

ARTICLE

Ethnicity and Performance on Language Tests in Indonesia: A Comparison between Five Ethnic Groups

Gilles van Luijtelaar^{1,2*} , Ni Made Swasti Wulanyani³ , Heni Gerda Pesau⁴ , Aria Immanuel^{3,5} , Augustina Sulastr⁶ 

¹ Donders Centre of Cognition, Radboud University, Nijmegen 6500 HD, The Netherlands

² Faculty of Psychology, Maranatha Christian University, Bandung 40164, Indonesia

³ Psychology Department, Faculty of Medicine, Udayana University, Bali 80361, Indonesia

⁴ Faculty of Psychology, Atma Jaya University, Makassar 12930, Indonesia

⁵ College of Education, University of Massachusetts Amherst, Amherst, MA 01003, USA

⁶ Faculty of Psychology, Soegijapranata Catholic University, Semarang 50234, Indonesia

ABSTRACT

Assessing cognitive performance across diverse ethnic groups in Indonesia poses a challenge due to the multitude of ethnic groups and their languages. Compounding this issue is the absence of distinct cultural norms for assessment measures. This study explored the effects of spoken language differences between five ethnic groups in Indonesia on three commonly used language tests. The ultimate objective was to determine whether new norms that incorporate Indonesian ethnic groups are warranted, akin to established practices for demographic factors such as age, and years of education. A dataset comprised of 690 individuals across five ethnic groups assessed in Bahasa Indonesia was used to examine performance differences on the Indonesian Boston Naming Test (I-BNT), a phonemic verbal fluency test, and the Token Test, and the role of spoken languages, at home and in public. Medium-sized ethnicity and small spoken language effects were observed with a Multivariate analysis of covariance, along with large education, medium age,

*CORRESPONDING AUTHOR:

Gilles van Luijtelaar, Donders Centre of Cognition, Radboud University, Nijmegen 6500 HD, The Netherlands; Faculty of Psychology, Maranatha Christian University, Bandung 40164, Indonesia; Email: Gilles.vanLuijtelaar@donders.ru.nl

ARTICLE INFO

Received: 5 July 2025 | Revised: 21 July 2025 | Accepted: 25 August 2025 | Published Online: 20 October 2025

DOI: <https://doi.org/10.30564/fls.v7i11.10906>

CITATION

van Luijtelaar, G., Wulanyani, N.M.S., Pesau, H.G., et al., 2025. Ethnicity and Performance on Language Tests in Indonesia: A Comparison between Five Ethnic Groups. *Forum for Linguistic Studies*. 7(11): 255–271. DOI: <https://doi.org/10.30564/fls.v7i11.10906>

COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

and small sex effects in performance. In particular, the I-BNT and the fluency test showed opposite ethnic differences. The variations in these differences across ethnic groups suggest that multiple factors contribute to the observed score disparities. It can be concluded that the study's findings underscore a need to develop new norms for two of the three language tests for different ethnic groups. Future research will be imperative to determine if there may be performance differences in other cognitive domains to warrant further adaptations of the Indonesian norms.

Keywords: Ethnicity; Cognitive Assessment; Language Tests; Neuropsychology

1. Introduction

A crucial component of neuropsychological assessment is the use of normative comparisons ^[1,2]. This process involves comparing a client's score with those of a reference group, typically matched for age, sex, and educational attainment. Currently, most neuropsychological measures and normative data are derived from White, educated, industrialized, rich, and democratic (WEIRD) populations ^[3]. It is widely recognized that WEIRD-derived normative data may not be appropriate for use in other cultures due to differences in sociocultural composition, health disparities, and education quality ^[4]. While some normative data attempt to account for sociocultural and linguistic effects by matching cultural and ethnic backgrounds ^[5], these data are not universally valid or appropriate across culturally diverse groups. Typically, most cognitive tests in Indonesia are only translated, and norms collected decades ago in a vastly different cultural context are used. This poses a risk of both false positives (misdiagnosing cognitively intact individuals as impaired) and false negatives (misdiagnosing cognitive deficits as normal). Such diagnostic errors can have significant repercussions, including inappropriate treatments, inequities in access to resources, reduced access to educational programs, and limited career opportunities ^[6].

Indonesia, the fourth most populous nation globally and the largest in Southeast Asia, boasts an estimated population of nearly 280 million people. The country is home to multiple ethnic groups, most of whom speak a local language ^[7], posing a challenge for accurate assessments and establishing correct norm scores for all Indonesians. It is not known whether ethnic origin and the diversity of languages spoken affect cognitive test scores in the Indonesian context. Differences in cognitive test performance among various ethnic groups have been extensively reported in the USA ^[8–11], with a consistent observation that African Americans (Blacks) exhibit lower performance compared

to European Americans (Whites) while Hispanic Americans score in between ^[11–19]. A meta-analysis confirmed lower scores on measures of quantitative and verbal ability for African Americans compared to White Americans in personnel and job selection settings ^[20]. Access to healthcare, health insurance, acculturation, literacy, food consumption, socioeconomic status, and quality of education—including the presence of sufficiently qualified teachers, their salaries, availability of teaching materials, and type of employment—among other social and environmental experiences over the life course, can potentially impact scores on cognitive tests ^[21–27]. Others have proposed that ethnic performance differences may depend on the importance of tested cognitive skills in different cultures and attitudes toward the assessment process, encompassing attitudes toward achievements in general, differences in strategies such as a preference for speed versus accuracy, familiarity with speed-based cognitive tests, and differences in socio-linguistic and ideologic characteristics of the test items and the assessment process ^[28–33].

Few studies regarding ethnic differences have been conducted in Southeast Asia. One identified study established an association between living in a low socioeconomic neighborhood and higher rates of cognitive impairment ^[34]. The ethnic differences observed have led to discussions on whether normative data of cognitive tests primarily collected for a white USA population need adjustments based on ethnic group differences to facilitate a fair interpretation of cognitive assessment results ^[12,27], similar to the common adjustments made for age and education, as was also done in the Indonesian context ^[35].

It is not known whether different normative scores should be developed for the various ethnic groups in Indonesia. Indonesia is different from Western countries in the sense that it has a huge variety of ethnic groups, and all persons have to acquire Bahasa Indonesia at elementary school. It is the lingua franca ^[36], even for the dominant

Javanese-speaking group that comprises 40% of the population. This implies that Bahasa Indonesia is the second language for all ethnic groups, which differs markedly from Western countries, where the ethnic language of the majority group is typically the lingua franca, and immigrants have to learn the language of the dominant group. Moreover, it is common in Indonesia that the local ethnic language is the main language in daily conversation at home (such as Bahasa Bali on the island of Bali, Bahasa Banjar in Southeast Kalimantan, or Bahasa Jawa in Central and East Java). The national Indonesian language is commonly used in public [7]. This implies that most Indonesians are bilingual, speaking a local language at home and Bahasa Indonesia in public venues. This distinction between the language used at home and in public is highly characteristic in the Indonesian context, and distinct roles were found for these two linguistic factors in explaining differences in language test performance among Indonesians [37–39].

A second issue is the potential variation in local language dominance among different ethnic groups. The Javanese, the largest ethnic group in Indonesia, exhibit strong dominance of the Javanese language [36,40]. In contrast, smaller ethnic groups, such as the Dayak and Banjarese, experience less dominance of their local languages and use Bahasa Indonesia more frequently in daily life because they are surrounded by many people from other ethnic groups. Unlike the Dayaks and Banjarese, who live in ethnically heterogeneous regions, the island of Bali has a rather homogenous Balinese population, 85.5% of which is Balinese, resulting in high dominance of speaking Bahasa Bali. Balinese residents predominantly speak Balinese at home and during cultural practices, but they speak Bahasa Indonesia in other settings [41]. The Chinese ethnic group presents a different case; they were immigrants throughout various periods, and due to past political restrictions, Bahasa Indonesia became the dominant language due to an enforced assimilation policy [42]. Consequently, the Chinese are less bilingual than the indigenous ethnic groups and use Bahasa Indonesia at home more often than any other Indonesian ethnic group [38,39].

