Psikohumaniora: Jurnal Penelitian Psikologi
Vol 9, No 1 (2024): 125–146. DOI: 10.21580/pjpp.v9i1.20957
ISSN 2527-7456 (online); 2502-9363 (printed)

Psychometric properties of the Five-executive Function Tests in Indonesian samples

Ni Made Swasti Wulanyani,[1*] **Lucia Trisni Widhianingtanti**,[2] **Aria Saloka Immanuel**,[1] **Aireen Rhammy Kinara Aisyah**,[3] **Marc P. H. Hendriks**,[3,4] **Yohana Ratrin Hestyanti**,[3] **Angela Oktavia Suryani**,[3] **Gilles van Luijtelaar**[4,5]

[1]Department of Psychology, Faculty of Medicine, Universitas Udayana, Denpasar – Indonesia; [2]Department of Psychology, Faculty of Psychology, Universitas Katolik Soegijapranata, Semarang – Indonesia; [3]Department of Psychology, Faculty of Psychology, Universitas Katolik Indonesia Atma Jaya, Jakarta – Indonesia; [4]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen – the Netherlands; [5]Department of Psychology, Faculty of Psychology, Universitas Kristen Maranatha, Bandung – Indonesia

Abstract: Difficulties in executive function (EF) may be causative factors in clinical conditions, including learning disabilities, depression, and anxiety. However, defining the structure of EF is an ongoing issue, although Miyake's three-factor structure is widely accepted. This study aims to discover the underlying factor structure of EF domains, as measured by five tests and the scores of 840 participants, to establish the domains' validity and reliability. Five neuropsychological tests were included in the exploratory and confirmatory factor analyses (EFA and CFA): the Digit Span (DS); the Stroop Color Word Test (Stroop Test); the phonemic Verbal Fluency Test (pVFT); the Five Point Test (5PT); and the Trail Making Test (TMT). EFA could be meaningfully performed (KMO = .862, Bartlett's test: χ^2=1393.169, df 45, $p < .001$) and yielded three factors closely mirroring the three core domains of EF. CFA of the model in which the 5PT loaded on factor 2 showed the best goodness of fit indicators, although the difference with the other three-factor models was small. Currently, these EF tests are available in Indonesia and suitable for use, although obtaining normative scores adjusted for demographic factors is necessary.

Keywords: clinical condition; executive function; neuropsychological adapted test

To cite this article (APA Style):
Wulanyani, N. M. S., Widhianingtanti, L. T., Immanuel, A. S., Aisyah, A. R. K., Hendriks, M. P. H., Hestyanti., Y. R., Suryani, A. O., & van Luijtelaar, G. (2024). Psychometric properties of the Five-executive Function Tests in Indonesian samples. *Psikohumaniora: Jurnal Penelitian Psikologi*, *9*(1), 125-146. https://doi.org/10.21580/pjpp.v9i1.20957

*__Corresponding Author:__ Ni Made Swasti Wulanyani (swastiwulan@unud.ac.id), Department of Psychology, Faculty of Medicine, Universitas Udayana, Jl. P. B. Jend. Sudirman, Denpasar, Bali, 80232 – Indonesia.

Submitted: 6 Mar 2024; Received in revised form: 23 May 2024; Accepted: 26 May 2024; Published regularly: May 2024

Introduction

Executive functions (EFs) are a rather broad concept, referring to a cognitive domain often used in the neuropsychological literature based on varying descriptions from different authors. Chan et al. (2008) propose that EF includes a variety of cognitive and behavioral processes, such as verbal reasoning, problem-solving, planning, the ability to maintain attention, resistance to distractions; the ability to perform multiple tasks at once the ability to act and respond quickly to tasks of varying complexity and the ability to face new challenges. Lezak et al. (2004) defined EF as the capacity to adapt to unfamiliar situations and considered them as the basis of many cognitive, emotional and social abilities.

Executive function is included in the field of cognitive psychology, which studies how humans receive, retain, change and use knowledge. Its focus includes attention, perception, learning, memory, judgment, decision making and problem solving. It is impossible to talk about human life without considering the capacity to absorb knowledge. All behavior must have a mental aspect, even in its simple forms. Cognition involves unique human language, personal identity and culture (Peterson, 2006). The context and quality of life determines on what and how human focusing attention (Koci & Donaldson, 2023).

Many researchers acknowledge that EF refers to a range of cognitive regulatory processes involved in the capacity to organize ideas and behavior, and includes three main domains: working memory, inhibitory control, and cognitive flexibility (Diamond, 2013; Logue & Gould, 2014; Miyake et al., 2000; Ren et al., 2017). Synder et al. (2015) hypothesize that the prefrontal cortex (PFC) supports the cognitive control processes of EF. Such processes enable self-regulation and self-directed behavior toward goals. Cognitive regulatory processes include breaking unwanted habits; making decisions and evaluating risks; planning for the future; prioritizing and sequencing actions; and coping with novel situations. Executive function is related to human problem-solving abilities, which are a great aid to humans in facing difficulties in their lives, and help to reduce depression, and increase optimism, hope, self-esteem and emotional well-being (Lebon, 2014). Problems with EF may have a wide range of consequences in individuals' lives, from how they interact with others to how they study and work (Snyder et al., 2015).

EF provides a distinct, added benefit beyond intelligence. It enables common cognitive tasks such as decision-making and problem-solving in ambiguous, unaccounted-for situations, often with time constraints (Chan et al., 2021). Executive dysfunctions hinder physical and mental health, and in clinical settings poor or impaired EF could co-occur with multiple forms of psychopathology (Snyder et al., 2015). Early developmental difficulties in EF could be causative factors in different developmental clinical conditions, including learning disabilities, attention-deficit/ hyperactivity disorder, conduct disorder, autism spectrum disorder; obsessive–compulsive disorder, depression; and anxiety (Zelazo, 2020). Adequate measurement and quantification of EF are necessary, considering that its effective development is important in many aspects of an individual's personal and social life, their professionalism, and even predictive for adequate leadership (Ramchandran et al., 2016).

Defining the underlying cognitive structures of EF has become an ongoing controversy. Various studies have proposed different factor models, recently, for example, Pires et al. (2019). Im-Bolter et al. (2006) proposed a two-layer, four-factor model, while Himi et al. (2021) put forward five-factor and two-layer-six-factor ones. Other researchers have agreed on the three EF core elements: working memory, inhibitory control and cognitive flexibility (Diamond, 2013; Logue & Gould, 2014; Miyake et al., 2000; Ren et al., 2017), although such a three-factor model continues to develop. It should also be realized that the underlying model is dependent on the type and number of EF tests included in the factor analysis, and possibly also on the population being studied. Moreover, the commonly used EF tests were developed a long time before the development of the underlying models.

Working memory, the first core element of the Miyake et al. (2000) model, is a cognitive process that can control and encode recently received information, extract, sort, manipulate and select it, and combine it with other information that is currently available to accomplish the ongoing mental task (Huizinga et al., 2006). The EF component of working memory is responsible for retaining information for a certain period and is a domain-general supervisory or executive attention mechanism (Engle, 2010). Inhibitory control, the second element of Miyake's EF model, is defined as the ability to actively inhibit or delay dominant reactions to achieve goals (Morasch & Bell, 2011). Cognitive flexibility, the third core element, refers to the ability to switch between tasks, situations or actions and to think creatively (Diamond, 2013; Miyake et al., 2000).

