Proceeding of Biopsychosocial Issues
The 2nd International Conference on Biopsychosocial Issues
Semarang, 23 – 24 June 2022

Indonesia Neuropsychological Test Battery: Normative Score, Reliability, Age and Education Effects

Shinta Estri Wahyuningrum, Donders Centre for Cognition, Radboud University, the Netherlands & Computer Science Faculty, Soegijapranata Catholic University, Indonesia
Marc P.H. Hendriks, Donders Centre for Cognition, Radboud University, the Netherlands & Academic Centre for Epileptology (ACE), Kempenhaeghe, Heeze, the Netherlands
Gilles van Luijtelaar, Donders Centre for Cognition, Radboud University, the Netherlands
Augustina Sulastri, Faculty of Psychology, Soegijapranata Catholic University, Indonesia

: s.wahyuningrum@donders.ru.nl

Abstract
Neuropsychological tests are sensitive tools for measuring cognitive abilities, and they refer to the Assessment of healthy persons and various categories of patients. Some cognitive abilities are influenced by age and education, and some are not. Recently, ten neuropsychology tests were adapted for Indonesia, forming the Indonesian Neuropsychological Test Battery (INTB). It is the first neuropsychological test battery administered in Bahasa, Indonesia. This study presents preliminary normative scores, test-retest reliability, and the effects of age and education on each test. Data from four hundred and ninety healthy participants from Java (Jakarta, Semarang, and Surabaya), stored in a dynamic database (Indonesian-ANDI), were used. Preliminary normative scores of INTB are presented. All tests showed a moderate to good test-retest correlation coefficient ranging from 0.51 to 0.84, and an exception was the short and long-term recall scores of the RAVLT. Analysis of variance revealed that eighteen subtests were significantly age-dependent, and the scores tended to decline with the ageing process. Only the time to complete the Bourdon and RAVLT learning over trials did not decrease. In contrast, cognitive performance was increased along with a higher education level. The only exceptions were the time to complete the Bourdon and the RAVLT's learning over trials and delayed recall. These different effects of age and education on the tests of the INTB demonstrate the necessity to correct the normative score of the tests in a tailored way for these factors.

Keywords: INTB, age effect, education effects, neuropsychology, normative score

Introduction

Neuropsychological tests assess and evaluate human cognitive abilities (Zilmer, Spiers, & Culbertson, 2008). It is widely recognized that demographic factors may influence a person's cognitive ability. The most influential factors are age and education (Jansen et al., 2021; Murman, 2015). These influences can be positive or negative (Weber & Skirbekk, 2014; Jansen et al., 2021; Guerra-Carrillo, Katovich, & Bunge, 2017). Most cognitive abilities decline along with normal ageing, whereas some do not, or there might be an increase followed by a decrease (Glisky, 2007). Moreover, the peak and decline in cognitive performance vary widely for individuals or populations (Hartshore & Germine, 2015). In the majority of the studies, education has a positive impact on cognitive abilities: the performance on most of the tests will increase with higher education levels. However, there is also the condition when the ageing process and education do not change cognitive performance, or the change is seemingly insignificant.

Neuropsychological tests have been widely used worldwide, showing adequate evidence to measure individual cognitive abilities. However, in Indonesia, this is rarely done. The interpretation of the test scores requires normative data, and recently collected normative scores are mostly lacking for the Indonesian population. Therefore, recently a group of researchers adapted a series of neuropsychological tests for Indonesia, the Indonesian Neuropsychological Test Battery (INTB). The INTB consists of ten neuro-psychological tests covering three significant domains: learning and memory, language and executive function, including attention. The INTB was administered as a series of tests using "Bahasa Indonesia, or called Bahasa hereafter," in six different mainly urbanized parts of West, Middle and East Java, Bali, South Sulawesi, and East Kalimantan. Data gathered before the covid19 pandemic were collected from 890 healthy subjects. All data were stored in a "dynamic" database and an online platform called the Indonesian Neuropsychological Dynamic Infrastructure (I-ANDI) (Wahyuningrum et al., 2021). Like many other psychological tests to be used by practitioners, the INTB should have normative scores to interpret the tests' results correctly. A normative score is the expected test score from a sample mimicking the clients' demographic factors as much as possible (Zilmer, Spiers, & Culbertson, 2008). We used this data set to propose normative scores of all these ten tests for Indonesia, which are now standard practices worldwide (Lezak et al., 2012; Lövdén et al., 2020). We wanted to know whether the newly adapted tests for Indonesia are sensitive to demographic effects. If that is the case, the normative scores must be corrected for these demographic factors. In case age and education effects are found, and they mimic what is internationally reported, this contributes to the validity of the tests. Two earlier studies on a smaller sample, one on the Indonesian version of the Boston Naming Test (Sulastri et al. in 2019) and one on the Trail Making Test (Widhianingtanti et al., 2022) already suggested that this is indeed the case, now this will be investigated for all tests in the battery and a much larger sample.