Few studies have been conducted in Southeast Asia regarding factors that explain ethnic differences. A single study established an association between living in a low socioeconomic neighborhood and higher rates of cognitive

impairment in an urban Asian society [34]. The ethnic differences observed have led to discussions on whether normative data of cognitive tests primarily collected for a white USA population need adjustments based on ethnic groups for a fair interpretation of cognitive assessment results [12,27], similar to the common adjustments made for age and education, also in the Indonesian context [35].

Bilingualism is known to affect cognition. Although bilingualism is associated with superior performance on tasks requiring executive functions, opposite effects have been reported on phonemic fluency tasks [43] and on picture naming tasks [44], including the Indonesian-adapted Boston Naming Test [45]. These effects may be due to the increased cognitive load of managing two linguistic systems and the potential for cross-language interference [46].

An Indonesian consortium of psychologists from six universities collected data from four cognitive domains, including two word production tests and a language comprehension test, as well as demographic information, including languages spoken at home and in public, from healthy individuals across three regions of Java Island, Bali, East Kalimantan, and South Sulawesi. This ensured the inclusion and comparison of diverse ethnic groups. The purpose of this study is to investigate whether there are differences in performance on the language tests between five ethnic groups and how spoken languages may affect the scores. It is proposed that Bahasa-speaking groups will perform better on language tests compared to groups that speak less Bahasa, given that the language of assessment is Bahasa Indonesia. Another factor contributing to the lower performance of certain ethnic groups could be the frequently reported disadvantage of bilingualism in word production tests. Consequently, the smallest ethnic groups, Banjarese and Dayaks, who are predominantly bilingual, are expected to perform less well in these tests.

2. Methods

2.1. Participants

Healthy participants aged 16 to 80 who spoke Bahasa Indonesia were recruited to complete ten cognitive tests of the Indonesian Neuropsychological Test Battery (INTB) [47,48], including three language tests: the Indonesian Boston Naming Test (I-BNT), a phonemic verbal fluency test, and

the Token Test. The recruitment of participants followed an identical protocol across the country: trained research assistants (3rd year BA students with proven proficiency in test theory, psychometrics, data collection, and research methods) received a two-day training, followed by a supervised single assessment. These assistants enrolled Bahasa-speaking participants without knowledge of their ethnicity through word-of-mouth methods. A stratified sampling technique was employed by recruiting participants from six different locations in Indonesia to ensure the diversity of ethnic backgrounds. Individuals were mostly recruited from urbanized areas of four Indonesian islands (Java, Kalimantan, Sulawesi, and Bali). Data were collected from 2020 to 2022. The database ^[48] contained approximately 840 subjects, representing more than 40 different ethnic groups, and this number is expanding. Five ethnic groups with a sufficient number of participants were identified among the 840 subjects (ranging from close to 50 to over 200 participants). The classification regarding ethnicity, proposed by Ananta et al. ^[7], was the same as that used by the Indonesian Bureau of Statistics. A total of 690 participants representing five ethnic groups—Javanese, Balinese, Banjarese, Dayaks, and Chinese—were included. The research was conducted following the Declaration of Helsinki, and all research activities were approved by the ethics committee (Soegijapranata University, Faculty of Psychology, 001B/B.7.5/FP.KEP/IV/2018). Written informed consent was obtained from all participants.

2.2. Procedure

Bahasa Indonesia was the language of communication, instruction, and the examinees' required responses. The tests were administered in a standardized order by the trained local psychology students, familiar with assessment procedures and well-versed in the INTB. Tests took place inside participants' homes in a quiet room that minimized distractions. The total assessment time of the complete INTB was approximately less than two hours, including a break if desired. Participants received seventy-five thousand rupiahs (equivalent to five US dollars) as compensation for completing the tests. Additionally, participants completed a questionnaire regarding demographics, including the ethnicity of their parents, health, and use of languages spoken in public and at home. This report presents the results of the three language tests of the INTB.

2.3. Instruments

Indonesia-adapted Boston Naming Test (I-BNT)

We used the Indonesian-adapted version of the BNT, the I-BNT ^[45]. Participants were asked to name in Bahasa Indonesia 60 different objects within 20 seconds, presented in visual pictures, ordered from easiest to most difficult. If the participants failed at an object, a phonemic cue was given by the test assistant and if the participants failed again, a phonemic cue was given. Responses from participants were recorded using a voice recorder and written in the answer sheets. The number of spontaneously generated words, the total time to complete the 60 items, and the number of correct words were the dependent variables. For more details, see Pesau et al., 2023 and Immanuel and van Luijtelaa, 2025 ^[37,41].

Phonemic Verbal Fluency Test (pVFT)

For the pVFT, also called the Letter Fluency Test, participants were asked by the test-assistant to name as many words as possible in one minute, starting with the letter T, next with the letter K, and finally with the letter S ^[49,50]. Geographical names were not allowed. The number of correct words per letter and the sum of the correct words of the three letters were the dependent variable, as commonly done. The assistant recorded all words spoken by the participants. The number of correct and incorrect words, including repetitions per letter were determined.

Token Test (TT)

The TT is an auditory language comprehension test developed by De Renzi and Vignolo (1962) ^[51]. Participants followed oral instructions given by the test-assistants to touch, reposition, or remove various plastic tokens placed in front of them. We used the version adapted from Spreen and Benton (1969) ^[52], which was also included in the Neurosensory Center Comprehensive Examination for Aphasia (NCCEA). It contains 39 items in total, distributed across six subscales, A through F. The score for each correct response is based on the item's difficulty, with weights as follows: 1 point for simpler items (e.g., Subscale A with its seven items), 2 points for each of the four items of scale B, 3 points per correct item of scale C, 4 points for each of the four items of scale D and E, and 6 points for the sixteen items of the F scale. The cumulative scores for all six subscales add up to 163 points if every item is answered correctly. The number of correct responses for each of the six subscales and the sum of the six subscales were the de-

pendent variables.

2.4. Categories

Participants were categorized based on age and education. Groups based on decades of age were created, namely: (i) age 20–29 years, (ii) age 30–39 years, (iii) age 40–49 years, and (iv) age 50–59 years. Data for people over 60 years were pooled (v), and for the 16–19 year old participants (vi). Therefore, we distinguished six age groups for analysis. Most of our participants were in the 20 to 29 years age bracket (35.9%), while the three mid-life age brackets between 30 to 59 were rather equally represented (15.2% to 17.0%). The youngest and eldest groups had the fewest proportions of individuals (9.3% and 6.9%, respectively).

Years of education were defined according to the Indonesian education system. Category 1: only Elementary School (ES) or less, Category 2: between 7–9 years encompassing Junior High School (JHS), Category 3: between 10–12 years (Senior High School (SHS) or equivalent), Category 4: between 13–16 years (Undergraduate (UG) or equivalent), and Category 5: more than 16 years (Graduate and postgraduate). Half of the participants had an undergraduate or comparable type of education (49.3%), while 34.3% had completed Senior High School. Very few had lower levels of educational attainment (6.7% Junior High School; 4.6% Elementary School). In total, our sample represented the relatively young Indonesian middle level of education. Almost all (93.5%) of the participants spoke Bahasa Indonesia daily in public: 41.7% spoke only Bahasa Indonesia, and a larger group (51.7%) spoke Bahasa Indonesia plus one or two other languages, and 6.6% only the local language. At home, more participants (39.4%) spoke the local language.

2.5. Statistical Analysis

Differences between the ethnic groups in demographic characteristics were analyzed using Chi2 tests and ANOVAs, followed by post-hoc tests (Bonferroni). Because 15% of participants reported being of mixed ethnicity (having parents from different ethnic groups), it was first analyzed whether performance on the I-BNT, pVFT, and TT differed between participants who were of mono-ethnic and mixed-ethnic backgrounds. To that end, a MANCOVA

with mono- versus mixed-ethnicity as a between-subjects factor and the demographic factors of years of education, age, sex, and languages spoken at home and in public as cofactors was performed. MANCOVAs were also utilized to determine the effect size of the mother's and father's ethnicity (as a between-subjects factor) on performance while controlling for all other demographic factors by using them as cofactors. These analyses were conducted to determine the necessity of stratifying subsequent primary analyses for differences between ethnic groups. For this analysis of differences between the five ethnic groups and the role of spoken languages in performance on the I-BNT, pVFT, and TT, ANCOVAs were employed as well to assess their potential effects. Ethnicity and spoken language in public and at home were used as between-subjects factors, and age and education as covariates. Post-hoc tests were conducted to delineate specific group differences and interactions.