As discussed, EF involves several different subdomains and therefore cannot be measured by a single test (Danielsson et al., 2010, 2012). Consequently, some have researchers used tests covering different components of EF (Fleming et al., 2016; Friedman et al., 2016; Ito et al., 2015), whereas others have used more than one EF test to assess a single EF component or sub-component (Huizinga et al., 2006).

Recently, ten neuropsychological tests were introduced in Indonesia by a consortium of six universities, known as the Indonesia Neuropsychological Test Battery (INTB) (Wahyuningrum et al., 2022). These ten tests cover a wide range of cognitive functions. They are used internationally and have clinical relevance. They were adapted where necessary and preliminary normative scores for the Indonesian population were proposed (Wahyuningrum et al., 2022). Five of these tests, namely the Digit Span (DS); the Stroop Color Word Test (Stroop Test) adapted for Indonesia; the phonemic Verbal Fluency Test (pVFT) adapted; the Five Point Test (5PT); and the Trail Making Test (TMT), are employed to measure different aspects of EF (Sherman et al., 2022).

In Indonesia, validated tests and normative data adapted for age, education, sex, and perhaps for the daily language spoken and ethnicity, are currently lacking, which hampers their appropriate use in the country, where most often normative data from decades ago from western countries are still used. Having valid, reliable, sensitive instruments that have been suitably psychometrically validated and are able to assess the EF of individuals in Indonesia is vital. This will result in appropriate interpretation of the test scores of Indonesian people. Therefore, the primary aim of this study is to determine the construct validity of the five tests listed above. These were selected and included in the INTB because they represent different aspects of executive functioning, such as inhibition, working memory, and different aspects of mental flexibility, and because of their widespread proven utility in clinical settings, with their huge diversity of neuropsychological patients, and their associated symptoms and cognitive dysfunctions.

This research exploratory and confirmatory factor analysis (EFA and CFA), as proposed by researchers such as Kyriazos (2018) and Canivez et al. (2019). The study aims to identify the factor structure underlying the five EF tests. If the results are able to identify a similar factor structure to that previously reported internationally, for example, the three-factor model of Miyake (Miyake et al., 2000), this will contribute to the validity of the five EF tests in our battery. The second study aim is to establish a reliability measure of the identified factors, while the third aim is to establish whether the tests are sensitive to demographic factors such as age and education through correlation analysis, and whether there are sex differences. If this is the case, this will also contribute to the validity of the tests, considering that education and age are known to have a major impact on the outcomes of EF tests.

Methods

Participants

The data utilized in the study were gathered from five neuropsychological tests. The data were kept in a dynamic database that included demographic information about the participants (Wahyuningrum et al., 2023), who were recruited through purposive sampling from six different regions of Indonesia: West, Central, and East Java, Bali, East Kalimantan and South Sulawesi. They declared themselves to be healthy and with no history of neurological or psychiatric diseases, brain injury or drug abuse, either currently or in the past. All the participants completed all of the tests and a questionnaire that included demographic variables such as sex, age, years of education, place of birth, marital status, and ethnic group of both parents. In total, 840 participants took part in the study, of whom 518 were females, and the mean age was 35.45 (SD age =15.25), representing the relatively young Indonesian population mainly from the west.

Age was classified into six groups: 16-19 years (n = 87); 20-29 years (n = 298); 30-39 years (n = 141); 40-49 years (n = 126); 50-59 years (n = 128); and >60 years (n = 60). Years of education were classified into five groups based on the Indonesian education system: 0-6 years of education (elementary school; n = 36); 7-9 years of education (junior high school; n = 58); 10-12 years of education (senior high school; n = 300); 13-16 years of education (at least undergraduate; n = 399); and >17 years of education (graduate or postgraduate; n = 47).

Research Procedure

The tests were administered in the participants' homes, with the research assistants ensuring that they were conducted in a quiet environment free from distracting noise. Prior to beginning the tests, the assistants provided an explanation of the study procedures, including the use of the data for scientific purposes and their anonymous storage in our database. The participants provided their consent to take part in the study. Upon completing the tests, they received a payment of 75,000 rupiahs, equivalent to five US dollars. The research assistants were third-year psychology students who had received training in test administration and psychometrics. The tests were conducted in the official language of Indonesia, "Bahasa Indonesia," which is used nationwide in education, media, administration and business. All the participants completed the full battery of tests. The research was conducted in compliance with the Helsinki Declaration, and the ethics committee of Soegijapranata University granted clearance for the project (University Ethical Clearance number: 001B/B.7.5/FP.KEP/IV/2018).

Instruments

Five EF tests, part of the Indonesian Neuropsychology Test Battery (INTB), were used. The data were recently collected and stored in our dynamic database (Wahyuningrum et al., 2023).

Digit Span Test (DS)

Digit Span (DS) is an assessment commonly used to measure verbal short-term and working memory, and this encompasses an attention component. DS consists of three subscales, forward, backward and sequence, and is based on the Indonesian version of the Wechsler Adult Intelligence Scales (Suwartono et al., 2014). The forward subscale assesses the capacity of the auditory memory buffer and attention. In contrast, the backward and sequence subscales require more cognitive control, such as manipulating information within the memory buffer, and are considered as working memory and therefore represent this core construct (Hale et al., 2002).

In the DS test, participants are verbally presented with a series of numbers, ranging from two to nine digits, and they must repeat them in the correct order, receiving a point for each correct repetition of a series. Only the forward subscale DS Forward (DS_FW) begins with a three-digit sequence. The DS Backward (DS_BW) subscale assesses the ability to repeat the series of numbers in reverse order, while the DS Sequence (DS_SQ) subscale requires participants to repeat the series of numbers in ascending order. The testing of each subscale is stopped after two consecutive incorrect series of digits of the same length. The sum of the scores of the three subtests (DS_TOT) is commonly used in clinical settings.

Trail Making Test (TMT)

The second EF test, the Trail Making Test (TMT), requires cognitive flexibility, as it asks individuals to shift between two distinct categories: digits and letters. Cognitive flexibility is one of the three core constructs of EF. To establish a baseline evaluation of motor and visual search speed, Part A of the TMT is administered first, with Part B administered next. Executive abilities are evaluated by assessing the capacity to be flexible and switch between two mental sets (Gläscher et al., 2012; E. Strauss et al., 2006). Since participants are asked to remember the previous letter or digit while performing the test, attention and working memory may also be involved in these tasks. Recently, normative scores were published for this test from young Javanese individuals (Widhianingtanti et al., 2022). The primary output variable is the time taken to complete the two parts. Besides cognitive flexibility, the test requires working memory (WM), as participants must recall the previous letter or digit and update this during the execution of the test (Lezak, in Giovagnoli et al., 1996; Gläscher et al., 2012; E. Strauss et al., 2006). Part A of the TMT requires participants to consecutively connect numbered circles on paper in ascending order from 1 to 25 using a pencil, as quickly as possible and without lifting the pencil from the paper. This provides a baseline assessment of motor and visual search speed, visual scanning, number identification, numeric sequencing, and motor speed. In contrast, Part B is a more direct measure of executive abilities, involving both numbers (1-13) and letters (A-L) and necessitating alternation between the two categories (e.g., 1-A-2-B-3-C, etc.). The time to complete Part B minus that to complete Part A is considered to be a purer measurement of cognitive flexibility. If there is an error in following the sequence, the tester points it out and asks the participant to correct it.