Here we report the effects of education and age on the whole battery of the Javanese sample. More specifically, we investigate whether there are also age-dependent increases in cognitive performances, the age of a putative peak and a decline in the performances. Next, we also present preliminary normative scores and outcomes of the reliability analyses of all ten tests.

Methods

INTB was conducted using the paper and pencil method. The tests were: The Token Test (TT), Boston Naming Test (BNT), and phonemic Verbal Fluency Test (VFT), which were intended to assess language domain. The Stroop Test (ST), Bourdon Test (BWT), Trail Making Test (TMT), and Five Point test (FPT) were administered and aimed to assess the attention and executive function domain. The Digit Span (DS), Figural Reproduction (FR), and Rey Auditory Verbal Learning (RAVLT) represented the learning and memory domain. All participants completed the test battery, which took approximately two and a half hours.

Here we report the data from the Javanese population; four hundred and ninety healthy participants from three big cities in Java Island, Indonesia (Jakarta, Semarang, and Surabaya) were involved. The age range of the participants was from 16-80 years, with education levels ranging from elementary school (6 years) to doctoral programs (over 17 years). The age was categorized into six with decade intervals except for the youngest and oldest groups. The categories were: 16-19 years, 20-29 years, 30-39 years, 40-49 years, 50-59 years, and 60-80 years. Education was divided into four categories: 0-9 years, 10-12 years, 13-16 years, and above 17 years. Tests were conducted by a research assistant, a second-grade student from the psychology faculty who trained to collect data. Data collection was conducted in compliance with the Helsinki Declaration, and the ethics committee of Soegijapranata University gave clearance for this research project (University Ethical Clearance number: 001B/B.7.5/ FP.KEP/IV/2018).

Another fifty participants were involved in determining the test-retest reliability of the INTB. We used *Pearson Correlation Coefficient* with a different dataset. The test-retest interval was three weeks. The score of each test was converted into a z-score to facilitate comparisons between scores of the various tests. The inverse scores of all variables that measured time and number of errors were used.

Analyses of Variance (*ANOVA)* were used to determine the age and education effects of the ten test scores. Basic rules of thumb for *partial eta squared* (η^2) as a measure of the effect size are that η^2= .01 indicates a small effect; η^2= .06 indicates a medium effect; η^2= .14 indicates a significant effect.

Results

The minimum and maximum score, mean and standard deviation of 490 participants for the ten tests are shown in Table 1. Means and standard deviations were calculated using data from all participants (N= 490). Normative scores for different combinations of age and education could not be given as yet because the number of participants in certain groups is still less than 50 (Bridges & Holler, 2007). For example, in the low education group, with a government program of nine years of compulsory education, it becomes challenging to find participants in this group. In addition, minimum and maximum scores explain the range of performance for all participants.

