects (as inferred from $\acute{\eta}^2$) of education compared to age were found (see Table 3). Correlation coefficients (data not given, it would need another table filled with only positive correlations with education and only negative correlations with age) between age and the performance in all three language tests were negative, and they were positive between years of education and performances in all three language tests. In general, age worsened and education improved performance in all three language tests.

The effect of language(s) spoken at home on the three language tests

We analyzed next whether the factor 'language spoken at home' affected the performance measures of all three language tests while controlling for the differences in education and age by using them as covariates. The effects of the demographic factors on the language tests were revealed by the outcomes of the ANOVAs. Rather similar effects of the demographic factors were found for the factor 'spoken language in public'. In particular, significant effects for the factor 'language spoken at home' were found for I-BNT and TT test, but not for the VFT.

The ANCOVA for the factor 'daily language spoken at home' showed significant group effects for most of the I-BNT variables and again on the same TT variables – the score on the F-scale and the total score – and again no effects of this factor were found on the VFT; details of the ANCOVA are presented in Table 4.

Significant group effects for the I-BNT regarding the number of spontaneous correct answers ($F = 9.87$, df 2.84, $p < .001$, $\acute{\eta}2 = .02$), the number of correct answers after-phonemic cues ($F = 5.67$, df 2.84, $p < .01$, $\acute{\eta}2 = .01$), the total number of correct responses ($F = 6.51$, df 2.84, $p < .01$, $\acute{\eta}2 = .02$) as well as total time to complete the BNT ($F = 6.48$, df 2.84, $p < .01$, $\acute{\eta}2 = .02$). All these variables showed the disadvantage as revealed by the first Helmert contrast (A1 vs A2, A3) for not-speaking Bahasa at home: that is, a smaller number of spontaneous and correct items, a slower speed to complete the test, and more used phonemic cues. The same disadvantage was also found for the two variables of the TT: its F-values were $F = 6.80$, df 2.84, $p < .01$, $\acute{\eta}2 = .02$ and $F = 6.39$, df 2.84, $p < .01$, $\acute{\eta}2 = .02$ respectively for the F-scale and Total score, since the first contrast confirmed in both cases worsened performance: fewer correct items for those who do not speak Bahasa at home and were assessed in Bahasa. The second contrast concerned the disadvantage of bilingualism: here the total score of the TT was affected (for F see above) and a lower score for the bilingual individuals was obtained. No group effects were found for the VFT.

The ANCOVA showed, in addition to the previously mentioned group effects, age and education effects as well for all three tests. The results are presented in Table 5. It is particularly striking that almost all variables of the three tests show the effects of education, with the size of this effect being larger than that of age. Striking as well is the finding that all subscales of the TT except one are sensitive to education effects. There were also rather large effects for education on the VFT.

Table 2. The effect of daily spoken language(s) in public on the three language tests while controlling for age, and education as covariates (ANCOVA).

Language Tests	G1	G2	G3	G4	Mean Square	F (3,83)	Sig.	Partial Eta Squared	Group differences
Boston-Naming Test (I-BNT)									
BNT Spontaneous	46.7 (8.7)	51.0 (5.6)	48.2 (7.6)	48.6 (7.8)	173.72	4.68	.00	.02	G2 > (G3, G4)
BNT A-phonemic	3.5 (2.8)	2.2 (2.3)	2.3 (2.6)	3.7 (3.6)	61.27	9.51	.00	.03	G2 < (G3, G4) G4 > G3
BNT Phonemic	3.3 (2.9)	2.4 (2.3)	3.1 (2.8)	3.2 (2.6)	23.16	3.60	.01	.01	G2 < (G3,G4)
BNT Total Score	53.6 (6.1)	55.6 (4.2)	53.6 (5.4)	55.5 (5.0)	78.52	4.49	.00	.01	G3 < G4
BNT Total Error	7.9 (10.4	5.4 (6.4)	7.5 (6.5)	5.6 (5.5)	91.75	2.29	.08	.00	
BNT Total Time	453.5 (361.5)	338.6 (200.4	502.3 (295.7)	432.4 (277.4)	884177.53	16.19	.00	.06	G2 < (G3, G4)
Token Test (TT)									
TT Part A	6.9 (.5)	6.9 (.6)	6.9 (.8)	6.9 (.4)	.16	.35	.79	.00	
TT Part B	7.8 (.9)	7.9 (.5)	7.9 (.6)	7.9 (.6)	.07	.21	.89	.00	
TT Part C	11.9 (.5)	11.9 (.7)	11.7 (1.1)	11.8 (.6)	.91	1.4	.25	.00	
TT Part D	15.5 (1.4)	15.6 (1.7)	15.4 (2.0)	15.4 (1.3)	1.00	.33	.81	.00	
TT Part E	22.5 (3.1)	23.4 (1.9)	23.2 (2.9)	23.8 (.7)	9.85	1.88	.13	.00	
TT Part F	68.6 (27.1	85.0 (15.9	80.3 (19.0	86.6 (13.0	2543.73	8.62	.00	.03	G1 < (G2,G3,G4)
TT Total Score	133.0 (28.0	150.7 (17.7)	145.4(22.2	152.4(13.4	2739.88	7.47	.00	.03	G1 < (G2,G3,G4)
Verbal Fluency Test									
VFT Letter S Correct	11.9 (4.8)	13.4 (5.5)	12.4 (5.2)	13.2 (5.1)	6.03	.25	.86	.00	
VFT Letter K Correct	12.1 (5.1)	14.1 (5.2)	13.3 (4.8)	14.4 (4.4)	5.80	.28	.84	.00	
VFT Letter T Correct	11.0 (4.7)	12.2 (5.0)	11.7 (4.8)	12.6 (4.8)	10.91	.54	.65	.00	
VFT Total Correct	35.0 (13.4	39.7 (13.9)	37.3 (13.3	40.2 (12.8	29.85	.20	.90	.00	
Note:									
G1	Speak in public other than Indonesian Language								
G2	Speak Indonesian in public daily								
G3	Speak Indonesian plus one other language daily in public								
G4	Speak Indonesian plus two or more languages daily in public								