3. Results

3.1. The Demographic Characteristics of the Five Ethnic Groups

Table 1 presents the demographic characteristics across the ethnic groups in percentages. The χ^2 tests of independence revealed significant associations between ethnic group and age ($\chi^2 = 34.50$, $df = 20$, $p < 0.05$), years of education ($\chi^2 = 39.38$, $df = 16$, $p < 0.001$), daily language spoken in public ($\chi^2 = 131.5$, $df = 12$, $p < 0.001$), and daily language spoken at home ($\chi^2 = 152.84$, $df = 8$, $p < 0.001$), indicating that the ethnic groups differed significantly in their demographic characteristics. ANOVAs for age ($F = 2.66$, $df = 4$, 685 , $p < 0.05$) and education ($F = 6.27$, $df = 4$, 685 , $p < 0.001$), followed by Bonferroni post-hoc tests, showed that Balinese individuals were older than all other groups; the age difference between Balinese and Javanese as well as Chinese individuals was significant. The post-hoc tests for years of education revealed that Dayaks had significantly fewer years of education than Javanese, Balinese, and Chinese participants. These age and education differences necessitated the inclusion of these two factors as covariates in the primary analyses. There were no significant sex differences between the ethnic groups.

Table 1. Demographic characteristics in percentages by ethnicities and post hoc tests for age and (Bonferroni).

Variables and Outcomes of Post-hoc Tests for Age and Education and Chi² tests for Language Spoken	Javanese N = 260	Dayak N = 56	Banjarese N = 48	Balinese N = 95	Chinese N = 231	Total N = 690
Sex						
Male	38.08	37.50	31.25	36.84	41.56	38.55
Female	61.92	62.50	68.75	63.16	58.44	61.45
Age (in years): BAL > JAV, CH						
16–19	5.77	10.71	8.33	7.37	13.85	9.28
20–29	40.38	33.93	35.42	21.05	37.66	35.94
30–39	18.08	19.64	18.75	21.05	12.99	16.96
40–49	15.77	14.29	22.92	18.95	11.69	15.22
50–59	15.77	12.50	10.42	20.00	15.58	15.65
≥60	4.23	8.93	4.17	11.58	8.23	6.96
Education (in years): JAV, CH, BAL > DAY						
ES, 0–6	2.69	14.29	8.33	6.32	3.03	4.64
JHS, 7–9	8.08	7.14	8.33	3.16	6.06	6.67
SHS, 10–12	29.62	48.21	33.33	42.11	33.33	34.35
UG, 13–16	51.92	30.36	47.92	44.21	53.25	49.28
PG, ≥17	7.69	0.00	2.08	4.21	4.33	5.07
Language Spoken in Public						
Local language	9.23	7.14	4.17	10.53	2.16	6.52
Only Bahasa Indonesia	43.08	30.36	16.67	11.58	60.61	41.74
Bahasa Indonesia plus one other language	44.62	57.14	68.75	66.32	21.65	42.61
Bahasa Indonesia plus two other languages	3.08	5.36	10.42	11.58	15.58	9.13
Language Spoken at Home						
Local language	40.38	60.71	47.92	77.89	15.58	39.42
Only Bahasa Indonesia	39.62	37.50	47.92	21.05	71.86	48.26
Bahasa Indonesia plus one other language	20.00	1.79	4.17	1.05	12.55	12.32

Note: Values are presented as percentages. ES = Elementary school; JHS = Junior High School; SHS = Senior High School; UG = Undergraduate or equivalent; PG = Graduate/Postgraduate. The χ^2 test of independence showed no significant associations between Ethnic group and Sex, and significant associations between Ethnic group and Age ($\chi^2 = 34.50$, $df = 20$, $p < 0.05$), Ethnic group and years of Education ($\chi^2 = 39.38$, $df = 16$, $p < 0.001$), Ethnic group and Daily language spoken in public ($\chi^2 = 131.5$, $df = 12$, $p < 0.001$) and Ethnic group and Daily language at home ($\chi^2 = 152.84$, $df = 8$, $p < 0.001$). BAL was the group with the smallest percentage of individuals who only spoke Bahasa, both in public and at home. JAV = Javanese, CH = Chinese, BAL = Balinese, BAN = Banjarese, DAY = Dayak. ALL = All groups.

3.2. Languages Spoken by the Ethnic Groups

Although most Indonesians (93.4%) reported speaking Bahasa in public, significant ethnic differences were observed in the languages spoken both in public ($F = 6.99$, $df = 4$, 685 , $p < 0.001$) and at home ($F = 29.54$, $df = 4$, 685 , $p < 0.001$). In public, the Chinese were predominantly monolingual, with 60.6% speaking only Bahasa Indonesia, followed by Javanese, of whom 43.1% were monolingual. In contrast, the majority of the Banjarese (79.1%), Balinese (77.9%), and Dayaks (62.5%) used multiple languages in

public. Similarly, at home, most Chinese (71.8%) spoke only Bahasa Indonesia, while less than half of the Banjarese (47.9%), Javanese (39.6%), Dayaks (37.5%), and Balinese (21%) did so, with local languages being predominantly spoken at home in these four groups. Overall, the Chinese participants were mostly monolingual in Bahasa Indonesia. In contrast, the Balinese exhibited the greatest discrepancy between the language spoken at home (mainly Bahasa Balinese) and in public, where multiple languages were spoken with Indonesian and Western tourists.

3.3. Mono- vs. Mixed Ethnicity and Ethnicity of Mother Versus Father

MANCOVA (Pillai's trace, $F = 1.06$, $df = 36$, 1632, $p > 0.05$) indicated no significant differences in performances on the I-BNT, pVFT, and TT between mono- and mixed-ethnicity participants. Next, the effect size expressed by partial eta squared on performances was the same for the ethnicity of the father and the ethnicity of the mother, both medium-sized (see Table S1). Therefore, these results indicate no differences between monoracial and mixed-ethnicity participants on the three tests and between the influence of the ethnicities of the father and mother. This implies that these factors can be ignored in further analyses. The moth-

er's ethnic background was selected as the default value in the subsequent primary analyses. Nonetheless, age, education, and sex only to a minor extent, showed significant effects on the language tests (see Table S1) and were controlled for in our primary analyses on differences between ethnic groups.

3.4. Ethnic Differences in Performance and Effects of Spoken Languages

The primary analyses (ANCOVA) showed significant ethnic differences for two of the three language tests. The mean performances, standard errors, and statistics including post-hoc tests, are presented in Table 2.

Table 2. Each ethnicity's performances on language tests, outcomes of the ANCOVA regarding ethnicity, and outcomes of the post-hoc tests.

Variables	Javanese	Dayak	Banjarese	Balinese	Chinese	F (df = 4643)	p	Partial Eta Squared	Post-hoc Tests
	Adjusted Mean (SE)	Adjusted Mean (SE)	Adjusted Mean (SE)	Adjusted Mean (SE)	Adjusted Mean (SE)				
I-BNT number of spontaneous answers	50.5 (0.69)	51.3 (1.18)	51.0 (1.17)	44.1 (1.0)	49.8 (0.75)	6.90	<0.001	0.04	BAL < all
I-BNT total time	410 (28.6)	447 (49.0)	528 (48.8)	553 (41.6)	407 (31.2)	2.99	<0.05 n.s.	0.02 0.01	BAL > JAV
I-BNT number correct	55.2 (0.48)	55.4 (0.83)	54.7 (0.83)	53.0 (0.70)	55.1 (0.52)	2.01			
Token Test total score	146 (2.4)	151 (4.1)	151 (4.1)	150 (3.5)	152 (2.6)	1.15	n.s.	0.007	
pVFT K correct	14.3 (0.57)	11.3 (0.98)	11.7 (0.97)	14.8 (0.83)	13.6 (0.62)	2.13	n.s.	0.01	
VFT T correct	12.2 (0.57)	10.2 (0.98)	11.3 (0.97)	14.2 (0.83)	11.5 (0.62)	2.58	<0.05	0.02	BAL > DAY
pVFT S correct	13.1 (0.62)	9.6 (1.1)	11.5 (1.1)	14.8 (0.98)	13.1 (0.68)	2.90	<0.05	0.02	BAL > DAY
pVFT total correct	39.5 (1.5)	31.1 (2.6)	34.5 (2.6)	43.8 (2.2)	38.3 (1.7)	3.40	<0.01	0.02	BAL > DAY

Note: I-BNT = Indonesian version of Boston Naming Test; pVFT = Phonemic Verbal Fluency Test; BAL = Balinese; JAV = Javanese; DAY = Dayak; SE = Standard error.