Table 1. Mean, standard deviation, minimum, and maximum of ten tests (n = 490)

Variables	Min; Max	Mean	Standard Deviation
BNT score	34; 60	51.37	4.95
BNT Time	60; 848	308.81	161.274
FR score	4; 15	11.87	2.54
DS forward	1; 14	7.48	2.24
DS backwards	0; 15	6.28	2.40
DS sequence	0; 16	7.74	2.89
BWT score	118; 327	193.14	13.82
BWT error	0; 58	10.07	9.57
ST score card 3	81; 100	97.76	3.01
ST score card 3-2	-19; 18	-1.16	3.37
ST total error	0; 22	3.29	3.56
RAVLT Mean A1-A5	4; 14.60	10.05	2.00
RAVLT LOT	-5; 46	16.87	7.80
RAVLT STPR	37.5; 150	90.01	16.68
RAVLT LTPR	36.36; 137.50	88.62	17.17
VFT total score	6; 90	40.59	13.36
TT score	9; 163	146.86	22.85
FPT unique number	3; 58	25.97	9.52
TMT Time A	9; 134	44.99	18.79
TMT Time B	17; 426	87.52	50.67

Pearson correlation coefficient revealed the test-retest reliability of the ten tests. The data are presented in Table 2. Good reliability (> .80) was introduced by TT, BNT, and TMT time B scores. Moderate coefficients were revealed by DS forward, DS backwards, DS sequence, FR score, FPT unique number, TMT time A, ST card 3, BWT score, RAVLT mean A1-A5, and VFT score with a range of .52 to .78. At the same time, a low correlation coefficient was found for RAVLT LOT, RAVLT STPR, and RAVLT LTPR, ST card 3-2, ST error with coefficient values less than .49.

Table 2. Pearson correlation coefficient of ten tests (n=50)

Variables	r
BNT score	.88
BNT Time	.84
FR score	.73
DS forward	.78
DS backwards	.76
DS sequence	.60
BWT score	.54
BWT error	.86
ST score card 3	.52
ST score card 3-2	.35
ST total error	.49
RAVLT Mean A1-A5	.78
RAVLT LOT	-.15
RAVLT STPR	.19
RAVLT LTPR	.10
VFT total score	.72
TT score	.84
FPT unique number	.67
TMT Time A	.66
TMT Time B	.81

Table 3 shows the results of the ANOVA on the age effects for ten tests. Age had a large impact as expressed by a measure of effect size on RAVLT mean A1-A5 (η^2= .210), FPT unique number (η^2 = .141), TMT time A (η^2 = .228), and TMT time B (η^2 = .203), the other variables showed a medium to large effect size. There was no statistically significant age effect for the BWT score and RAVLT LOT (p=.207 and p =.308, respectively).

Table 3. Age effects of ten tests.

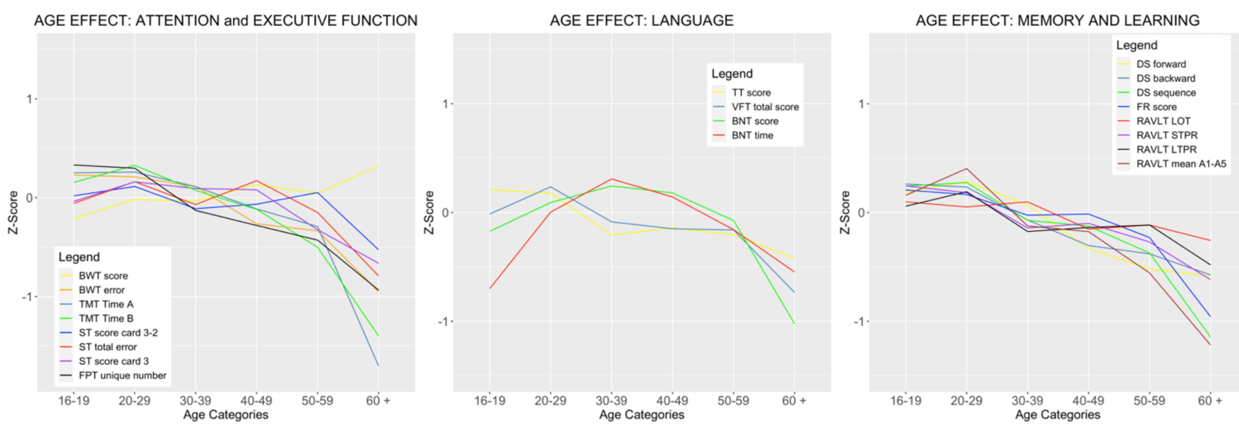
Variables	F	sig	Partial eta squared (η^2)
BNT score	8.98	.000	.085
BNT Time	3.92	.002	.039
FR score	8.47	.000	.108
DS forward	11.77	.000	.108
DS backwards	8.71	.000	.083
DS sequence	15.84	.000	.083
BWT score	1.44	.207	.015
BWT error	11.417	.000	.106
ST score card 3	5.52	.000	.054
ST score card 3 - 2	2.49	.031	.025
ST total error	5.84	.000	.057
RAVLT mean A1-A5	25.68	.000	.210

Table 3. Age effects of ten tests.