Note: This table shows the effect of daily spoken language(s) in public on the Boston Naming Test (BNT), Token Test (TT), and Verbal Fluency Test (VFT) while controlling for age, and education as covariates (ANCOVA). Given are the Mean and SD of the four groups regarding the spoken language(s). The Mean Square, F (with df's), p-value and partial eta squared (effect size), followed by the outcomes of the Helmert contrasts as post-hoc tests.

Discussion

Indonesia is truly a society in which many people use more than one language and this is represented in our sample. While almost 95% speak Bahasa Indonesia in public, at home the percentage of Bahasa-speaking participants is much lower (64%). Next, many (47.5%) Indonesians speak another language in addition to Bahasa in public. Those that did not speak Bahasa in public were somewhat older and less educated than the Bahasa speaking groups. Among the Bahasa speaking groups it was the youngsters who only spoke Bahasa; the older participants were more often bilingual. The same tendencies were found regarding the language spoken at home: the youngsters and well-educated people more often spoke Bahasa at home compared to older and less well-educated individuals. This result fits in the general tendency that most of the population speak Indonesian next to their local language and that Indonesian is becoming more common as a

first language (Cohn & Ravindranath, 2014). Under the assumption that middle class families are better educated than their working-class counterparts, the outcomes of our study agree with an earlier Indonesian study showing that middle class parents and children in Central Java use Bahasa much more than their working-class counterparts (Kurniasih, 2006).

The level of education and the age of the participants across the groups was not the same and considering that it is well documented that age and education have major effects on many cognitive tests, including language tests, it is imperative to control for age and education when the groups' linguistic performance is compared. The major outcome of the comparisons between the language groups is that language spoken at home has a moderate effect on the performance of two of the three language tests. More specifically, a clear disadvantage was found for participants who do not speak Bahasa Indonesia at home on a language

Table 3. The AGE and EDUCATION effects as obtained in the ANCOVA towards the effects of daily spoken languages in public on the three language tests.

Language Tests	AGE effects			EDUCATION effects		
	F (1,83)	P -value	Partial Eta. Squared	F (1,83)	P -value	Partial Eta. Squared
Boston-Naming Test (I-BNT)						
BNT Spontaneous	24.66	0.00	0.03	160.60	0.00	0.16
BNT Aphonemic	20.91	0.00	0.02	3.83	0.05	0.01
BNT Phonemic	0.49	0.49	0.00	8.19	0.00	0.01
BNT Total Score	16.29	0.00	0.02	241.48	0.00	0.23
BNT Total Error	10.35	0.00	0.01	62.71	0.00	0.07
BNT Total Time	15.07	0.00	0.02	144.77	0.00	0.15
Token Test (TT)						
TT Part A	0.29	0.59	0.00	0.21	0.64	0.00
TT Part B	2.96	0.09	0.00	21.15	0.00	0.03
TT Part C	6.03	0.01	0.01	35.86	0.00	0.04
TT Part D	0.95	0.33	0.00	18.45	0.00	0.02
TT Part E	4.30	0.04	0.01	15.03	0.00	0.02
TT Part F	7.06	0.01	0.01	36.93	0.00	0.04
TT Total Score	8.05	0.01	0.01	44.64	0.00	0.05
Verbal Fluency Test (VFT)						
VFT Letter S	14.72	0.00	0.02	98.46	0.00	0.11
VFT Letter K	10.30	0.00	0.01	135.19	0.00	0.14
VFT Letter T	19.20	0.00	0.02	102.20	0.00	0.11
VFT Total Correct	18.96	0.00	0.02	145.23	0.00	0.15

Note: F with df's, and p-value and effect size are given.

comprehension test and on a word production test (I-BNT) when assessed in Bahasa Indonesia. The same disadvantage was also found for those that do not speak Bahasa Indonesia in public on the same language comprehension test, but not for the I-BNT.