Phonemic Verbal Fluency Test (pVFT).

The phonemic Verbal Fluency Test (pVFT) is a measure of the verbal production of unique words within a limited time frame (one minute), starting with a specific letter. It requires individuals to access and retrieve words from their mental lexicon while switching between different semantic categories (Aita et al., 2019). Therefore, this task also measures cognitive flexibility.

In Indonesia, the letters S, K, and T were previously selected and used as subscales (Hendrawan & Hatta, 2010; Pesau & Luijtelaar, 2021). Scores are based on the number of correct and unique words generated in each of the three subscales, as well as the sum of the three subscales.

Verbal fluency tests are widely used in clinical neuropsychological practice as a sensitive marker of brain dysfunction. They involve the neuro-anatomical areas associated with executive and language function, including self-monitoring, inhibition and working memory (Whiteside et al., 2016).

Five Point Test (5PT)

The 5PT is also a cognitive flexibility test. It assesses spontaneous figural fluency, or flexibility and creative thinking. The test measures a person's ability to generate a diverse range of new ideas and solutions (Ionescu, 2012; E. Strauss et al., 2006). It consists of a configuration of 40 five-point matrices that must be connected using a pencil. The participant's task is to create as many unique patterns as possible by connecting at least two points in each matrix within a limited amount of time, usually three minutes. The number of unique patterns and of perseverance errors are the output variables.

Stroop Test

The Stroop Color Word Testis an EF task used to assess inhibition, the third cognitive EF domain. The test measures the ability to stop or inhibit dominant, automatic or pre-potent responses, including stopping an action or thought, as well as the sensitivity to interference (Ferguson et al., 2021; Kurniawan & Kusrohmaniah, 2018; Panerai et al., 2014; Testa et al., 2012). The Stroop test has three subtests, each consisting of 100 stimuli. Subtest 1 requires participants to read the names of colors printed in black. Subtest 2 is a color-naming task using three differently colored rectangles, while subtest 3 is an incongruent word-color naming task with incongruent word-color combinations, in which the color words are printed in different ink colors (e.g., the word "red" is printed in blue). In this subtest, participants are asked to name the color (blue) and ignore the word "red." A commonly used outcome variable is the difference in time to complete subtest 3 minus the time to complete subtest 2. The number or percentage of correct answers produced by the participants then include to be analyzed.

Data Analysis

The dataset (n = 840) was separated into two parts: the odd-numbered subjects were included the odd dataset (n = 420), and the even ones in the even dataset (n = 420). Two different and independent data sets were used to explore and confirm the factor structure of the five EF tests in the Indonesian sample. It also avoided overfitting goodness of fit indicators on the EFA and CFA in the sample (Fokkema & Greiff, 2017).

First, EFA was used to explore and identify the number of factors in the odd dataset. The variables that were included in the factor analyses were the number of correctly generated items for the three subscales of the DS (FW, BW, SQ); the number of unique figures in the 5PT; the number of correct words in pVFT (Letter S, K, T); the number of correct items in the Stroop Card 3, and the Stroop Card 3 time minus Card 2 time, together with the time of TMT (Part B- Part A). These ten variables representing the five EF tests were included in the two types of factor analysis. Principal axis factoring with promax oblique (*oblimin*) rotation was used to measure the common variance between constructs and to reveal their latent structure (Mvududu & Sink, 2013; Tabachnick & Fidell, 2019).

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity were used to assess the suitability of the data for factor analysis. A KMO correlation above .50 is considered adequate for a meaningful analysis of the EFA output (Field, 2013). A scree plot and parallel analyses were used to identify the number of factors constructed in the odd dataset by observing the point of inflection. Factor loadings larger than .25 were interpreted.

Second, CFA was used to confirm the models found by EFA, with the even dataset used for this purpose.

The sample size was sufficiently large (> 200), and the ratio of sample size to model variables should be equal to or higher than 10 (Myers et al., 2011) for CFA. Chi-square (with $p > .05$); comparative fit index (CFI) (recommended > .90); root mean square error of approximation (RMSEA) (recommended < .08); and goodness of fit index (GFI) (recommended > .90) were used in the CFA as indicators to determine the goodness of fit of the obtained models (Hu & Bentler, 1999). The modification index (MI) procedure was used; this identifies misfits by correlating covariance errors among variables within the factors (Brown, 2015), thereby contributing to the choice of model. In addition to the indicators, the Akaike information criterion (AIC), and Bayesian information criterion (BIC) were also used. Several MI procedures were performed by freeing parameters with the highest modification indexes one at a time in sequential order, while observing the improvement in each CFA indicator. A critical value of 3.84 (Brown, 2015) was used as the cutoff point for MI. The open-source software package JASP version 0.14.1 (https://jasp-stats.org) was used for the various EFA and CFA procedures. In order to facilitate interpretation of the obtained constructs, the variables in which time was measured were reversed. That was done for the scores of the completion time of TMT B-A and of Stroop Card 3-2.

Third, Cronbach's alpha (ICC) was used to determine the reliability index for the extracted constructs representing one aspect of reliability: internal consistency among the variables within a construct. A coefficient above .70 was considered to be acceptable (Chadha, 2009).

Finally, the role of demographic factors (e.g., age and education) in the EF tests was analyzed with Pearson product-moment correlations, while the role of sex in the EF tests was analyzed using the student's t-test; p-values below .05 were used to infer significant correlations and significant sex differences.

Results

Properties of the Five EF tests

Descriptive statistics for the EF tests and their variables for the full, odd and even datasets are presented in Table 1.

Construct Validity and Reliability of Executive Function Measurements

Construct validity was examined through EFA and CFA using the two different datasets. To avoid multicollinearity, some variables, including some redundant ones, were excluded from the factor analyses: DS total score, 5PT perseverance errors, pVFT total score, TMT A and TMT B, and Stroop percentage correct. EFA was first conducted using the odd dataset, which indicated that the factor structure was meaningful (KMO = .862, Bartlett's test: χ^2 = 1393.169, df 45, p < .001).

The number of factors were extracted from the parallel analysis. The screen plot identified a point of inflection at three factors between the real and simulated data (see Figure 1). The results revealed two three-factor models that explained slightly more than 31 % of the total variance.

The EFA results, as presented in Table 2, identified three factors. The first of these was dominated by the pVFT, with a minor loading of the 5PT, measuring cognitive fluency elements. The second factor was dominated by the DS, representing WM, with smaller loadings of the TMT (B-A) and the 5PT. The third factor had a moderate loading from the interference factor typical for inhibition of conflicting information and a smaller negative loading on the speed factor.

The factor structure suggested two possible three-factor models since the 5PT unique loaded about equally on two factors. Next, CFA, with its goodness of fit indicators, was used to compare the two three-factor models and those of Miyake's three-factor model (Miyake et al., 2000). This last model has the three DS variables as indicators of working memory, the two Stroop variables representing inhibition, the TMT (B-A), the three pVFT variables, and the number of correct items of the 5PT as indicators of cognitive flexibility. The goodness of fit indicators of the three models are presented in Table 3.