Variables	F	sig	Partial eta squared (η^2)
RAVLT LOT	1.20	.308	.012
RAVLT STPR	6.07	.000	.059
RAVLT LTPR	3.86	.002	.038
VFT total score	6.56	.000	.063
TT score	4.45	.001	.044
FPT unique number	15.85	.000	.141
TMT Time A	28.66	.000	.228
TMT Time B	24.73	.000	.203

Note: BWT score and RAVLT LOT have no significant age-effects

Figure 1, left below, illustrates the domain attention and executive function age trends. The overall trend for this domain tends to decline except for performance on the BWT time score. The time to complete the test was not age dependent; in contrast, the number of errors increased with ageing. The lines with colours light blue (TMT time A), green (TMT time B), and black (FPT unique) show the age-dependent effects of these tests. The three tests show the same onset decline at 30 years old and a significant decrease in the most senior age group (60+). The three variables of the Stroop Test show a substantial reduction in the age above 60, although, for the Stroop card 3, the decline started a bit earlier, at the age of 50 and continued with a slight decrease at age 60 and above.



The X-axis has six age categories, and Y-axis is the z-score. Colour lines show the performance of each variable for the various tests in the three domains.

Figure 1. Age-effects for attention and executive learning domain (left side), language domain (middle side), and memory and learning (right side).

The following graph on the middle illustrates the language domain's age trends—all subtests on this domain show medium to significant effects. The oldest category showed the most significant decline. Interestingly, we found the peak in the performance on the BNT, not for the younger category but the 30-39 years category. In contrast, VFT and Token tests showed a decline at 30, but both tests' performances remained stable in the following age category.

The last graph illustrated the memory and learning domain. All subtests show a declining trend except for RAVLT LOT, the line with the light red colour. As mentioned in Table 2, LOT was not significantly found to be affected by the ageing process; the figure shows a straight line. This means that this verbal learning test's "learning over trials" was relatively stable for all age categories. In contrast, RAVLT represents A1-A5 or the mean of the recall of the five first recall trials, which showed a significant age effect. This trend, demonstrated by the line-colour dark red, shows that the peak performance occurred at age 20 and continues with two declines, first at age 30 and later at 50. Other variables show a similar trend, starting at age 30 and significantly declining for the oldest categories at 60+.

Table 4. Education effects of ten tests.

Variables	F	sig	Partial eta squared (η^2)
BNT score	43.04	.000	.210
BNT time	30.42	.000	.158
FR score	18.25	.000	.101
DS forward	22.52	.000	.122
DS backwards	18.73	.000	.104
DS sequence	22.88	.000	.124
BWT score	.92	.429	.006
BWT Error	10.60	.000	.061
ST score card 3	12.67	.000	.073
ST score card 3 - 2	10.54	.000	.061
ST total error	17.61	.000	.098
RAVLT mean A1-A5	14.27	.000	.081
RAVLT LOT	.66	.576	.004
RAVLT STPR	5.08	.002	.030
RAVLT LTPR	.043	.988	.000
VFT total score	41.19	.000	.203
TT score	22.43	.000	.122
FPT unique number	17.38	.000	.097
TMT time A	27.13	.000	.143
TMT time B	64.34	.000	.284