A second major outcome was that bilingualism is a disadvantage in word production as measured with the I-BNT but not with the VFT. Furthermore, this bilingualism disadvantage effect, or monolingualism advantage, was found for both speaking only Bahasa at home and for only speaking Bahasa in public. Thirdly, large education and moderate age effects were found for all three language tests. Sex was not found to have any influence on any of the performance measures of the three language tests.

The disadvantage of not being tested in the daily language one speaks in public was evident for the language comprehension test but only for the total score of the TT. This disadvantage of being assessed in Bahasa while Bahasa is not the language one regularly uses in public was confirmed and extended by the analyses regarding the languages spoken at home: again, a lower TT performance, not only on the test's F-scale (the most difficult scale) and the total score, but now also on the I-BNT. Almost all I-BNT variables showed this disadvantage: both language comprehension and word production were negatively affected. Both the time to complete the task and the number of correct items showed this effect. Interestingly, such an effect was not found for the other word production test, the VFT. Other research in

this area showed that Swedish immigrants when tested in the Swedish language performed worse on a phonemic VFT task compared to age-matched native Swedes, demonstrating that not being tested in one's original language lowers the performance scores in word production (Stålhammar, Hellström, Eckerström, & Wallin, 2020). The reasons for the lack of effects on the VFT test in the present study might be: that although Bahasa is not spoken daily, the familiarity of the subjects with Bahasa or the proficiency in Bahasa might be high, considering that almost all Indonesians are educated in Bahasa for nine years; that the majority of our subjects finished at least senior high school or college (undergraduate); and that the large exposure to mass media has a large and compensatory impact on their ability to produce words in Bahasa Indonesia. It also seems that the BNT, with its increase in the level of difficulty of the items, is more challenging than the VFT for the detection of word production performance.

The disadvantage in word production of bilinguals found specifically for the I-BNT was earlier found by others for the BNT both regarding the decrease in the number of correct items or in a lengthening of the time to complete the BNT (Bialystok, Craik, & Luk, 2008; Luo et al., 2010; Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007; Sandoval, Gollan, Ferreira, & Salmon, 2010). Many other studies found that bilingualism appears to slow down lexical processing as measured in VF tasks, including those constrained by letter-cues (Gollan, Montoya, & Werner, 2002; Portocarrero, Burright, & Donovick, 2007). The longer retrieval

Table 4. The effect of spoken language at home while controlling for age, and education (ANCOVA) on the three language tests.

Language Tests	A1	A2	A3	Mean Square	F (2,84)	Sig.	Partial Eta Squared	Group differences
Boston-Naming Test (I-BNT)								
BNT Spontaneous	47.1 (7.9)	50.7 (6.1)	51.4 (5.5)	363.44	9.87	.00	.02	A1 < (A2,A3)
BNT A-phonemic	2.7 (2.6)	2.4 (2.8)	1.8 (1.9)	21.94	3.32	.04	.01	A1 > (A2,A3) A2 > A3
BNT Phonemic	3.3 (2.9)	2.5 (2.3)	2.7 (2.3)	36.44	5.67	.00	.01	A1 > (A2,A3)
BNT Total Score	53.1 (5.6)	55.6 (4.3)	55.9 (4.2)	113.73	6.51	.00	.02	A1 < A2,A3
BNT Total Error	7.4 (6.9)	5.4 (6.0)	7.2 (8.5)	107.26	2.68	.07	.01	
BNT Total Time	499.7 (320.0)	370.6 (228.6)	363.1 (187.9)	368416.10	6.48	.00	.02	A1 > (A2,A3)
Token Test (TT)								
TT Part A	6.9 (.7)	6.9 (.7)	6.88 (.75)	.01	.03	.97	.000	
TT Part B	7.9 (.6)	7.9 (.5)	7.82 (.82)	.46	1.44	.24	.00	
TT Part C	11.8 (.9)	11.9 (.6)	11.67 (1.39)	1.30	1.93	.15	.01	
TT Part D	15.4 (1.9)	15.6 (1.4)	15.38 (2.64)	2.09	.69	.50	.00	
TT Part E	23.1 (2.8)	23.5 (1.5)	23.16 (3.58)	4.56	.87	.42	.00	
TT Part F	78.0 (21.1)	85.6 (14.0)	81.6 (22.0)	2034.11	6.80	.00	.02	A1 < (A2, A3); A2 > A3
TT Total Score	143.0 (23.3)	151.5 (14.9)	146.5 (27.6)	2351.80	6.348	.00	.02	A1 < (A2 + A3); A2 > A3
Verbal Fluency Test (VFT)								
VFT Letter S Correct	11.8 (5.0)	13.4 (5.5)	13.8 (5.0)	28.79	1.19	.30	.00	
VFT Letter K Correct	12.7 (5.0)	14.2 (4.9)	14.6 (4.8)	20.15	.98	.38	.00	
VFT Letter T Correct	11.2 (4.6)	12.3 (5.0)	12.5 (4.7)	5.76	.29	.75	.00	
VFT Total Correct	35.8 (13.1)	39.9 (13.8)	40.8 (12.9)	130.16	.88	.42	.00	
Note:								
A1	Not speaking Indonesian at home							
A2	Speaking Indonesian at home							
A3	Speaking Indonesian at home plus another language							