Table 1

Descriptive Statistics (Mean and Standard Deviation) for Each of the EF Tests

	Full Dataset (n = 840)		Odd Dataset (n = 420)		Even Dataset (n = 420)	
Variable	Mean	SD	Mean	SD	Mean	SD
Digit Span						
Digit Span Forward	7.23	2.22	7.24	2.23	7.22	2.21
Digit Span Backward	6.00	2.33	6.00	2.30	6.01	2.36
Digit Span Sequence	7.54	2.94	7.60	2.95	7.49	2.94
Digit Span Total Score	20.77	6.17	20.83	6.04	20.72	6.31
Five Point Test						
Unique Number	24.30	9.60	24.99	9.54	23.62	9.61
Percentage Perseverance Errors	3.19	5.60	3.08	4.70	3.29	6.38
Trail Making Test						
Part A Time	50.01	26.64	49.73	28.12	50.30	25.10
Part B Time	98.80	68.66	98.50	76.68	99.09	59.68
Time B-Time A	47.78	54.44	46.77	60.10	48.76	48.20
Phonemic Verbal Fluency Test						
Letter S Correct Score	12.88	5.32	13.13	5.39	12.64	5.23
Letter K Correct Score	13.70	4.99	14.01	5.03	13.40	4.94
Letter T Correct Score	11.95	4.85	12.20	4.89	11.70	4.81
Phonemic VFT Total Score	38.54	13.57	39.34	13.69	37.73	13.41
Stroop Test						
Card 3 Correct Score	97.25	5.48	97.56	4.66	96.95	6.19
Time Card 3 minus Time Card 2	29.20	20.01	28.39	20.44	30.01	19.56

Figure 1

Number of Factors Extracted from the Parallel Analysis

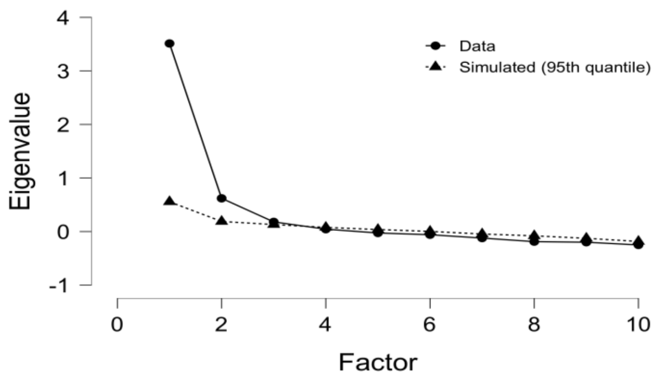


Table 2

Factor Structure Matrix and Loadings of the Various Tests based on EFA Using Oblique Rotation

Variable	Factor 1	Factor 2	Factor 3	Uniqueness
DS Forward		.553		.566
DS Backward		.838		.364
DS Sequence		.556		.559
pVFT Letter S	.854			.278
pVFT Letter K	.841			.305
pVFT Letter T	.804			.308
Stroop Card 3-2 Time			-.356	.850[b)]
Stroop Card 3 Correct			.625	.619
5PT Unique[a)]	.270	.293		.627
TMT B-A Time[a)]		.356		.756

Notes:
[a] = loading factor coefficients were loaded in multiple factors; [b] = high values of uniqueness coefficients

Table 3

Confirmatory Factor Analysis

Model	Goodness of Fit Indicator							
	χ^2	df	p	CFI[a]	RMSEA[b]	GFI[c]	AIC	BIC
First Model (*5PT loads on factor 1*)								
Basic Model	134.346	32	<.001	.929	.087	.941	10554.274	10647.200
1st Modification Indices DS Forward and Sequence	117.251	31	<.001	.940	.081	.948	10539.179	10636.146
2st Modification Indices pVFT Letter K and Letter T	111.053	30	<.001	.944	.080	.951	10534.981	10635.988
Second Model (*5PT loads on factor 2)*								
Basic Model	127.191	32	<.001	.934	.084	.939	10547.119	10640.045
1st Modification Indices 5PT and DS Backward	92.976	31	<.001	.957	.069	.958	10514.904	10611.870
Third Model (Miyake's Model)								
Basic Model	135.615	32	<.001	.928	.088	.934	10555.544	10648.470
1st Modification Indices 5PT and TMT B-A Time	109.182	31	<.001	.946	.077	.952	10531.111	10628.077

Notes:
[a]=expected value>.90; [b]=expected value<.08; [c]=expected value > .90 (Hu & Bentler, 1999);
[+]=expected values obtained

In order to improve the goodness of fit indicators, multiple modification index (MI) procedures were performed by correlating the two covariance errors among variables that were constructed in the same factors for both the three-factor models and Miyake's model. The results, as presented in Table 3, indicate that the goodness of fit indicators progressively improve with each MI.

The χ^2 p-values of the three models were below .01 (significant), but χ^2 is fairly sensitive to the large number of subjects included in the study, so other goodness of fit measures needed to be explored (Alavi et al., 2020). RMSEA, (CFI and GFI) showed that the three basic models already had reasonable to good goodness of fit indicators. The MI procedure used to improve these. This was performed by correlating the two covariance errors among variables that were constructed in the same factors for both the three-factor models Two MIs. were necessary to obtain the expected values for each of the goodness of fit indicators of the first three-factor model. In comparison, the second three-factor model and Miyake's model needed one MI to obtain the expected values for each of the indicators and were therefore preferred.

The comparisons between the three models showed that the second, in which 5PT loaded on the second factor with DS Forward, DS Backward, DS Sequence and TMT B-A, fitted the data best [χ^2 (31,840) = 92.976, $p < .001$, RMSEA = .069, CFI = .957, GFI = .958, AIC = 10514.904, BIC = 10611.870] in comparison to the first and the third models (Miyake's model). The structure of the best fitted three-factor model (Model 2) is presented in Figure 2.

Cronbach's alpha (α) was used to determine the reliability (internal consistency of the items of a construct) coefficient of the three factors by using the full dataset (n=840). The results of the reliability analysis can be seen in Table 4. Two factors showed an acceptable to good ICC coefficient: .875 for factor 1 and .759 for factor 2, which were both above .70 (Chadha, 2009), while the reliability coefficient (α) for factor 3, covering only the Stroop test, time to complete card 3-2 and the number of errors, was .351. This was due to the fact that speed and number of errors, the two items of the Stroop test included in the factor analysis, are inversely related and measure different aspects (speed versus accuracy); therefore, the internal consistency of factor 3 is low. An earlier study based on 480 subjects also showed that the test-retest reliability of the Indonesian Stroop test was satisfactory and close to .70 (.68 for the time difference between Card 3 and Card 2). Another reliability measure, the standard error of measurement (1.05) (Geerinck et al., 2019), showed a satisfactory value, while the ICCs for Card 3, the time difference (3 minus 2), and score Card 3 ranged between .62 and .91 (Wahyuningrum et al., 2022).