Significantly lower I-BNT scores were found for Balinese individuals compared to all other ethnic groups in the number of spontaneously generated words. Additionally, Balinese participants needed more time to complete the task than Javanese participants. In contrast, the Balinese performed better than the Dayaks on two of the three scales and the total score of the pVFT.

Main effects for language spoken at home and in public were absent in the ANCOVA; however, a significant in-

teraction effect was found between the two language factors for the pVFT total score ($F = 2.49$, $df = 5$, 643, $p < 0.05$, $\eta^2 = 0.02$) and the I-BNT total score ($F = 2.76$, $df = 6,643$, $p < 0.05$, $\eta^2 = 0.02$). The post-hoc test following the interaction showed for the pVFT an disadvantage for those who speak two languages in public and only Bahasa at home. Next, there was clear advantage for those who speak only Bahasa at home and in public above other combinations as well an advantage for bilinguals at home and in public. The I-BNT

showed a disadvantage for those who use two languages in public and a local language at home as well. There were no other group differences and interactions.

The ANCOVA also revealed significant effects of age and education on the language tests, as summarized in **Table S2**. Age and years of education influenced the scores of all three language tests in the expected directions, that is that age negatively affected the test scores, and education positively. The effect of sex was confined to the TT, with males scoring higher than females.

4. Discussion

4.1. Ethnic Differences

Performance differences on two of the three language tests were observed among the ethnic groups. Their effect sizes were moderate and comparable to the effect size of age, smaller than that for educational attainment, and larger than the impact of spoken languages. There were no significant differences in performance associated with parental ethnicity (mother vs. father) or mono- vs. mixed-ethnicity. Adjusting for age and years of education differences did not eliminate the ethnic performance differences. Next, differences in language use among the ethnic groups were found: Chinese were predominantly speaking Bahasa Indonesia, both at home and in public.

The present findings reveal significant ethnic differences in performance on language-based neuropsychological tests, consistent with patterns reported in previous research^[37,53,54]. On the Indonesian version of the Boston Naming Test (I-BNT), Balinese participants produced significantly fewer spontaneous responses and required more time than all other groups, although their total number of correct responses after prompting did not differ significantly. Conversely, on the phonemic verbal fluency test (pVFT), Balinese participants outperformed Dayak participants with a significant advantage in total correct responses. The Token Test, assessing language comprehension, did not reveal significant group differences.

What could be the reasons for the ethnic differences? We do not think at beforehand that ethnic differences in cognitive abilities exist, on the other hand, we do see clear differences among ethnic groups in their performance on language tests. Blommaert^[31], Pennycook^[33], and García and

Wei^[32] use a critical sociolinguistic perspective, and collectively argue that language assessments cannot be treated as neutral cognitive measures, as they are embedded in social and ideological structures. Blommaert advocates that test items and norms represent specific linguistic and cultural capitals and that these resources are unequally distributed among Indonesia's ethnic groups. For instance, the I-BNT relies on standardized Bahasa Indonesia and assumes familiarity with object names common across all ethnic groups and equally in urbanized and rural areas. Balinese participants' slower and less spontaneous naming may reflect a linguistic repertoire in which equivalent lexical items are more accessible in Balinese than in Bahasa Indonesia, requiring additional retrieval effort. From a critical applied linguistics perspective, Pennycook^[33] argues that language tests legitimize certain language practices while marginalizing others. The current tests privilege monoglossic norms that align with national language policy but not with Indonesia's multilingual reality, thereby positioning certain groups—such as Dayak or Banjarese speakers—at a systematic disadvantage. This ideological dimension reframes lower scores not as cognitive deficits but as outcomes of language policy and historical inequality in access to the dominant language. García and Wei's^[32] notion of translanguaging challenges the compartmentalized view of bilingualism underpinning these tests. Indonesian multilingual speakers typically draw from an integrated repertoire rather than maintaining strict separation between languages. Tasks like pVFT, which constrain responses to Bahasa Indonesian words beginning with a specific letter, impose artificial boundaries that disrupt natural lexical access. Dayak participants' weaker performance on fluency tasks may therefore reflect testing conditions that penalize their translanguaging practices rather than limitations in verbal ability. Taken together, these perspectives underscore that ethnic differences in language test performance are sociolinguistically mediated. Future research should explore translanguaging-based adaptations or alternative scoring practices to capture multilingual speakers' full linguistic resources.

Limited studies have compared ethnic groups on neuropsychological tests outside the USA. A single study reported differences in age-standardized cognitive impairment prevalence among Malays, Indians, and Chinese in Singapore^[55]. No differences were found in a comparison

between Javanese and Sundanese living close together on Java island on the same three language tests as currently used ^[56]. Notably, differences in test performance among ethnic groups should not imply superiority or inferiority but rather highlight the need for specific normative data tailored to some ethnic groups ^[5]. Establishing such norms could account for sociocultural and linguistic factors underlying these differences ^[31–33], facilitating fairer test interpretation across the large diversity of Indonesian ethnic groups.

Disparities in education and healthcare are significant contributors to cognitive performance differences across ethnic groups in the USA ^[24,57,58]. Comparable quantitative cross-cultural research on cognition is completely lacking in Indonesia, where neuropsychology is virtually nonexistent. Additionally, there is no research regarding the possible causes of performance differences among different ethnic groups. Since Indonesia declared its independence, the official policy has been expressed in the national motto “*Bhinneka Tunggal Ika*” (“Unity in diversity”), which is referenced in Indonesia’s constitution, promoting harmony among its ethnic groups and does not emphasize differences. Therefore, regional differences in economic situation and health facilities are extensively reported by the Indonesian Bureau of Statistics (IBS), but not ethnic differences in relevant factors such as education, socioeconomic status, and access to and quality of healthcare. There is no direct relation between ethnicity and borders of regions and provinces. Moreover, the transmigration policy has contributed to a mixed ethnic population in different parts of the country. Additionally, Indonesia’s large cities, where a substantial part of the data was collected, are a melting pot of different ethnicities. Only indirect data from the IBS are available per province, such as maternal and infant mortality as an indirect quality of healthcare index, access to healthcare facilities, life expectancy at birth, literacy, access to secondary and tertiary education, employment, and income. The Human Development Index (HDI), a mix of economic, health, and education parameters also used by WHO, shows large differences between Indonesian provinces, with lower scores in East Indonesia than in West Indonesia and Bali. However, in the urbanized districts of Java (Jakarta, Semarang, Surabaya) and Kalimantan (Samarinda), and Bali (Denpasar), where the data were collected, the HDI indexes were all high ($>0.80 < 0.83$, IBS)

and not different from each other (BPS-Statistics Indonesia. Welfare Indicators 2022) ^[59]. Despite this, it is expected that the quality of education and healthcare will have a major effect on cognition. More and different data are necessary to explain the ethnic differences reported here. Identifying the reasons for these differences will require systematic research that considers a multitude of environmental as well as socio-linguistic factors specific to the local Indonesian context.

4.2. Languages Spoken and Other Effects

Language use at home and in public has been identified as factors influencing performance in the INTB’s language tests ^[37–39]. Interestingly, the Chinese are the most monolingual and the most likely to speak Bahasa Indonesia both at home and in public; in contrast, individuals from the smaller ethnic groups (Dayak, Banjarese, and Balinese) were more often bilingual in public settings and monolingual at home. The Balinese were the least likely to speak Bahasa Indonesia, both in public and at home, which may have negatively impacted their low I-BNT scores. Another contributing linguistic factor to this low performance could be that speaking Balinese at home and two or three languages in public creates conflict between lexical systems, reducing both the number of spontaneously generated words and the speed of completing the I-BNT.