Note: This table demonstrates the effect of spoken language at home while controlling for age, and education (ANCOVA) on the Boston Naming Test (BNT), Token Test (TT), and Verbal Fluency Test (VFT). Given are the Mean and SD of the three groups (A1, A2, A3), the MS, F (with df's) p and eta squared, followed by the outcomes of the Helmert contrasts as post-hoc tests.

times needed by bilinguals have been interpreted as a consequence of competition or interference between lexical entries of the bilinguals' two languages, or, alternatively, to lower word usage frequencies. And there is indeed quite a body of evidence that, when a bilingual individual is assessed: that lexical access is influenced by properties of the lexicon of the other language (Dijkstra & van Heuven, 2002); that the lexicon of the two languages is not accessed separately; and that between-language similarities regarding phonological, semantic and orthographic similarities can trigger the activation of the other language (Sanoudaki & Thierry, 2015).

Interestingly, here again the second word production test, the VFT, was not sensitive, but this lack of effects was predicted (Bialystok et al., 2012; Gollan et al., 2002; Luo et al., 2010; Rosselli, Ardila, Salvatierra, Marquez, Luis, & Weekes, 2002). The phonemic VFT is also considered an executive function task and bilinguals should therefore have here an advantage above monolinguals. Therefore, the lack of a bilingual disadvantage in the VFT task as found here may be a consequence of the interplay between the linguistic disadvantages and the

executive control advantages in bilinguals (Sandoval et al., 2010). Put in other words, the bilingual disadvantage in word production might have been compensated for by the bilingual advantage in executive function. The bilingual disadvantage for the I-BNT as found here is not caused by age and education effects, as demonstrated by the outcomes of the ANCOVA (where the effects persisted after controlling for these demographics), although it is plausible that some non-language aspects on which bilingual and monolingual speakers might differ such as parents' education level, socioeconomic status, and ethnicity could also have contributed to a bilingual disadvantage, as was proposed by Antoniou (2018).

A comparison of individuals who speak either one or two languages in addition to Bahasa revealed a significant difference in just one I-BNT variable: the number of a-phonemic cues. This type of cue was used most often by those who speak two more languages in addition to Bahasa, followed by those who speak a single additional language and then those who speak only Bahasa. Young (2016) found that the recall time of a picture-

Table 5. The AGE effects and EDUCATION effects in the ANCOVA towards the effects of daily language spoken at home on the three language tests.

Language Tests	The AGE effects			The EDUCATION effects		
	F (1,83)	P -value	Partial Eta. Squared	F (1,83)	P -value	Partial Eta. Squared
Boston-Naming Test (BNT)						
BNT Spontaneous	26.96	0.00	0.03	134.87	0.00	0.14
BNT Aphonemic	24.54	0.00	0.03	2.04	0.15	0.00
BNT Phonemic	0.63	0.43	0.00	4.48	0.04	0.01
BNT Total Score	16.01	0.00	0.02	215.31	0.00	0.21
BNT Total Error	10.90	0.00	0.01	63.22	0.00	0.07
BNT Total Time	18.33	0.00	0.02	122.53	0.00	0.13
Token Test (TT)						
TT Part A	0.27	0.60	0.00	0.27	0.61	0.00
TT Part B	2.83	0.09	0.00	20.92	0.00	0.02
TT Part C	5.27	0.22	0.01	35.25	0.00	0.04
TT Part D	0.72	0.40	0.00	15.69	0.00	0.02
TT Part E	4.60	0.03	0.01	15.02	0.00	0.02
TT Part F	8.45	0.00	0.01	33.41	0.00	0.04
TT Total Score	9.34	0.00	0.01	40.81	0.00	0.05
Verbal Fluency Test (VFT)						
VFT Letter S	14.43	0.00	0.02	88.90	0.00	0.10
VFT Letter K	10.40	0.00	0.01	129.06	0.00	0.13
VFT Letter T	18.33	0.00	0.02	94.30	0.00	0.10
VFT Total Correct	18.56	0.00	0.02	134.71	0.00	0.14

Note: F with df's, p-value, and effect size are given

naming task increased gradually as the number of languages increased. Their evidence supports a hypothesis that learning more languages does increase difficulty with word retrieval, the more languages a person knows, and the slower he or she recalls individual words phonetically and semantically. Getting access to a broader lexicon and finding the appropriate word seems more difficult and cues are more necessary compared to those who only speak Bahasa.