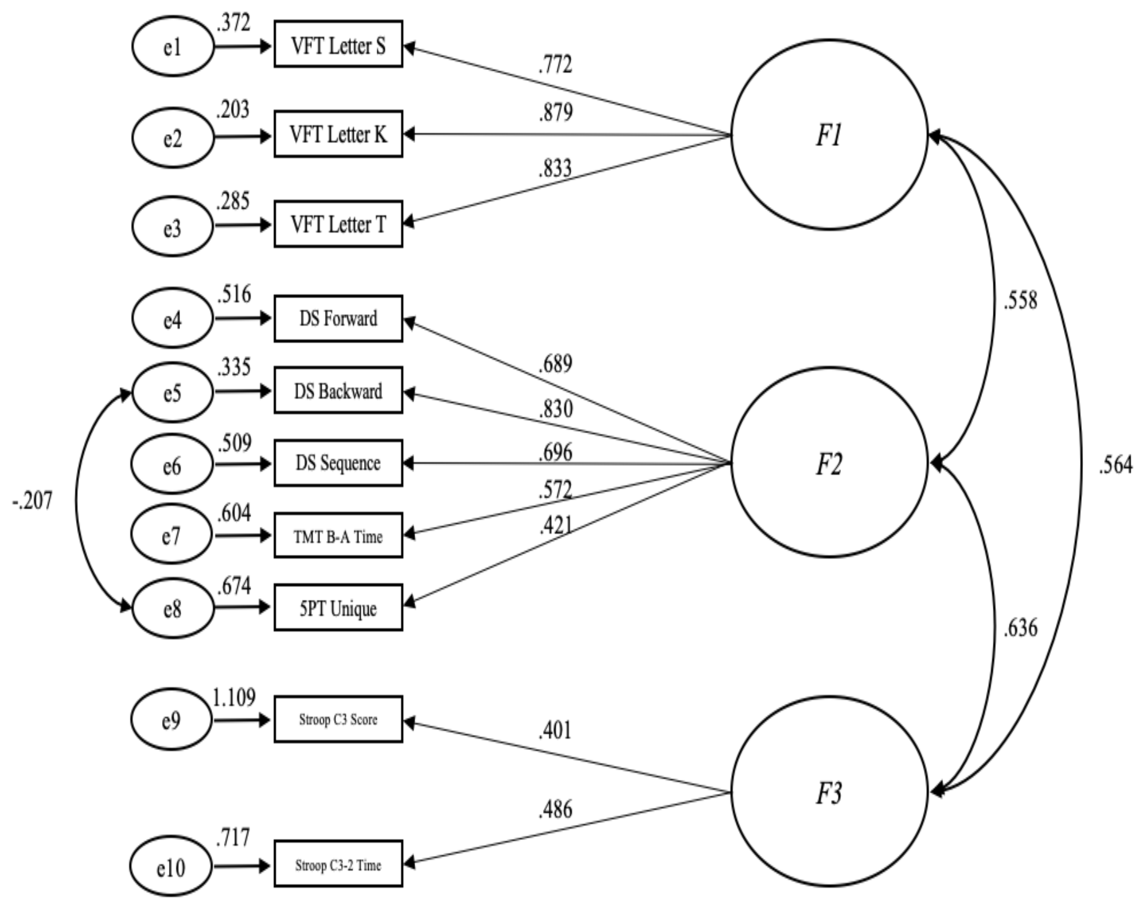
Demographic Analysis

The outcomes of the three bivariate correlation analyses showed correlation coefficients between the demographic factors of age, education and sex on the one hand, and on the other hand the EF variables, as reported in Table 5. The age factor showed moderate but significant negative correlations, which was the case for all five EF tests, apart from the perseverance variable errors of the 5PT. The older people become, the lower their performance. The difference scores (TMT B-A and Stroop 3-2) correlated positively with age, showing that the difference becomes greater with advancing age. Overall, the results indicate age-dependent declining performance in the EF measurements.

Education showed positive correlations with the EF variables, an exception again being the perseverance score of the 5PT test. A longer period of education was accompanied by better scores on all the EF tasks. Finally, sex differences were found in all three TMT variables, the DS backward, and the 5FP unique number. In TMT, males were faster than females (TMT A; Male 46.11, Female: 52.44; TMT B: Male: 88.26, Female: 104.99; B-A Male: 42.72 Female: 52.55). Similarly, at 5PT the scores of males were higher than those of females (Male: 25.2; Female: 23.8), and males also outperformed females in DS Backward (Male: 6.22 and Female: 5.87).

Figure 2

Best Fitting Model of the Cognitive Structure of Executive Function



Note:
χ^2(31,840) = 92.976, p <. 001
RMSEA = .069, CFI = .957, GFI = .958, AIC = 10514.904, BIC = 10611.870.
F1: cognitive fluency, F2: working memory, F3: inhibition

Table 4

Reliability Coefficient of EF Measurement

Factors	Cronbach's α
Factor 1	.875
pVFT Letter S	
pVFT Letter K	
pVFT Letter T	
Factor 2	.759
DS Sequence	
DS Backward	
DS Forward	
TMT B-A	
5PT Test	
Factor 3	.351
Stroop Card 3-2 Time	
Stroop Card 3 Correct	

Table 5

Bivariate Analyses (Correlation Coefficients) of the Demographic Data (Age and Education) and EF Measurements for the whole Sample, together with the Outcomes of the t-test for Sex Differences

Variable	r_{Age}	p_{Age}	r_{Edu}	p_{edu}	t_{Sex}	p_{Sex}
Digit Span (DS)						
Digit Span Forward	-.310	.000[a]	.335	.000[a]	-1.807	.071
Digit Span Backward	-.287	.000[a]	.291	.000[a]	-2.102	.036[b, c]
Digit Span Sequence	-.363	.000[a]	.346	.000[a]	-1.052	.293
Digit Span Total Score	-.393	.000[a]	.395	.000[a]	-1.945	.052
Five Point Test (5PT)						
Unique Design Number	-.394	.000[a]	.362	.000[a]	-2.020	.044[b, c]
Perseveration	.064	.063	-.004	.907	1.303	.193
Trail Making Test (TMT)						
Part A Time	.387	.000[a]	-.342	.000[a]	3.372	.001[a, d]
Part B Time	.407	.000[a]	-.453	.000[a]	3.337	.001[a, d]
B-A Time	.324	.000[a]	-.404	.000[a]	-2.553	.011[b, d]
Phonemic Verbal Fluency Test (pVFT)						
Letter S Correct Score	-.221	.000[a]	.349	.000[a]	-1.089	.276
Letter K Correct Score	-.212	.000[a]	.396	.000[a]	-.229	.765
Letter T Correct Score	-.229	.000[a]	.362	.000[a]	-1.808	.071
VFT Total Score	-.255	.000[a]	.425	.000[a]	-1.183	.237
Stroop Test						
Card 3 Correct Score	-.195	.000[a]	.219	.000[a]	1.087	.277
Card 3 minus 2 Time	.255	.000[a]	-.229	.000[a]	.173	.083

Note: [a]p<.01; [b]<.05; [c]=male > female; [d] =male < female

Discussion

The first research aim was to validate five EF tests (DS, pVFT, TMT, 5PT, and Stroop test) for its use in the Indonesian population. These tests are often used internationally and were included in the INTB and some adapted (pVFT and the Stroop test) into Indonesian versions by the Indonesian Neuropsychological Consortium. Previous studies show that the five tests measure different aspects of EF (Testa et al., 2012; Tucha et al., 2012). Our EFA showed meaningful three factor models with a total variance explaining slightly more than 31%. Even though this is below 50%, the KMO coefficient is more than .50, which means the factor structure developed is meaningful (Field, 2013). The goodness of fit indicators (Chi-square, RMSEA, CFI, GFI, AIC and BIC) of the models were more than acceptable (Hu & Bentler, 1999).