The results of the pVFT showed the opposite pattern: the Balinese had the highest scores and differed significantly from the Dayak. Thus, no single ethnic group consistently scored the highest or lowest across all tests. It can therefore be stated that the results of these tests argue against a one-factor explanation for the observed ethnic differences. The pVFT is sensitive to the lexicon of the assessment language ^[60]. Testing all participants in Bahasa Indonesia ensured lexical uniformity; however, speaking more than one language introduces lexical competition. This competition may vary among ethnic groups due to large differences in the lexicons of the four Austronesian languages (with Chinese not being an Austronesian language), in particular in the number of words beginning with the three letters S, K, and T. This linguistic factor warrants further investigation. The sensitivity of the pVFT to ethnicity was also reported in a U.S. sample ^[61].

The interactions between language-use factors on the

word production tests revealed an expected advantage on the pVFT for participants who used the same language(s) at home and in public. This advantage was observed both for monolingual Bahasa speakers and for bilinguals. The first finding suggests that being assessed in one's mother tongue without interference from another language facilitates performance. The second indicates that consistent bilingual use across settings may mitigate the assumed conflict between two competing lexical systems.

In contrast, lower I-BNT scores were found among participants who spoke a local language at home but were bilingual in public, suggesting that switching languages in public while maintaining a different home language is disadvantageous. Interestingly, the reverse situation—switching at home but not in public—did not produce this effect. Overall, the findings suggest that language use in the home environment exerts a stronger influence on performance than language use in public.

Other explanations for the observed ethnic differences in language test performance remain elusive as well. Ethnic differences in attitudes toward time and familiarity with timed tasks may also contribute to these performance differences. Comparative studies have suggested that Western cultures prioritize speed, whereas Russian and Asian cultures emphasize accuracy^[28,29]. Other studies reported differences among Western samples, hypothesizing that samples may differ in their performance on cognitive tests requiring a balance between speed and accuracy^[62]. Time pressure was emphasized only in the pVFT, not in the I-BNT or Token Test. Performance differences in the pVFT can indeed be attributed to differences in attitude and experience among Indonesian ethnic groups. However, whether this holds for other Indonesian ethnic groups remains to be determined.

Another source of ethnic differences, commonly cited in cross-cultural studies, is racial or administration bias^[63]. These are unlikely in our study, as participants were tested by trained individuals from their local communities. It is acknowledged that the BNT was adapted for Indonesia in a Javanese sample^[49], and this may not fully accommodate Balinese cultural and linguistic contexts. Whether differential item functions play a role in Indonesia in cross-cultural picture-naming tests needs to be evaluated. In sum, it is quite possible that ethnic disparities on language tests

in Indonesia likely reflect structural linguistic inequalities rather than genuine language impairments. These findings lend empirical support to Canagarajah's (2005)^[64] critique of monolingual testing regimes and underscore the importance of localizing assessment tools—not only linguistically, but also ideologically—in order to foster greater equity and validity in cross-ethnic language assessment, although the feasibility in Indonesia in multi-cultural and multi language society can be questioned.

Age and education effects on I-BNT, TT, and pVFT scores were consistent with international findings^[47,65,66]. Ethnic group differences in performance highlight the need for normative adjustments for specific groups, in addition to adjusted norms for age and education, as commonly practiced in neuropsychological assessment^[67,68], to ensure fairness in testing. In other countries, different normative scores have been developed for diverse ethnic groups^[5,24,67–76]. Developing separate normative data for the heterogeneous Indonesian population is imperative to improve the validity and fairness of the assessments.

Efforts to create culture-fair tests, such as those by the European Consortium on Cross-Cultural Neuropsychology (ECCrON), aim to minimize reliance on culturally specific, often school-based skills^[77]. In Indonesia, tests like the Culture Fair Intelligence Test (CFIT) for children and the AJT Cog Test^[78,79] provide normative data for urban and rural populations but remain limited in scope^[80]. Further work is required to ensure fair and valid interpretations across Indonesia's diverse population. In South-East Asia, however, initiatives to develop culture-fair measures are on hold because of the high heterogeneity of ethnic groups and the multitude of languages spoken. Therefore, our current strategy is to introduce clinically relevant neuropsychological tests covering various cognitive domains and constructs^[22], and to choose the national language, Bahasa Indonesia, for the assessment. Bahasa Indonesia is becoming the dominant language in the country^[42], especially for the younger generations. The 2020 Sensus reported that 97% of the Indonesian population speaks and understands Bahasa Indonesia. We are currently exploring whether adaptations of the normative data in our dynamic database are necessary. On the other hand, we acknowledge the limitations of our strategy, seek solutions for obtaining representative normative test scores, and strive for fair interpretation of the test scores.

4.3. Limitations of the Study

Data for several ethnic groups were aggregated from various regions, including individuals of Chinese and Hokkien descent from multiple islands having different mother-tongues (Mandarin, Cantonese, Teochow, Hokkien), Javanese from West and Central Java, and Yogyakarta, and members of different Dayak tribes. Consequently, intra-group differences may exist within some ethnic categories, such as between Cantonese and Mandarin speaking Chinese Indonesians, and between Dayak tribes. Additionally, group sizes vary, and statistical outcomes are inherently influenced by sample size. While detecting differences between smaller groups is more challenging, significant differences were still observed in groups with fewer than 100 participants, suggesting that the effects were substantial and warrant consideration. As additional data become available in the future, we aim to conduct more nuanced analyses of within-group differences (Chinese from different parts of the archipelago and Dayak's different ethnic tribes) and ethnic differences among Indonesia's diverse ethnic groups with more balanced sample sizes.

Ethnicity was determined through self-report, allowing individuals to identify the ethnicities of both parents rather than restricting them to predefined categories. This classification approach aligns with the methodology used in national Indonesian surveys [7]. Only participants who conformed to these established ethnic categories were included in the study.

Our sample used in this study exhibited some imbalances in the distribution of age groups and educational attainment levels across ethnic groups, although we controlled for this in our statistical analyses. It was predominantly drawn from urbanized areas, which may have influenced the findings. Well known is that access to healthcare, quality of education, literacy and social economic situation in Indonesia's extensive rural regions is less than in urban areas, and this undoubtedly has influenced the scores on the tests negatively. Nevertheless, age and education were statistically controlled in the analyses, and medium effect sizes among ethnic groups were still observed.

A final limitation is that many ethnic groups were not assessed and included in the INTB database, mainly residing in Sumatera and East Indonesia. Therefore, the ongoing collection of normative data that better reflects the entire

Indonesian population is vital. This is particularly important as cultural and other environmental factors shaping cognition, such as transmigration, urban migration, healthcare quality, and acculturation, may evolve. Next, socio-linguistic factors are likely to influence the performance on the language tests as well.

4.4. Practical Consequences

A first practical takeaway of this study is that the effect size of ethnicity on the language comprehension task was negligible, indicating no structural comprehension gap. A second one is that the two expressive tasks requiring rapid lexical retrieval under standardized constraints show ethnic differences, and may be caused by e.g. sociolinguistic disparities, besides environmental factors. The significant and clinically relevant performance differences among some ethnic groups, a difference in seven spontaneously generated items on the I-BNT, 146 seconds in completion time for the I-BNT, and 12.7 words generated on the pVFT (Table 2) warrant the adaptation of norms for various Indonesian ethnic groups. This leads to more accurate and fair interpretations of test scores. Ideally, each cognitive test should have a comprehensive formula that accounts for all relevant demographic adjustments, based on regression or other norming methods. It also needs to be determined for which ethnic groups these adaptations are necessary. Ethnicity-specific norms and adjustments require more data and stratified sampling by ethnicity and region. Our ultimate goal is to propose normative scores for Indonesians across a range of cognitive tests, incorporating potential demographic adjustments as is standard internationally. This approach aims to ensure a more equitable assessment and interpretation of test scores for a broader population.

5. Conclusions

The performance differences on two language expression tasks (I-BNT and pVFT) among the five ethnic groups highlight the need for adapted normative scores for ethnicity in addition to age and education. This will minimize potential bias in the interpretation of the scores of these tests. While culturally fair tests may become more available in the future, current best practices should emphasize the cultural competence of the practitioner, an understanding that

culture shapes cognition, and that socio-linguistic and environmental factors are major factors shaping the performance on cognitive tests.

Supplementary Materials

Interpretation of effect sizes as expressed by partial eta squared was according to Richardson (2011).

Table S1. Effect size of the factors affecting the scores of the three language tests.