All three language tests were more sensitive for education than for age with moderate effects, while sex effects were not found on any of the performance measures of the three language tests (Lucas, Ivnik, Smith, Ferman, Willis, Petersen, & Graff-Radford, 2005; Scheuringer, Wittig, & Pletzer, 2017). The relatively small, though significant, age effects on the TT are also in agreement with literature (e.g., Aranciva, Casals-Coll, Sánchez-Benavides, Quintana, Manero, Rognoni, Calvo, Palomo, Tamayo, & Peña-Casanova, 2012; Lucas et al., 2005; Moreira et al., 2011) showing that verbal comprehension is usually better preserved during ageing than the productive aspects of language, such as word fluency and picture naming, but some difficulties are encountered when the material is complex (Juncos-Rabadán, Facal, Rodríguez, & Pereiro, 2010) mainly due to a diminishment in the phonemic loop of working memory. Also interesting, in our case, the age-related effects on the TT were found for the more difficult C, E, F scales and Total score, but not for the easiest ones (A and B).

Corrections for age and education have become common practice for cognitive tests when normative scores are reported. As mentioned, effects of education and age on all three language tests were found, in agreement with a large body of literature. The size of the reported effects of the languages spoken at home on the language tests are comparable in size with those of age and cannot be denied. Therefore, it is more than reasonable that corrections for language spoken will be incorporated in the normative scores, next to education and age effects. The present outcomes have the practical consequence that when assessed with Indonesian language tests, different normative data should be used for those participants that do not speak Indonesian at home, as well, perhaps, for participants that do not use Indonesian daily in public. This is the case for the I-BNT and to a lesser degree for the Token Test, but not for the VFT. No effects of spoken language were found for the VFT when we controlled for age and education effects. The lack of spoken language effects for the VFT may suggest that age and education have mediating effects on the factors of language spoken at home and in public, and indeed those who did not speak Bahasa at home were older and less educated.

In clinical neuropsychological practice, persons may often be tested in a language different from their home language. Those people who do not have Bahasa as a home language, but who are assessed in Bahasa, have an immediate disadvantage. Putative lower scores on the I-BNT and TT might lead to

underestimation of their linguistic skills, or in an overdiagnosis in cases of pathology. It remains to be determined whether corrections for ethnicity are imperative as well. This awaits further empirical research addressing this issue. It is imperative regarding a fair assessment that the clients' and patients' diversity be respected by offering them ethnicity-adapted tests, as well as providing recent normative data from groups that mimic their demographic characteristics as closely as possible. A fair assessment implies as well that one is tested in one's primary language and by a representative of one's own culture or ethnic group. This is necessary to prevent neuropsychological health-care disparities and underestimation of a persons' cognitive skills (Cory, 2021; Rabin, Brodale, Elbulok-Charcape, & Barr, 2020).

It can be concluded that the two factors "daily language" and "language spoken at home" that were explored in this study have rather similar effects on the three different language tests, although the factor language spoken at home seems more influential. The effects have consequences for clinical practices. First, and this might be preferable, to avoid underdiagnosis or overdiagnosis of language problems due to the language effects found here, the tests should be administered in one's most familiar language. When this is not possible or when the tests have not been properly adapted for the first language of the testee, then the tests can be administered by a neuropsychologist with an interpreter or else directly administered by multilingual research assistants (Nielsen, Segers, Vanderaspoilden, Bekkhus-Wetterberg, Minthon, Pissiota, Bjørkløf, Beinhoff, Tsolaki, Gkioka, & Waldemar, 2018). Secondly, consideration needs to be given to correcting the normative data for the TT and I-BNT for "language spoken at home." Such correction is particularly called for if the cognitive assessment is done in Bahasa Indonesia and Bahasa is not the language spoken at home. These corrections of the scores would be similar to the well-known corrections for age and education. For the phonemic VFT corrections for age and education are all that seem imperative in this Indonesian sample.

Competing interests. The author(s) declare none

References

Antoniou M (2018) The advantages of bilingualism debate. *Annual Review of Linguistics* 5, 395–415. https://doi.org/10.1146/annurev-linguistics-011718-011820

Aranciva F, Casals-Coll M, Sánchez-Benavides G, Quintana M, Manero RM, Rognoni T, Calvo L, Palomo R, Tamayo F and Peña-Casanova J (2012) Estudios normativos españoles en población adulta joven (Proyecto NEURONORMA jóvenes): normas para el Boston Naming Test y el Token Test. *Neurología* 27, 394–399. https://doi.org/10.1016/j.nrl.2011.12.016

Baldo J, Shimamura A, Delis D, Kramer J and Kaplan E (2001) Verbal and design fluency in patients with frontal lobe lesions. *Journal of the International Neuropsychological Society* 7, 586–596. http://doi.org/10.1017/S1355617701755063

Bialystok E, Craik F and Luk G (2008) Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34, 859–873. http://doi.org/10.1037/0278-7393.34.4.859