The meaningfulness of the solutions mainly concerns how well the factors can be interpreted. The first factor is primarily based on the results of the pVFT, with the scores of the pVFT letters S, K and T all loading highly on this factor. This factor is referred to as 'verbal flexibility'. The reliability of the first construct was also deemed to be acceptable. Cognitive flexibility, one of the three primary domains of EF distinguished by scholars such as Miyake et al. (Miyake et al., 2000), can be assessed through the pVFT, as participants are required to search for words from various semantic categories and have to make switches (Troyer et al., 1997). Additionally, the pVFT necessitates access to one's lexicon, lexical knowledge, memory monitoring, and inhibitory control (Aita et al., 2019; McCabe et al., 2010). All of these aspects contribute to the uniqueness of this construct among the EF tasks. It is also worth noting that the 5PT loaded > .25 on this factor in Model 1, a task that also measures cognitive flexibility and creativity. As a result, this factor is primarily regarded as cognitive flexibility, one of the three primary domains of EF (Miyake et al., 2000).

The second factor consisted mainly of the outcome of DS Forward, DS Backward and DS Sequence, together with a minor aspect of the 5PT and TMT. The three DS subtests loaded on the same factor and showed that the construct also had an acceptable reliability coefficient. This factor is referred to as 'working memory'. Hantoro et al. (2019) also found that a computerized DS Backward showed a high-reliability coefficient in an Indonesian sample. The Central Executive, a key element in the working memory model of Baddeley and Hitch (1974), plays a significant role in DS performance (Bourrier et al., 2018). DS Forward is also indicative of the size or capacity of the short-term auditory memory buffer. DS Backward and DS Sequence are cognitively considered to be more complex than DS Forward. DS Backward and DS Sequence measure individuals' ability to manipulate information in their working memory.

Therefore, the ability to perform well on the DS requires attention to the presented digits, short-term memory, and working memory, including the ability to manipulate and update currently processed information. A memory component and the updating of information are also necessary for the TMT (B-A); the task requires switching from letters to digits and memorizing the previous letter and digit, while the 5PT requires memorizing the previously-made patterns. This second factor is close to Miyake's working memory domain of EF (Miyake et al., 2000), although our factor analysis also shows elements of cognitive flexibility: the difference between TMT B-A requires scanning and switching while controlling for visual-motor speed. Previous research (Tucha et al., 2012) found significant and moderate correlation of 5PT with the TMT test parts A and B, suggesting that both tests measure a partly shared construct. Matias-Guiu et al. (2022) also argue that fluency can be examined by using non-verbal content. That was one of the reasons why design fluency tasks have been developed, one of which is the 5PT.

The third factor, 'inhibition', consisted of the correct score of Stroop Card 3 (Word Color

Naming) and the difference in time between Stroop Cards 3 and 2. This factor showed a low internal consistency coefficient, indicating that the two dependent variables of the Stroop test, one reflecting speed, the other accuracy (number of correct items), measure distinct cognitive aspects The simple fact that the Stroop test has different response scales (accuracy and time) might also be a reason for the low ICC value. Finally, the high 'uniqueness' value of these two measures also supports the uniqueness of both speed and accuracy. To be more precise, the uniqueness scores of the Stroop Card 3 - Card 2 are higher than all the other variables, indicating that the internal consistency coefficient (ICC) as a reliability measure is not suited for the construct measured by factor 3. A better indicator of the reliability of the Stroop test could be the test-retest correlation of the various dependent variables, or the standard error of measurement of these. Research on the Stroop test indicates a good test-retest reliability for all its different versions, including ours (Widhianingtanti et al., 2022). In addition, different strategies in performing the task, a trade-off between speed and accuracy, are common in many different cognitive tasks; for example, in the Flanker, the Simon task, go/no-go, and many kinds of judgment tasks, such as lexical, semantic, alphabet and numeric decision ones (Vandierendonck, 2021). This could be the cause of the low reliability of factor 3.

In theoretical terms, inhibitory control is the third main pillar of the three-domain theory of EF (Diamond, 2013; Logue & Gould, 2014; Miyake et al., 2000). Scarpina and Tagini (2017) and G. P. Strauss et al. (2005) confirmed that the Stroop test provides information about inhibitory control, alongside selective attention, cognitive flexibility, and processing speed. The difference between the time to accomplish the congruent trial (Card 2) and that to accomplish the incongruent trial (Card 3) is utilized to measure the sensitivity for interference induced by conflicting information (Kurniawan & Kusrohmaniah, 2018; MacLeod, 2010). Overall, factor 3, represented the inhibition domain of EF (accuracy), as well as the speed of processing, no other tests of variables loaded on this factor.

As stated in the introduction, many experts (e.g., Miyake et al.) refer to the three main domains of EF as working memory, inhibition and cognitive flexibility. Our results approximate very closely to Miyake's three-correlated factor model, showing separable but correlated sub-EF functions. The results of this study also confirm the existence of these three fairly similar factors in the Indonesian context: cognitive flexibility was mainly represented by the pVFT; working memory mainly by the DS; and inhibition by the Stroop test. Based on the EFA and CFA, it can be concluded that the outcomes demonstrate that the three core constructs often mentioned in the literature are represented in these five EF tests, which contributes to the validity of the EF tests in the INTB. Other researchers and clinicians in Indonesia can employ the tests or extend the number of EF tests and re-evaluate the factorial structure, although the five tests already capture the three main domains of EF. Practicing psychologists can also use the five subtests of the INTB to measure EF, since now, besides the validity of the tests, normative data on the Javanese population have been published (Wahyuningrum et al., 2022).

The third research question in the study concerned exploration of the demographic factors related to EF tests. Our results indicate similar findings to previous research, in that age and education had effects on the pVFT, DS, 5PT, TMT and Stroop test (Goebel et al., 2009; Matias-Guiu et al., 2022). Age, education and sex effects on the TMT A, TMT B and TMT (B-A) were recently reported for the adult Javanese population (Widhianingtanti et al., 2022). Previous studies found that age and education explained approximately 20–25% of the total variance of the scores of the TMT (Khalil, 2010; Tucha et al., 2012).

Research by Vicente et al. (2021) and Widhianingtanti et al. (2022) also showed that Stroop's interference score was negatively influenced by age, which means that the older the individual, the higher the amount of interference. Age effects on EF tests can be explained by physiological aspects, including neuro-anatomical changes associated with aging, which are mostly seen in the prefrontal cortex. McCabe et al. (2010) explain that age has a greater effect on changes in the frontal cortex compared to many other brain areas, although it should be noted that not all frontal areas decline at similar rates with advancing adult age. Overall, it can be safely assumed that changes in the prefrontal cortex accompanying normal aging significantly affect scores in EF tests.

Sex differences were found in the TMT, DS Backward, and 5FP test unique number of correct items. The results showed that men are better at figural tasks, and at tasks requiring attention, specifically in relation to transformation and manipulation of information while simultaneously storing information. This finding is similar to those of Jorm et al. (2004), Dadin et al. (2009) and Choi et al. (2014). The advantage of males may be related to the fact that they generally use serial type strategies for coding verbal material, while females use semantic type strategies. Social role influences should be investigated more in the future. The same-sex differences for the TMT were also found in a smaller Javanese sample (Widhianingtanti et al., 2022), which can be generalized to a mixed Javanese and non-Javanese population. Vicente et al. (2021) state that for Stroop Word-color scores, an interaction between age and sex was found, in which men outperform women up to the age of 55, but after the age of 65, women perform better than men.