Factors	Partial eta Squared	Factors	Partial eta Squared	Classifications
Ethnicity of the mother	0.10	Ethnicity of the father	0.10	Both Medium
Effect of Age	0.08	Effect of Age	0.07	Both Medium
Effect of Education	0.31	Effect of Education	0.32	Both Large
Effect of Sex	0.03	Effect of Sex	0.04	Both Small
Effect of language spoken at home	0.03	Effect of language spoken at home	0.03	Both Small
Effect of language spoken in public	0.05	Effect of language spoken in public	0.05	Both Small

Note: The effect sizes were established with MANCOVA's.

Table S2. Demographic factors significantly affecting the performance on the three language tests.

Factors	Language Tests	Variables Affected
Age	I-BNT	Spontaneous, Total correct, Total time
	TT	B-, C-, F-scales, Total score
	pVFT	S-, K-, T-scales, Total score
Education	I-BNT	Spontaneous, Total correct, Total time
	TT	B-, C-, D-, E-, F-scales, Total score
	pVFT	S-, K-, T-scales, Total score
Sex	I-BNT	n.s.
	TT	F-scale, Total score
	pVFT	n.s.
Language Spoken in Public	I-BNT	Total time
	TT	n.s.
	pVFT	n.s.
Language Spoken at Home	I-BNT	Spontaneous
	TT	n.s.
	pVFT	n.s.

Note: I-BNT = Indonesian version of Boston Naming Test, Spontaneous is the number of spontaneously generated words, Total correct is the total number of correct items, Total time is the time to complete the I-BNT. TT = Token Test, given are the subscales showing a significant effect, Total is the result of all six subscales. pVFT = phonemic Verbal Fluency Test, S, K, and T refer to its subscales, Total refers to the sum of the three subscales. n.s. = no significant outcomes for this test on any of the variables.

Author Contributions

Conceptualization, G.v.L.; methodology, G.v.L., N.M.S.W., H.G.P., A.I., A.S.; validation, G.v.L., N.M.S.W., H.G.P., A.I., A.S.; formal analysis, G.v.L.; investigation, H.G.P., N.M.S.W., A.S.; resources, A.S.; data curation, G.v.L., A.I.; writing—original draft preparation, G.v.L.; writing—review and editing, G.v.L., N.M.S.W., H.G.P.,

A.I., A.S.; supervision, G.v.L.; project administration A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by DIKTI grant named Development of I-ANDI.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of SOEGIJAPRANATA University (protocol code 001B/B.7.5/FP.KEP/IV/2018) for studies involving humans.

Informed Consent Statement

Informed consent was obtained before the start of the study from all subjects involved in the study.

Data Availability Statement

The data used in this study are available on any reasonable request and should be addressed to Dr. Augustina Sulastri, Faculty Psychology, Soegijapranata University, Semarang, Indonesia.

Acknowledgments

The authors want to thank the members of the Indonesian Neuropsychology Consortium, the test assistants for collecting the data, Mr. Muhammad Anjar Gagahriyanto, Udayana University, Denpasar, Bali, for editorial assistance and Pierre Soeren, Radboud University, for statistical advice.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Harvey, P.D., 2012. Clinical applications of neuropsychological assessment. *Dialogues in clinical neuroscience*. 14(1), 91–99. DOI: <https://doi.org/10.31887/DCNS.2012.14.1/pharvey>
- [2] Huizenga, H.M., Agelink van Rentergem, J.A., Grasman, R.P., et al., 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(6), 611–629. DOI: <https://doi.org/10.1080/13803395.2015.1132299>
- [3] Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? *Behavioral and brain sciences*. 33(2–3), 61–83. DOI: <https://doi.org/10.1017/S0140525X0999152X>
- [4] Siciliano, M., Chiorri, C., Battini, V., et al., 2019. Regression-based normative data and equivalent scores for Trail Making Test (TMT): an updated Italian normative study. *Neurological Sciences*. 40(3), 469–477. DOI: <https://doi.org/10.1007/s10072-018-3673-y>
- [5] Casaletto, K.B., Umlauf, A., Beaumont, J., et al., 2015. Demographically corrected normative standards for the English version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*. 21(5), 378–391. DOI: <https://doi.org/10.1017/S1355617715000351>
- [6] Smith, H.J., Chen, J., Liu, X., 2008. Language and rigour in qualitative research: problems and principles in analyzing data collected in Mandarin. *BMC medical research methodology*. 8(1), 1–8. DOI: <https://doi.org/10.1186/1471-2288-8-44>
- [7] Ananta, A., Arifin, E.N., Hasbullah, M.S., et al., 2015. Demography of Indonesian's Ethnicity. *Institute of Southeast Asian studies*: Singapore.
- [8] Mehta, K.M., Simonsick, E.M., Rooks, R., et al., 2004. Black and white differences in cognitive function test scores: what explains the difference? *Journal of the American Geriatrics Society*. 52(12), 2120–2127. DOI: <https://doi.org/10.1111/j.1532-5415.2004.52575.x>
- [9] Sheffield, K.M., Peek, M.K., 2011. Changes in the prevalence of cognitive impairment among older Americans, 1993–2004: overall trends and differences by race/ethnicity. *American Journal of Epidemiology*. 174(3), 274–283. DOI: <https://doi.org/10.1093/aje/kwr074>
- [10] Sharma, S., Hale, J.M., Myrskylä, M., et al., 2023. Racial, ethnic, nativity, and educational disparities in cognitive impairment and activity limitations in the United States, 1998–2016. *Demography*. 60(5), 1441–1468. DOI: <https://doi.org/10.1215/00703370-10941414>
- [11] Diehr, M.C., Heaton, R.K., Miller, W., et al., 1998. The paced auditory serial addition task (PASAT): Norms for age, education, and ethnicity. *Assessment*. 5, 375–387. DOI: <https://doi.org/10.1177/107319119800500407>
- [12] Gasquoine, P.G., 2009. Race-norming of neuropsychological tests. *Neuropsychology Review*. 19, 250–262. DOI: <https://doi.org/10.1007/s11065-009-9090-5>
- [13] Heaton, R.K., Miller, S.W., Taylor, M.J., et al., 2004. Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted

- neuropsychological norms for African American and Caucasian adults. Psychological Assessment Resources: Lutz, FL, USA.
- [14] Norman, M.A., Evans, J.D., Miller, W.S., et al., 2000. Demographically corrected norms for the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*. 22(1), 80–94. DOI: [https://doi.org/10.1076/1380-3395\(200002\)22:1;1-8;FT080](https://doi.org/10.1076/1380-3395(200002)22:1;1-8;FT080)
- [15] Norman, M.A., Moore, D.J., Taylor, M., et al., 2011. Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. *Journal of Clinical and Experimental Neuropsychology*. 33(7), 793–804. DOI: <https://doi.org/10.1080/13803395.2011.559157>
- [16] Díaz-Venegas, C., Downer, B., Langa, K.M., et al., 2016. Racial and ethnic differences in cognitive function among older adults in the USA. *International Journal of Geriatric Psychiatry*. 31(9), 1004–1012. DOI: <https://doi.org/10.1002/gps.4410>
- [17] Gershon, R.C., Fox, R.S., Manly, J.J., et al., 2020. The NIH Toolbox: Overview of development for use with Hispanic populations. *Journal of the International Neuropsychological Society*. 26(6), 567–575. DOI: <https://doi.org/10.1017/S1355617720000028>
- [18] Morlett Paredes, A., Gooding, A., Artiola i Fortuny, L., et al., 2021a. The state of neuropsychological test norms for Spanish-speaking adults in the United States. *The Clinical Neuropsychologist*. 35(2), 236–252. DOI: <https://doi.org/10.1080/13854046.2020.1729866>
- [19] Morlett Paredes, A., Carrasco, J., Kamalyan, L., et al., 2021b. Demographically adjusted normative data for the Halstead category test in a Spanish-speaking adult population: Results from the Neuropsychological Norms for the U.S.-Mexico Border Region in Spanish (NP-NUMBRS). *The Clinical Neuropsychologist*. 35(2), 356–373. DOI: <https://doi.org/10.1080/13854046.2019.1709660>
- [20] Roth, P.L., Bevier, C.A., Bobko, P., et al., 2001. Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*. 54(2), 297–330. DOI: <https://doi.org/10.1111/j.1744-6570.2001.tb00094.x>
- [21] Mishra, R.C., 2001. Cognition across cultures. In: Matsumoto, D. (eds.). *The Handbook of Culture and Psychology*. Oxford University Press: Oxford, UK. pp. 119–135.
- [22] Ardila, A., 2018. Historical development of human cognition: A cultural-historical neuropsychological perspective. Springer: Singapore. pp. 135–159.
- [23] Ostrosky-Solís, F., Ramirez, M., Ardila, A., 2004. Effects of culture and education on neuropsychological testing: A preliminary study with indigenous and nonindigenous population. *Applied neuropsychology*. 11(4), 186–193. DOI: https://doi.org/10.1207/s15324826an1104_3
- [24] Lucas, J.A., Ivnik, R.J., Smith, G.E., et al., 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(2), 243–269. DOI: <https://doi.org/10.1080/13854040590945337>
- [25] Glymour, M.M., Weuve, J., Chen, J.T., 2008. Methodological challenges in causal research on racial and ethnic patterns of cognitive trajectories: measurement, selection, and bias. *Neuropsychology Review*. 18(3), 194–213. DOI: <https://doi.org/10.1007/s11065-008-9066-x>
- [26] Park, D.C., Huang, C.M., 2010. Culture wires the brain: A cognitive neuroscience perspective. *Perspectives on Psychological Science*. 5(4), 391–400. DOI: <https://doi.org/10.1177/1745691610374591>
- [27] Medina, L.D., Torres, S., Gioia, A., et al., 2021. Reporting of demographic variables in neuropsychological research: An update of O'Bryant et al.'s trends in the current literature. *Journal of the International Neuropsychological Society*. 27(5), 497–507. DOI: <https://doi.org/10.1017/S1355617720001083>
- [28] Agranovich, A.V., Puente, A.E., 2007. Do Russian and American normal adults perform similarly on neuropsychological tests? Preliminary findings on the relationship between culture and test performance. *Archives of Clinical Neuropsychology*. 22(3), 273–282. DOI: <https://doi.org/10.1016/j.acn.2007.01.003>
- [29] Agranovich, A.V., Panter, A.T., Puente, A.E., et al., 2011. The culture of time in neuropsychological assessment: exploring the effects of culture-specific time attitudes on timed test performance in Russian and American samples. *Journal of the International Neuropsychological Society*. 17(4), 692–701. DOI: <https://doi.org/10.1017/S1355617711000592>
- [30] Melikyan, Z.A., Agranovich, A.V., Puente, A.E., 2019. Fairness in psychological testing. In: Goldstein, G., Allen, D.N., DeLuca, J. (eds.). *Handbook of Psychological Assessment – 4th Edition*. Academic Press: Cambridge, MA, USA. pp. 551–572.
- [31] Blommaert, J., 2010. *The sociolinguistics of globalization*. Cambridge University Press: Cambridge, UK.

- [32] García, O., Wei, L., 2014. Translanguaging: Language, bilingualism and education. Palgrave Macmillan: London, UK.
- [33] Pennycook, A., 2001. Critical applied linguistics: A critical introduction. Routledge: New York, NY, USA.
- [34] Wee, L.E., Yeo, W.X., Yang, G.R., et al., 2012. Individual and area level socioeconomic status and its association with cognitive function and cognitive impairment (low MMSE) among community-dwelling elderly in Singapore. *Dementia and Geriatric Cognitive Disorders Extra*. 2(1), 529–542. DOI: <https://doi.org/10.1159/000345036>
- [35] Widhianingtanti, L.C., van Luijelaar, G., Suryani, A.O., et al., 2022. Indonesian Trail Making Test: Analysis of psychometric properties, effects of demographic variables, and norms for Javanese adults. *Jurnal Psikologi*. 49(2). DOI: <https://doi.org/10.22146/jpsi.68953>
- [36] Suryadinata, L., Arifin, E.N., Ananta, A., 2003. Indonesia's Population: Ethnicity and Religion in a Changing Political Landscape. Institute of Southeast Asian Studies: Singapore.
- [37] Pesau, H.G., Immanuel, A.S., Sulastri, A., et al., 2023. The role of daily spoken language on the performance of language tests: The Indonesian experience. *Bilingualism: Language and Cognition*. 26(3), 538–549. DOI: <https://doi.org/10.1017/S136672892200075X>
- [38] Sari, B.T., van de Vijver, F., Chasiotis, A., et al., 2019a. Contextualized bilingualism among adolescents from four different ethnic groups in Indonesia. *International Journal of Bilingualism*. 23(6), 1469–1482. DOI: <https://doi.org/10.1177/1367006918803678>
- [39] Sari, B.T., Chasiotis, A., van de Vijver, F.J., et al., 2020. The importance of language vocabulary and language usage for sociocultural adjustment among Indonesian adolescents from three bilingual ethnic groups. *Journal of Multilingual and Multicultural Development*. 41(6), 531–546. DOI: <https://doi.org/10.1080/01434632.2019.1630417>
- [40] Ananta, A., Utami, D.R.W.W., Purbowati, A., 2016. Declining dominance of an ethnic group in a large multi-ethnic developing country: The case of the Javanese in Indonesia. *Population Review*. 55(1), 1–26. DOI: <https://doi.org/10.1353/prv.2016.0000>
- [41] Immanuel, A., van Luijelaar, G., 2025. Proficiency of language and item familiarity: Their effects on the I-BNT in Balinese. *Jurnal Psikologi Ulayat*. 12(1), 5–29. DOI: <https://doi.org/10.24854/jpu951>
- [42] Wright, S., 2016. Nation Building in the Wake of Colonialism: Old Concepts in New Settings. In: Wright, S. (eds.). *Language Policy and Language Planning*. Palgrave Macmillan: London, UK. pp. 78–111.
- [43] Gollan, T.H., Montoya, R.I., Fennema-Notestine, C., et al., 2005. Bilingualism affects picture naming but not picture classification. *Memory & cognition*. 33, 1220–1234. DOI: <https://doi.org/10.3758/BF03193224>
- [44] Sheppard, C., Kousaie, S., Monetta, L., et al., 2016. Performance on the Boston Naming Test in Bilinguals. *Journal of the International Neuropsychological Society*. 22(3), 350–363. DOI: <https://doi.org/10.1017/S135561771500123X>
- [45] Sulastri, A., Utami, M.S.S., Hendriks, M., et al., 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 (IJSR)*. 8(11), 134–139. DOI: <https://doi.org/10.21275/ART20202307>
- [46] Bialystok, E., Craik, F.I., Luk, G., 2012. Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*. 16(4), 240–250. DOI: <https://doi.org/10.1016/j.tics.2012.03.001>
- [47] Wahyuningrum, S.E., Sulastri, A., Hendriks, M.P.H., et al., 2022. The Indonesian Neuropsychological Test Battery (INTB): Psychometric properties, preliminary normative scores, the underlying cognitive constructs, and the effects of age and education. *Acta Neuropsychologica*. 20(4), 445–470. DOI: <https://doi.org/10.5604/01.3001.0016.1339>
- [48] Wahyuningrum, S.E., van Luijelaar, G., Sulastri, A., 2023. An online platform and a dynamic database for neuropsychological assessment in Indonesia. *Applied Neuropsychology: Adult*. 30(3), 330–339. DOI: <https://doi.org/10.1080/23279095.2021.1943397>
- [49] Hendrawan, D., Hatta, T., 2010. Evaluation of stimuli for development of the Indonesian version of verbal fluency task using ranking method. *Psychologia*. 53(1), 14–26. DOI: <https://doi.org/10.2117/psysoc.2010.14>
- [50] Pesau, H.G., van Luijelaar, G., 2021. Equivalence of Traditional and Internet-Delivered Testing of Word Fluency Tasks. *Jurnal Psikologi*. 20(1), 35–49. DOI: <https://doi.org/10.14710/jp.20.1.35-49>
- [51] De Renzi, A., Vignolo, L.A., 1962. The token test: A sensitive test to detect receptive disturbances in aphasia. *Brain: a journal of neurology*. 85(4), 665–678. DOI: <https://doi.org/10.1093/brain/85.4.665>
- [52] Spellacy, F.J., Spreen, O., 1969. A short form of the Token Test. *Cortex*. 5(4), 390–397.
- [53] Immanuel, A.S., Pesau, H.G., Wulanyani, N.M.S., et al., 2024. The Role of Spoken Language on Performance of Cognitive Tests: the Indonesian Experience. *Journal of Cognition and Culture*. 24(3–4), 207–240. DOI: <https://doi.org/10.1163/15685373-12340187>