Bialystok E, Craik FI and Luk G (2012) Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences* 16, 240–50. http://doi.org/10.1016/j.tics.2012.03.001

Casaletto KB and Heaton RK (2017) Neuropsychological Assessment: Past and future. *Journal of International Neuropsychological Society* 23, 778–790. http://doi.org/10.1017/S1355617717001060

Cohn AC and Ravindranath M (2014) Local languages in Indonesia: Language maintenance or language shift. *Linguistik Indonesia* 32, 131–148. https://doi.org/10.26499/LI.V32I2.22

Cory JM (2021) White privilege in neuropsychology: An 'invisible knapsack' in need of unpacking? *The Clinical Neuropsychologist* 35, 206–218. https://doi.org/10.1080/13854046.2020.1801845

Costa A (2005) Lexical access in bilingual production. In *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford University Press, pp. 289–307.

Dijkstra T and Van Heuven W (2002) The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* 5, 175–197. http://doi.org/10.1017/S1366728902003012

Fernández AL and Marcopulos BA (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, 239–246. https://doi.org/10.1111/j.1467-9450.2008.00637.x

Gollan TH, Montoya RI and Werner GA (2002) Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology* 16, 562–576. https://doi.org/10.1037/0894-4105.16.4.562

Gollan TH, Montoya RI, Fennema-Notestine C and Morris SK (2005) Bilingualism affects picture naming but not picture classification. *Memory & Cognition* 33, 1220–1234. https://doi.org/10.3758/BF03193224

Gollan TH, Fennema-Notestine C, Montoya RI and Jernigan TL (2007) The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society* 13, 197–208. https://doi.org/10.1017/S1355617707070038

Hausdorf PA, LeBlanc MM and Chawla A (2003) Cognitive ability testing and employment selection: Does test content relate to adverse impact? *Applied H.R.M. Research* 7, 41–48.

Hendrawan D and Hatta T (2010) Evaluation of stimuli for development of the Indonesian version of verbal fluency task using ranking method. *Psychologia* 53, 14–26. https://doi.org/10.2117/psysoc.2010.14

Henrich J, Heine S and Norenzayan A (2010) The weirdest people in the world? *Behavioral and Brain Sciences* 33, 61–83. https://doi.org/10.1017/s0140525×0999152x

Huizenga HM, van Rentergem JAA, Grasman RPPP, Muslimovic D and Schmand B (2016) Normative comparisons for large neuropsychological test batteries: User friendly and sensitive solutions to minimize familywise false positives. *Journal of Clinical and Experimental Neuropsychology* 38, 611–629. http://dx.doi.org/10.1080/13803395.2015.1132299

Ivanova I and Costa A (2008) Does bilingualism hamper lexical access in speech production? *Acta Psychologica* 127, 277–88. http://doi.org/10.1016/j.actpsy.2007.06.003

Jahagirdar V (2014) Hemispheric differences: The bilingual brain. In *Foundations of bilingual memory*. NY: Springer, pp. 257–282 http://dx.doi.org/10.1007/978-1-4614-9218-4_13

Juncos-Rabadán O, Facal D, Rodríguez MS and Pereiro AX (2010) Lexical knowledge and lexical retrieval in ageing: Insights from a tip-of-the-tongue (TOT) study. *Language and Cognitive Processes* 25, 1301–1304. http://doi.org/10.1080/01690961003589484

Kessels RPC and Hendriks MPH (2016) Neuropsychological assessment. In *Encyclopedia of mental health (2nd ed.)*. Oxford: Elsevier, pp. 197–201.

Kiran S, Cherney LR, Kagan A, Haley KL, Antonucci SM, Schwartz M, Holland AL and Simmons-Mackie N (2018) Aphasia assessments: a survey of clinical and research settings. *Aphasiology* 32(sup1), 47–49. https://doi.org/10.1080/02687038.2018.1487923

Klumpp H and Deldin P (2010) Review of brain functioning in depression for semantic processing and verbal fluency. *International Journal of Psychophysiology* 75:77–85. https://doi.org/10.1016/j.ijpsycho.2009.10.003

Kohnert KJ, Hernandez AE and Bates E (1998) Bilingual Performance on the Boston Naming Test: Preliminary Norms in Spanish and English. *Brain and Language* 65, 422–440. https://doi.org/10.1006/brln.1998.2001

Kurniasih Y (2006) Gender, class and language preference: A case study in Yogyakarta. In *Selected papers from the 2005 Conference of the Australian Linguistic Society.* http://www.als.asn.au/proceedings/als2005/kurniasih.gender.pdf

Leite K, Miotto E, Nitrini R and Yassuda M (2017) Boston Naming Test (BNT) original, Brazilian adapted version and short forms: Normative data for illiterate and low-educated older adults. *International Psychogeriatrics* 29, 825–833. http://doi.org/10.1017/S1041610216001952