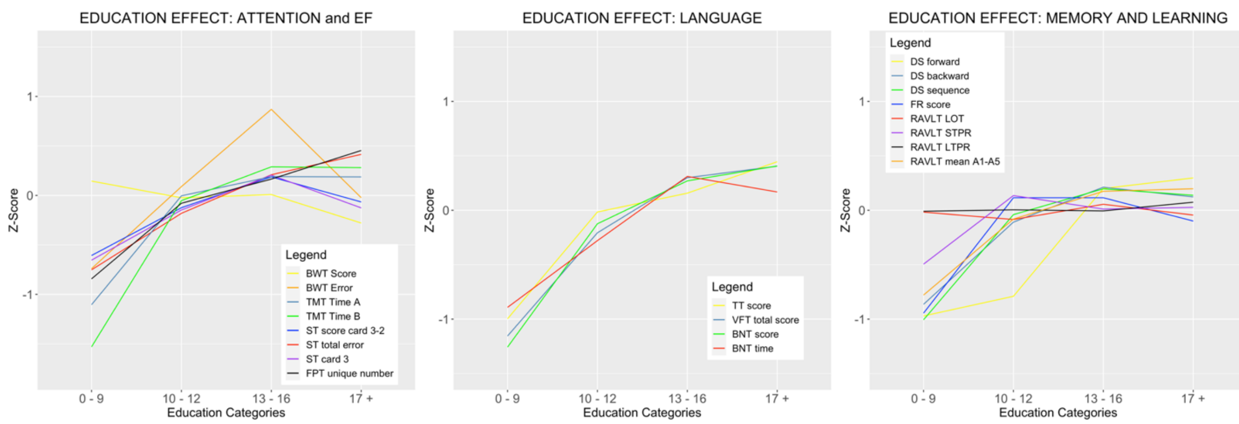
Note: BWT score, RAVLT LTPR, and RAVLT LOT have no significant education-effects

As presented in Table 4, the education effect revealed only three subtests were insignificant. In contrast, the rest of the subtests were significant, with a p-value below .01. The subtests were BWT score (p=.429), RAVLT LOT (p=.576), and RAVLT LTPR (p=.988). Interestingly, on the subtest, TMT time A, TMT time B, VFT score, BNT score, and BNT time found significant education effects with an effect size above .14.

The effect of education on the ten tests is illustrated in Figure 2. Overall, cognitive trends in education categories increased along with educational attainment. The worst performance was the lowest education categories, except for BWT score and RAVLT LOT and LTPR, which are not significantly affected by education.

The left side graph demonstrates performance in the attention and executive function domain. On TMT performance, both subtests show a significant increase in the second education category followed by a slight increase in the following categories and tend to be stable on the differences for the last two categories. The consistent increasing performance found on the FPT and Stroop total error, illustrated by black and red lines, shows a positive linear correlation. While on the three subtests (Stroop Card 3-2, Stroop score card 3, and Bourdon-Wiersma error), there is an increase in the first three education categories and continues with a slight decline for the highest category.

The language domain tests correlate positively with education attainment, which is presented on the middle side graph. Performance on language rose in the second and third categories but did not significantly increase for the highest category. The last domain, learning and memory, is shown in the right side graph in the middle. Some variables in this domain significantly increase from the lowest to the next education category, followed by a flatline. A different trend is DS forward, which shows a significant increase a bit later, at 13 years of education. In contrast, a trend is shown by the black and red flatline describing the performance on RAVLT LOT and LTPR, which have no significant changes in those performances along the education categories.



The X-axis has six education categories, and Y-axis is the z-score value. Colour lines show the performance of each variable subtest from each domain.

Figure 2. Education-effects for attention and executive learning domain (left side), language domain (middle side), and memory and learning (right side).

Discussions

We analyzed the effects of the demographic factors using 490 subjects to see whether the tests were sensitive to commonly reported age and education effects. The implication would be that the normative scores for these tests needed to be adapted and corrected for these demographic factors. We propose only preliminary scores of the whole Javanese sample because age and education effects were found, demanding different normative scores for different age and education groups. Since, for most of the combination of age and education categories, the number of subjects was less than 50, and this number is necessary to obtain a reliable and representative score per subcategory, we present only preliminary normative scores for ten tests (Bridges & Holler, 2007). Next, we offer the reliability of the tests, next to age and education effects.