- [54] Pesau, H.G., Immanuel, A.S., Sulastri, A., et al., 2022. The Influence of Ethnicity on Language Tests: A comparison between Balinese and Banjarese. Available from: <https://conference.unika.ac.id/index.php/ssic/iconbi/paper/viewFile/524/138> (cited 23–24 June 2022).
- [55] Wong, M.Y.Z., Tan, C.S., Venketasubramanian, N., et al., 2019. Prevalence and risk factors for cognitive impairment and dementia in Indians: A multiethnic perspective from a Singaporean study. *Journal of Alzheimer's Disease*. 71(1), 341–351. DOI: <https://doi.org/10.3233/JAD-190610>
- [56] Pali, V.H., van Luijtelaa, G., Wardani, R., 2024. Comparison of INTB Language Test Performance Between Javanese and Sundanese. In *Proceedings of The 3rd International Conference on Biopsychosocial Issues*, Semarang, Indonesia, 26–27 September 2024; pp. 185–198.
- [57] Snitz, B.E., Unverzagt, F.W., Chang, C.C., et al., 2009. Effects of age, gender, education and race on two tests of language ability in community-based older adults. *International Psychogeriatrics*. 21(6), 1051–1062. DOI: <https://doi.org/10.1017/S1041610209990214>
- [58] Fillenbaum, G.G., Heyman, A., Huber, M.S., et al., 2001. Performance of elderly African American and White community residents on the CERAD Neuropsychological Battery. *Journal of the International Neuropsychological Society*. 7(4), 502–509. DOI: <https://doi.org/10.1017/s1355617701744062>
- [59] BPS-Statistics Indonesia, 2022. Welfare Indicators 2022. Available from <https://www.bps.go.id/en/publication/2022/11/30/71ae912cc39088ead37c4b67/welfare-indicators-2022.html> (cited 5 December 2024).
- [60] Kempler, D., Teng, E.L., Dick, M., et al., 1998. The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society*. 4(6), 531–538. DOI: <https://doi.org/10.1017/s1355617798466013>
- [61] Gladsjo, J.A., Schuman, C.C., Evans, J.D., et al., 1999. Norms for letter and category fluency: demographic corrections for age, education, and ethnicity. *Assessment*. 6(2), 147–178. DOI: <https://doi.org/10.1177/107319119900600204>
- [62] Roivainen, E., 2019. European and American WAIS IV norms: Cross-national differences in perceptual reasoning, processing speed and working memory subtest scores. *Scandinavian Journal of Psychology*. 60, 513–519. DOI: <https://doi.org/10.1111/sjop.12581>
- [63] van de Vijver, F.J., Poortinga, Y.H., 1997. Towards an integrated analysis of bias in cross-cultural assessment. *European journal of psychological assessment*. 13(1), 29–37. DOI: <https://doi.org/10.1027/1015-5759.13.1.29>
- [64] Canagarajah, S., 2005. *Reclaiming the local in language policy and practice*. Routledge: New York, NY, USA.
- [65] Fernandez, A.L., Marcopulos, B.A., 2008. A comparison of normative data for the Trail Making Test from several countries: Equivalence of norms and considerations for interpretation. *Scandinavian Journal of Psychology*. 49(3), 239–246. DOI: <https://doi.org/10.1111/j.1467-9450.2008.00637.x>
- [66] Mitrushina, M., Boone, K.B., Razani, J., et al., 2005. *Handbook of normative data for neuropsychological assessment*. Oxford University Press: Oxford, UK.
- [67] Casaletto, K.B., Heaton, R.K., 2017. Neuropsychological assessment: Past and future. *Journal of the International Neuropsychological Society*. 23(9–10), 778–790. DOI: <https://doi.org/10.1017/S1355617717001060>
- [68] Zahodne, L.B., Watson, C.W., Seehra, S., et al., 2018. Positive psychosocial factors and cognition in ethnically diverse older adults. *Journal of the International Neuropsychological Society*. 24(3), 294–304. DOI: <https://doi.org/10.1017/S1355617717000935>
- [69] Zahodne, L.B., Sharifian, N., Kraal, A.Z., et al., 2021. Positive psychosocial factors and cognitive decline in ethnically diverse older adults. *Journal of the International Neuropsychological Society*. 27(1), 69–78. DOI: <https://doi.org/10.1017/S1355617720000648>
- [70] Gasquoin, P.G., Weimer, A.A., Estevis, E., et al., 2021. Survey of Spanish Language Neuropsychological Test Use in the Assessment of Hispanic Americans/Latino/as/x. *Archives of Clinical Neuropsychology*. 36(7), 1350–1360. DOI: <https://doi.org/10.1093/arpclin/aacl131>
- [71] Carone, D.A., 2024. *A Compendium of Neuropsychological Tests: Fundamentals of neuropsychological assessment and test reviews for clinical practice* (4th ed.): E. M. S. Sherman, J.E. Tan & M. Hrabok, Oxford University Press, New York, NY, 2022. *Applied Neuropsychology: Adult*. 31(3), 306–309. DOI: <https://doi.org/10.1080/23279095.2023.2260143>
- [72] Milani, S.A., Marsiske, M., Cottler, L.B., et al., 2018. Optimal cutoffs for the Montreal Cognitive Assessment vary by race and ethnicity. *Alzheimer's & dementia*. 10, 773–781. DOI: <https://doi.org/10.1016/j.dadm.2018.09.003>
- [73] Flores, I., Casaletto, K.B., Marquine, M.J., et al., 2017. Performance of Hispanics and Non-Hispanic Whites on the NIH Toolbox Cognition Battery: The Roles of Ethnicity and Language Backgrounds. *The Clinical Neuropsychologist*. 31(4), 783–797. DOI: <https://doi.org/10.1017/S1355617717000935>

- org/10.1080/13854046.2016.1276216
- [74] Rovner, B.W., Casten, R.J., Arenson, C., et al., 2012. Racial differences in the recognition of cognitive dysfunction in older persons. *Alzheimer disease and associated disorders*. 26(1), 44–49. DOI: <https://doi.org/10.1097/WAD.0b013e3182135f09>
- [75] Unverzagt, F.W., Hall, K.S., Torke, A.M., et al., 1996. Effects of age, education, and gender on cerad neuropsychological test performance in an african American sample. *Clinical Neuropsychologist*. 10(2), 180–190. DOI: <https://doi.org/10.1080/13854049608406679>
- [76] Manly, J.J., Byrd, D., Touradji, P., et al., 2004. Literacy and cognitive change among ethnically diverse elders. *International Journal of Psychology*. 39(1), 47–60. DOI: <https://doi.org/10.1080/00207590344000286>
- [77] Franzen, S., European Consortium on Cross-cultural Neuropsychology (ECCroN), Watermeyer, T.J., et al., 2021. Cross-cultural neuropsychological assessment in Europe: Position statement of the European Consortium on Cross-cultural Neuropsychology (ECCroN). *The Clinical Neuropsychologist*. 36(3), 546–557. DOI: <https://doi.org/10.1080/13854046.2021.1981456>
- [78] Tiatri, S., Jap, T., 2015. Preliminary investigation on the effectiveness of a thinking skill training in Indonesia: "Thinking skills training with digital technology". *Journal of Psychological & Educational Research*. 23(2), 41–53.
- [79] Akhtar, H., 2022. Measuring fluid reasoning and its cultural issues: A review in the Indonesian context. *Buletin Psikologi*. 30(2), 276–288. DOI: <https://doi.org/10.22146/buletinpsikologi.74475>
- [80] Halima, A., Dewi, E.M.P., Damayanti, E., et al., 2022. AJT Cognitive Test: Respon harapan baru terhadap tes inteligensi anak di Indonesia. *Jurnal Sipakalebbi*. 6(2), 92–104. DOI: <https://doi.org/10.24252/sipakalebbi.v6i2.33060>