Lim KB, Kim J, Lee HJ, Yoo J, Kim HS, Kim C and Lee H (2019) COWAT Performance of Persons with Alzheimer Dementia, Vascular Dementia, and Parkinson Disease Dementia According to Stage of Cognitive Impairment. *Physical Medicine and Rehabilitation* 11, 737–744. https://doi.org/10.1002/pmrj.12125

Lucas JA, Ivnik RJ, Smith GE, Ferman TJ, Willis FB, Petersen RC and Graff-Radford NR (2005) Mayo's Older African Americans Normative Studies: Norms for Boston Naming Test, Controlled Oral Word Association, Category Fluency, Animal Naming, Token Test, Wrat-3 Reading, Trail Making Test, Stroop Test, and Judgment of Line Orientation. *The Clinical Neuropsychologist* 19, 243–269. https://doi.org/10.1080/13854040590945337

Luo L, Luk G and Bialystok E (2010) Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition* 114, 29–41. https://doi.org/10.1016/j.cognition.2009.08.014

Manca R, Mitolo M, Stabile MR, Bevilacqua F, Sharrack B and Venneri A (2019) Multiple brain networks support processing speed abilities of patients with multiple sclerosis. *Postgraduate Medicine* 131, 523–532. http://doi.org/10.1080/00325481.2019.1663706

Mansour A and Lajiness-O'Neill R (2015) Call for an integrative and multi-disciplinary approach to traumatic brain injury (TBI). *Psychology* 6, 323–374. https://doi.org/10.4236/psych.2015.64033

Maseda A, Lodeiro-Fernández L, Lorenzo-López L, Núñez-Naveira L, Balo A and Millán-Calenti JC (2014) Verbal fluency, naming and verbal comprehension: three aspects of language as predictors of cognitive impairment, *Aging & Mental Health* 18, 1037–1045, https://doi.org/10.1080/13607863.2014.908457

Michael EB and Gollan TH (2005) Being and becoming bilingual: Individual differences and consequences for language processing. In *Handbook of Bilingualism: Psycholinguistic Approaches*, pp 389–409.

Mindt MR, Arentoft A, Kubo Germano K, D'Aquila E, Scheiner D, Pizzirusso M, Sandoval TC and Gollan TH (2008) Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. *Neuropsychology Review* 18, 255–268. https://doi.org/10.1007/s11065-008-9069-7

Mitrushina M, Boone KB, Razani J and D'Elia LF (2005) *Handbook of normative data for neuropsychological assessment*, 2nd Ed. New York: Oxford University Press.

Moreira L, Schlottfeldt CG, de Paula JJ, Daniel MT, Paiva A, Cazita V, Coutinho G, Salgado JV and Malloy-Diniz LF (2011) Normative study of the Token Test (short version): preliminary data for a sample of Brazilian seniors. *Archives of Clinical Psychiatry* (*São Paulo*) 38:97–101. https://doi.org/10.1590/S0101-60832011000300003

Nielsen TR, Segers K, Vanderaspoilden V, Bekkhus-Wetterberg P, Minthon L, Pissiota A, Bjørkløf GH, Beinhoff U, Tsolaki M, Gkioka M and 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*, 1411–1430. https://doi.org/10.1080/13854046.2018.1430256

Outtz JL (2002) The role of cognitive ability tests in employment selection. *Human Performance* 15, 161–171. http://dx.doi.org/10.1080/08959285.2002.9668089

Pasquier F, Lebert F, Grymonprez L, and Petit H (1995) Verbal fluency in dementia of frontal lobe type and dementia of Alzheimer type. *Journal of Neurology, Neurosurgery & Psychiatry* 58, 81–84. http://doi.org/10.1136/jnnp.58.1.81

Pathak LS, Rijal S and Pathak P (2021) Instruction in second language enhances linguistic and cognitive abilities in first language as well: evidence from public school education in Nepal. *Journal of Cultural Cognitive Science* 5, 287–310, https://doi.org/10.1007/s41809-021-00084-7

Paula JJ de., Bertola L, Nicolato R, Moraes EN de., and Malloy-Diniz LF (2012) Evaluating Language Comprehension in Alzheimer's disease: the

use of the Token Test. *Arquivos de Neuro-Psiquiatria* 70, 435–440. http://doi.org/10.1590/s0004-282×2012000600010

Paulesu E, Goldacre B, Scifo P, Cappa SF, Gilardi MC, Castiglioni I, Perani D and Fazio F (1997) Functional heterogeneity of left inferior frontal cortex as revealed by fMRI. *Neuroreport.* 27; 8:2011–7. http://doi.org/10.1097/00001756-199705260-00042.