In general, the reliability of the ten tests was also good enough, proving that the tests were reliable for the Indonesian population. However, we found a notable trend in the result of the test-retest reliability of the RAVLT on three of four investigation scores (LOT, STPR, and LTPR). Apart from the low reliability of STPR (0.19) and LTPR (0.10), we found negative results for LOT. The latter indicates that the mean score of the retest was lower than the test. The South African adolescent study also revealed a lower mean score on the second test (Blumenau, 2011). The low test-retest reliability in our analysis of the RAVLT was most probably due to the that we used the same word list in both sessions, and this had a consequence that the fifteen items were still in the participants' memory as expressed by higher memory scores and a lower increase over trials. The reliability score for the recall of the first five trials was close to .8, suggesting that this aspect was less bothered by using the same wordlist.

In general, the trend on the age effect shows that almost all tests showed the anticipated decline (Cohen et al., 2019; Elkana et al., 2015). In nearly all tests or variables, younger people outperform older, except in the language domain, where the middle age categories perform better than others. From all subtests, we conclude that the most senior category performed poorer in all tests, and an accelerated decline occurred around sixty. As expected, not all variables or tests start to decline at the same age, most start at age 30, and some come later (Glisky, 2007). A very general conclusion is that this study confirms that the Indonesian population's cognitive performance is influenced by ageing but that the ageing process is different for the different tests.

Regarding the education effect, we found that participants with low education performed the worst, in agreement with what is generally reported (Guerra-Carrillo et al., 2017). The significantly better performance was shown by those with at least ten years or more of education. For the domain language and executive function, including attention, better performance gradually increased until the category with the most years of schooling. Different from both domains, there is no significant improvement for memory and learning domains on the educational level after 13 years of education.

From the ten tests, we emphasize that the Bourdon-Wiersma score time and RAVLT LOT have no significant effects on either age or educational level. Bourdon-Wiersma score was used to measure the ability to maintain accuracy and concentration while looking for a group of 4 dots. Being attentive and scanning the lines fast is an ability that remains intact in older people and seems already present even after elementary school. The number of errors of the Bourdon-Wiersma test was age-dependent and did not deviate much from the cognitive decline found for other executive functions. Interestingly, although there is no significant age effect, the Bourdon scores show a slight increase in the last age categories. We expect our older participants to be more patient and careful when completing this test. RAVLT LOT is used to measure the ability to learn over the trial. Our population shows this ability remains stable even for the oldest and less educated groups.

We found that three subtests have a significant impact on the ageing process. The first was the RAVLT mean of trial A1 to trial A5. This variable represented the average immediate verbal recall. A sudden decline happened from the early young older category (40 years old) and continued to the most senior category. The second was FPT, a score of unique number design associated with creativity and mental flexibility, the decline starts earlier, at age 30, and the decline continues until the oldest category. The third was TMT time A and time B; both subtests were about speed to visually detect and recognize the sequence of numbers or letters. The knowledge that speed processing is also age-dependent decline (Salthouse, 2010) might be helpful for the next researcher on the neuropsychological test, which measures speed or visual or auditory ability with participants who are elderly, to have a pre-test for the perception abilities for fairness of the results.

Furthermore, education level largely influences the performance of VFT total score and BNT time and total correct score, and these tests belong to the language domain. Some previous studies conclude that performance in the language domain remained stable and increased along with educational attainment (Murman, 2015). In BNT performance, we expect that the vocabulary and knowledge were raised during the years of education and the better jobs that the better educated have. Other subtests which gave significant effects were TMT time A and B. In addition to measuring speed, this test also measures working memory. Concerning the VFT total score, a participant was asked to produce words beginning with a specific letter. Both test scores increase at higher levels of education (Troyer, 2000).

Conclusions

The scores on ten cognitive tests for the Javanese population highly depended on age and education. The performance of the elderly was the lowest compared to all age categories. Participants with the lowest education level performed poorly on almost all cognitive tests. Tests in the language domain were the most sensitive for education, while attention, executive function, memory, and learning became more stable after ten years of schooling. The age and education effects of the tests of the INTB imply that representative and valid normative data need to be developed for these different categories of Javanese people, adapted for age and education.

Acknowledgement

This work was supported by the Directorate of Higher Education General of Indonesia with Dikti grant number 0317/AK.04/2022. We also thanked the Indonesian Neuropsychological Consortium, research assistants, and all participants who contributed to the current project.