Overall, the sensitivity of age, education, and sex effects, in line with previous studies, demonstrates that the adapted EF tests are sensitive to these demographics, which also contributes to the validity of the five EF tests.

Strengths of the Study

The study included 840 participants from different parts of Java, Bali, East Kalimantan and South Sulawesi, which represent large Indonesian islands with sizable populations. The sample size is adequate for the generalization of results, as well as for preliminary norming studies of the Indonesian population. To date, there has been no agreement on EF measurements in Indonesia. This study can therefore act as a preliminary reference for psychology researchers in understanding EF measurement, and for psychologists' use of the five validated EF tests in Indonesia. The series of tests is also practical because it only takes approximately 35 minutes to administer them.

Limitations and Recommendations for Future Studies

Despite the significant implications of our findings, some limitations should be kept in mind when interpreting the results. In the study, most of the participants were from urbanized areas, many were relatively well-educated, and in the younger age group, so older and less educated people were less well represented. Future studies should expand the age range; the proportion of less educated participants; and recruit people from the rural population in order to capture the broad range of functioning that is present in the Indonesian community. Considering that the participants in our study were a nonclinical sample, future studies could attempt to replicate and extend our results to different clinical populations, establishing their clinical usefulness in the Indonesian context. Various categories of neuropsychological patients, neuropsychiatric patients, and patients with, among other conditions, diabetes, cardiovascular problems, autoimmune diseases, and kidney problems, could be used. Interaction effects between age, education, and sex on EF should be researched more deeply, as well as putative ethnic effects. A final limitation is that the EFA resulted in a three-factor solution, in which Cronbach's alpha indicated a low internal consistency for the construct based on the two Stroop variables measuring speed and accuracy. This suggests that the construct requires a further inhibition task such as Go-NoGo to increase its internal consistency. This research recommend that future researchers analyze invariant equivalence for various demographic aspects.

Conclusion

This research demonstrated with EFA and CFA that the five EF tests (pVFT, DS, TMT, 5PT and the Stroop test), adapted to the Indonesian context by our consortium, are multifactorial and valid, and that two of the three constructs are reliable. This research found a three-factor structure underlying the five tests, representing cognitive flexibility, working memory and inhibition, in agreement with what has been internationally reported. Moreover, the tests were sensitive to age, education and sex. The three-factor structure, and the age, education and sex effects contribute to the validity of the five EF tests. Considering the demographic effects, normative scores for the EF tests need to be adapted for the related factors.[]

Acknowledgment

The authors are grateful for the contribution of the tedious work of the research assistants who collected the data, and the other members of the Indonesian Consortium of Neuropsychologists, who were instrumental in the completion of the work.

Author Contribution Statement

Ni Made Swasti Wulanyani: Conceptualization; Resources; Writing Original Draft; Writing; Review & Editing. **Lucia Trisni Widhianingtanti:** Data Curation; Methodology; Resources; Writing Original Draft. **Aria Saloka Immanuel:** Data Curation; Formal Analysis; Writing Original Draft. **Aireen Rhammy Kinara Aisyah:** Writing Original Draft; Writing; Review & Editing. **Marc P. H. Hendriks:** Validation; Clinical Expertise. **Yohana Ratrin Hestyanti:** Writing Original Draft; Writing, Review & Editing. **Angela Oktavia Suryani:** Formal Analysis; Writing Original Draft; Methodology. **Gilles van Luijtelaar:** Conceptualization; Data Curation; Formal Analysis; Writing Original Draft; Writing, Review & Editing; Supervision.

References

Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2019). Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Applied Neuropsychology: Adult*, *26*(5), 441–451. https://doi.org/10.1080/23279095.2018.1439830

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing*, *76*(9), 2209–2211. https://doi.org/10.1111/jan.14399

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of Learning and Motivation* (pp. 47–89). Academic Press.

Bourrier, S. C., Berman, M. G., & Enns, J. T. (2018). Cognitive strategies and natural environments interact in influencing executive function. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01248

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Press.

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale For Children – Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, *89*(2), 195–224. https://doi.org/10.1111/bjep.12230

Chadha, N. K. (2009). *Applied psychometry*. SAGE Publications India Pvt Ltd. https://doi.org/10.4135/9788132108221

Chan, R., Shum, D., Toulopoulou, T., & Chen, E. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216. https://doi.org/10.1016/j.acn.2007.08.010

Chan, T., Wang, I., & Ybarra, O. (2021). Leading and managing the workplace: The role of executive functions. *Academy of Management Perspectives*, *35*(1), 142–164. https://doi.org/10.5465/amp.2017.0215

Choi, H. J., Lee, D. Y., Seo, E. H., Jo, M. K., Sohn, B. K., Choe, Y. M., Byun, M. S., Kim, J. W., Kim, S. G., Yoon, J. C., Jhoo, J. H., Kim, K. W., & Woo, J. I. (2014). A normative study of the digit span in an educationally diverse elderly population. *Psychiatry Investigation*, *11*(1), 39–43. https://doi.org/10.4306/pi.2014.11.1.39

Dadín, C. O., Salgado, D. R., & Fernández, E. A. (2009). Natural sex hormone cycles and gender differences in memory. *Actas Espanolas de Psiquiatria*, *37*(2), 68–74. http://www.ncbi.nlm.nih.gov/pubmed/19401854

Danielsson, H., Henry, L., Messer, D., & Rönnberg, J. (2012). Strengths and weaknesses in executive functioning in children with intellectual disability. *Research in Developmental Disabilities*, *33*(2), 600–607. https://doi.org/10.1016/j.ridd.2011.11.004

Danielsson, H., Henry, L., Rönnberg, J., & Nilsson, L.-G. (2010). Executive functions in individuals with intellectual disability. *Research in Developmental Disabilities*, *31*(6), 1299–1304. https://doi.org/10.1016/j.ridd.2010.07.012

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology*, *51*(S1), S17–S26. https://doi.org/10.1086/650572

Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports*, *11*(1), 1382. https://doi.org/10.1038/s41598-020-80866-1

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.

Fleming, K. A., Heintzelman, S. J., & Bartholow, B. D. (2016). Specifying associations between conscientiousness and executive functioning: Mental set shifting, not prepotent response inhibition or working memory updating. *Journal of Personality*, *84*(3), 348–360. https://doi.org/10.1111/jopy.12163

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble. *European Journal of Psychological Assessment*, *33*(6), 399–402. https://doi.org/10.1027/1015-5759/a000460

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*(2), 326–340. https://doi.org/10.1037/dev0000075

Geerinck, A., Alekna, V., Beaudart, C., Bautmans, I., Cooper, C., De Souza Orlandi, F., Konstantynowicz, J., Montero-Errasquín, B., Topinková, E., Tsekoura, M., Reginster, J.-Y., & Bruyère, O. (2019). Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: An analysis of subjects from 9 validation studies. *Plos One*, *14*(4), e0216065. https://doi.org/10.1371/journal.pone.0216065

Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacona, M., & Capitani, E. (1996). Trail making test: Normative values from 287 normal adult controls. *The Italian Journal of Neurological Sciences*, *17*(4), 305–309. https://doi.org/10.1007/BF01997792

Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., Paul, L. K., & Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences*, *109*(36), 14681–14686. https://doi.org/10.1073/pnas.1206608109