Peña-Casanova J, Quiñones-Úbeda S, Gramunt-Fombuena N, Aguilar M, Casas L, Molinuevo JL, Robles A, Rodríguez D, Barquero SM, Antúnez C, Martínez-Parra C, Frank-García A, Fernández M, Molano A, Alfonso V, Sol JM and Blesa R (2009) Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for Boston Naming Test and Token Test. *Archives of Clinical Neuropsychology* 24, 343–354. https://doi.org/10.1093/arclin/acp039

Peng P and Kievit RA (2020) The development of academic achievement and cognitive abilities: A bidirectional perspective. *Child Development Perspectives* 14, 15–20. http://doi.org/10.1111/cdep.12352

Pesau HG and van Luijtelaar G (2021) Equivalence of Traditional and Internet-Delivered Testing of Word Fluency Tasks, *Jurnal Psikologi* 20, 35–49. https://doi.org/10.14710/jp.20.1.35-49

Portocarrero JS, Burright RG and Donovick PJ (2007) Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology* 22, 415–422. http://doi.org/10.1016/j.acn.2007.01.015

Quiñones-Ubeda S, Peña-Casanova J, Böhm P, Gramunt-Fombuena N and Comas L (2004) Preliminary normative data for the second edition of the Boston Naming Test for young Spanish adults. *Neurologia* 19, 248–253.

Rabin LA, Brodale DL, Elbulok-Charcape MM and Barr WB (2020) Challenges in the neuropsychological assessment of ethnic minorities. In *Clinical cultural neuroscience: An integrative approach to cross-cultural neuropsychology.* Oxford University Press, pp. 55–80.

Richardson JTE (2011) Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6, 135–147. https://doi.org/10.1016/j.edurev.2010.12.001.

Rosselli M, Ardila A, Salvatierra J, Marquez M, Luis M and Weekes VA (2002) A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience* 112, 759–776. https://doi.org/10.1080/00207450290025752

Sandoval TC, Gollan TH, Ferreira VS and Salmon DP (2010) What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition* 13, 231–252. https://doi.org/10.1017/S1366728909990514

Sanoudaki E and Thierry G (2015) Language non-selective syntactic activation in early bilinguals: the role of verbal fluency. *International Journal of Bilingual Education and Bilingualism* 18, 548–560. https://doi.org/10.1080/13670050.2015.1027143

Sari BT, van de Vijver FJR, Chasiotis A and Bender M (2018) Contextualized bilingualism among adolescents from four different ethnic groups in Indonesia. *International Journal of Bilingualism* 27: 1. http://doi.org/10.1177/1367006918803678

Scheuringer A, Wittig R and Pletzer B (2017) Sex differences in verbal fluency: the role of strategies and instructions. *Cognitive Processing* 18, 407–417. https://doi.org/10.1007/s10339-017-0801-1

Schroeder SR, Marian V, Shook A and Bartolotti J (2016) Bilingualism and Musicianship Enhance Cognitive Control. *Neural Plasticity*, 1–11. https://doi.org/10.1155/2016/4058620

Siciliano M, Chiorri C, Battini V, Sant'Elia V, Altieri M, Trojano L and Santangelo G (2018) Regression-based normative data and equivalent scores for Trail Making Test (TMT): an updated Italian normative study. *Neurological Sciences.* https://doi.org/10.1007/s10072-018-3673-y

Smith GE, Ivnik RJ and Lucas J (2008) Assessment techniques: tests, test batteries, norms and methodological approaches. In *Textbook of Clinical Neuropsychology*, pp 38–58

Stålhammar J, Hellström P, Eckerström C and Wallin A (2020) Neuropsychological Test Performance among Native and Non-Native Swedes: Second Language Effects. *Archives of Clinical Neuropsychology* 00, 1–13. https://doi.org/10.1093/arclin/acaa043

Strauss E, Sherman EMS and Spreen OA. (2006) *Compendium of neuropsychological tests. Administration, norms, and commentary.* New York: Oxford University Press.

Sulastri A, Utami MSS, Jongsma M, Hendriks M and van Luijtelaar G (2019) The Indonesian Boston Naming Test: Normative data among healthy adults and effects of age and education on naming ability. *International Journal of Science and Research* **8**, 134–139.

Tikhomirova T, Malykh A and Malykh S (2020) Predicting academic achievement with cognitive abilities: Cross-sectional study across school selection. *Behavioral Sciences* **10**, 1–13. https://doi.org/10.3390/bs10100158

Wahyuningrum SE, van Luijtelaar G and Sulastri A (2021) An online platform and a dynamic database for neuropsychological assessment in Indonesia. *Applied Neuropsychology Adult* **13**:1–10. https://doi.org/10.1080/23279095.2021.1943397

Yeung N, Nystrom LE, Aronson JAA and Cohen JD (2006) Between-Task Competition and Cognitive Control in Task Switching. *Journal of Neuroscience* **26**, 1429–1438; http://doi.org/10.1523/JNEUROSCI.3109-05.2006

Young SE (2016) The Effects of Bilingualism and Multilingualism on Lexical Retrieval. Linguistics Senior Research Projects. 6. https://digitalcommons.cedarville.edu/linguistics_senior_projects/6