References

Zillmer, E.A., Spiers, M.V., & Culbertson, W.C. (2008). Principles of neuropsychology. Higher Education, 574. Retrieved from http://books.google.com/books?id=wIk1PwAACAAJ&pgis=1

Jansen MG, Geerligs L, Claassen JAHR, Overdorp EJ, Brazil IA, Kessels RPC and Oosterman JM. (2021). Positive Effects of Education on Cognitive Functioning Depend on Clinical Status and Neuropathological Severity. Front. Hum. Neurosci. 15:723728. DOI: 10.3389/fnhum.2021.723728

Murman, D. L. (2015). The Impact of Age on Cognition. Seminars in Hearing, 36(3), 111–121. https://doi.org/10.1055/s-0035-1555115

Weber, D. & Skirbekk, V. (2014). The Educational Effect on Cognitive Functioning: National versus Individual Educational Attainment. IIASA Interim Report. IIASA, Laxenburg, Austria: IR-14-008

Guerra-Carrillo, B., Katovich, K., & Bunge, S. A. (2017). Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample. In PLoS ONE (Vol. 12, Issue 8). https://doi.org/10.1371/journal.pone.0182276

Sulastri, A., Utami, M. S. S., Jongsma, M., Hendriks, M., & 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(11), 134–139.

Wahyuningrum, S.E., van Luijtelaar, G., & Sulastri, A. (2021). An online platform and a dynamic database for neuropsychological Assessment in Indonesia. Applied Neuropsychology: Adult, 0(0), 1–10. https://doi.org/10.1080/23279095.2021.1943397

Hartshorne, J.K. & Germine, L.T. (2015). When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. Psychological Science, 26(4), 433–443. https://doi.org/10.1177/0956797614567339

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). Neuropsychological Assessment (5th ed.). Oxford University Press.

Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020). Education and Cognitive Functioning Across the Life Span. *Psychological Science in the Public Interest*, 21(1), 6–41. https://doi.org/10.1177/1529100620920576

Glisky, E. (2007). Changes in Cognitive Function in Human Aging. Brain Aging: Models. Method, and Mechanisms (Issue April 2007, pp. 3–20). https://doi.org/10.1201/9781420005523.sec1

Salthouse T. A. (2010). Selective review of cognitive aging. Journal of the International Neuropsychological Society: JINS, 16(5), 754–760. https://doi.org/10.1017/S1355617710000706

Troyer A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. Journal of clinical and experimental neuropsychology, 22(3), 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370

Bridges, A. J, & Holler, K.A. (2007). How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. Child Neuropsychology, 13(6), 528–538. https://doi.org/10.1080/09297040701233875

Blumenau, J., & Broom, Y. (2011). Performance of South African Adolescents on Two Versions of the Rey Auditory Verbal Learning Test. South African Journal of Psychology, 41, 228 - 238.

Cohen, R. A., Marsiske, M. M., & Smith, G. E. (2019). Neuropsychology of aging. *Handbook of Clinical Neurology*, 167, 149–180. https://doi.org/10.1016/B978-0-12-804766-8.00010-8

Elkana, O., Eisikovits, O. R., Oren, N., Betzale, V., Giladi, N., & Ash, E. L. (2015). Sensitivity of neuropsychological tests to identify cognitive decline in highly educated elderly individuals: 12 months follow up. *Journal of Alzheimer's Disease,* 49(3), 607–616. https://doi.org/10.3233/JAD-150562

Guerra-Carrillo, B., Katovich, K., & Bunge, S. A. (2017). Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample. *PLoS ONE* (Vol. 12, Issue 8). https://doi.org/10.1371/journal.pone.0182276

Widhianingtanti, L.T, van Luijtelaar, G, Suryani A.O, Hestyanti, Y.R. & Sulastri A. (2022). Indonesian Trail Making Tets: Analysis of Psychometric Properties, Effects of Demographic Variables, and Norms for Javanese Adults. *Jurnal Psikologi.* Vol 49, no 2, pp 104-122, DOI: 10.22146/jpsi.68953.

Bridges, A. J., & Holler, K. A. (2007). How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. Child Neuropsychology, 13(6), 528–538. https://doi.org/10.1080/09297040701233875