Goebel, S., Fischer, R., Ferstl, R., & Mehdorn, H. M. (2009). Normative data and psychometric properties for qualitative and quantitative scoring criteria of the Five-point Test. *The Clinical Neuropsychologist*, *23*(4), 675–690. https://doi.org/10.1080/13854040802389185

Hale, J. B., Hoeppner, J.-A. B., & Fiorello, C. A. (2002). Analyzing digit span components for assessment of attention processes. *Journal of Psychoeducational Assessment*, *20*(2), 128–143. https://doi.org/10.1177/073428290202000202

Hantoro, W. D., Setyaningrum, C. T. S., & Dahlan, P. (2019). Reliabilitas dan validitas backward digit span test dalam screening gangguan atensi anak sekolah dasar. *Berkala NeuroSains*, *18*(2), 90–93. https://journal.ugm.ac.id/bns/article/view/54992

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. https://doi.org/10.2117/psysoc.2010.14

Himi, S. A., Bühner, M., & Hilbert, S. (2021). Advancing the understanding of the factor structure of executive functioning. *Journal of Intelligence*, *9*(1), 16. https://doi.org/10.3390/jintelligence9010016

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: The role of executive function. *Child Development*, *77*(6), 1822–1841. https://doi.org/10.1111/j.1467-8624.2006.00976.x

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, *30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, *108*(2), 187–218. https://doi.org/10.1037/a0038557

Jorm, A. F., Anstey, K. J., Christensen, H., & Rodgers, B. (2004). Gender differences in cognitive abilities: The mediating role of health state and health habits. *Intelligence*, *32*(1), 7–23. https://doi.org/10.1016/j.intell.2003.08.001

Khalil, M. S. (2010). Preliminary Arabic normative data of neuropsychological tests: The verbal and design fluency. *Journal of Clinical and Experimental Neuropsychology*, *32*(9), 1028–1035. https://doi.org/10.1080/13803391003672305

Koci, J., & Donaldson, S. I. (2023). *Well-being and success for university students: Applying PERMA+4*. Routledge - Taylor & Francis.

Kurniawan, A., & Kusrohmaniah, S. (2018). Performa inhibitory control dengan induksi Sing-a-Song Stress Test pada dewasa awal. *Jurnal Psikologi*, *45*(1), 40–51. https://doi.org/10.22146/jpsi.25846

Kyriazos, T. A. (2018). Applied psychometrics: Writing-up a factor analysis construct validation study with eExamples. *Psychology*, *9*(11), 2503–2530. https://doi.org/10.4236/psych.2018.911144

Lebon, T. (2014). *Achieve your potential with positive psychology: CBT, mindfulness and practical philosophy for finding lasting happiness*. Hachette UK.

Lezak, M. D. (2004). *Neuropsychological assessment*. Oxford University Press, USA.

Logue, S. F., & Gould, T. J. (2014). The neural and genetic basis of executive function: Attention, cognitive flexibility, and response inhibition. *Pharmacology Biochemistry and Behavior*, *123*, 45–54. https://doi.org/10.1016/j.pbb.2013.08.007

MacLeod, C. M. (2010). Stroop effect. In *The Corsini encyclopedia of psychology* (pp. 1–2). Wiley. https://doi.org/10.1002/9780470479216.corpsy0952

Matias-Guiu, J. A., Pytel, V., Delgado-Álvarez, A., Delgado-Alonso, C., Cortés-Martínez, A., Fernández-Oliveira, A., & Matías-Guiu, J. (2022). The Five-point Test: Normative data for middle-aged and elderly Spaniards. *Applied Neuropsychology: Adult*, *29*(6), 1323–1331. https://doi.org/10.1080/23279095.2021.1873137

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. https://doi.org/10.1037/a0017619

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *21*(2), 0169–0183. https://doi.org/10.1055/s-2000-7563

Morasch, K. C., & Bell, M. A. (2011). The role of inhibitory control in behavioral and physiological expressions of toddler executive function. *Journal of Experimental Child Psychology*, *108*(3), 593–606. https://doi.org/10.1016/j.jecp.2010.07.003

Mvududu, N. H., & Sink, C. A. (2013). Factor analysis in counseling research and practice. *Counseling Outcome Research and Evaluation*, *4*(2), 75–98. https://doi.org/10.1177/2150137813494766

Myers, N. D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in exercise and sport. *Research Quarterly for Exercise and Sport*, *82*(3), 412–423. https://doi.org/10.1080/02701367.2011.10599773

Panerai, S., Tasca, D., Ferri, R., Genitori D'Arrigo, V., & Elia, M. (2014). Executive functions and adaptive behaviour in autism spectrum disorders with and without intellectual disability. *Psychiatry Journal*, 1–11. https://doi.org/10.1155/2014/941809

Pesau, H. G., & Luijtelaar, G. Van. (2021). Equivalence of traditional and internet-delivered testing of word fluency tasks. *Jurnal Psikologi*, *20*(1), 35–49. https://doi.org/10.14710/jp.20.1.35-49

Peterson, G. (2006). Cultural theory and emotions. In *Handbook of the Sociology of Emotions* (pp. 114–134). Springer US. https://doi.org/10.1007/978-0-387-30715-2_6

Pires, L., Moura, O., Guerrini, C., Buekenhout, I., Simões, M. R., & Leitão, J. (2019). Confirmatory factor analysis of neurocognitive measures in healthy young adults: The relation of executive functions with other neurocognitive functions. *Archives of Clinical Neuropsychology*, *34*(3), 350–365. https://doi.org/10.1093/arclin/acy040

Ramchandran, K., Colbert, A. E., Brown, K. G., Denburg, N. L., & Tranel, D. (2016). Exploring the neuropsychological antecedents of transformational leadership: The role of executive function. *Adaptive Human Behavior and Physiology*, *2*(4), 325–343. https://doi.org/10.1007/s40750-016-0051-y

Ren, X., Schweizer, K., Wang, T., Chu, P., & Gong, Q. (2017). On the relationship between executive functions of working memory and components derived from fluid intelligence measures. *Acta Psychologica*, *180*, 79–87. https://doi.org/10.1016/j.actpsy.2017.09.002

Scarpina, F., & Tagini, S. (2017). The Stroop Color and Word Test. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00557

Sherman, E., Tan, J. E., & Hrabok, M. (2022). *A compendium of neuropsychological tests: Fundamentals of neuropsychological assessment and test reviews for clinical practice* (4th ed.). Oxford University Press.

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00328

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.

Strauss, G. P., Allen, D. N., Jorgensen, M. L., & Cramer, S. L. (2005). Test-retest reliability of Standard and Emotional Stroop Tasks. *Assessment*, *12*(3), 330–337. https://doi.org/10.1177/1073191105276375

Suwartono, C., Halim, M. S., Hidajat, L. L., Hendriks, M. P. H., & Kessels, R. P. C. (2014). Development and reliability of the Indonesian Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). *Psychology*, *5*(14), 1611–1619. https://doi.org/10.4236/psych.2014.514171

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Testa, R., Bennett, P., & Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Archives of Clinical Neuropsychology*, *27*(2), 213–224. https://doi.org/10.1093/arclin/acr112

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Tucha, L., Aschenbrenner, S., Koerts, J., & Lange, K. W. (2012). The Five-point Test: Reliability, validity and normative data for children and adults. *Plos One*, *7*(9), e46080. https://doi.org/10.1371/journal.pone.0046080

Vandierendonck, A. (2021). On the utility of integrated speed-accuracy measures when speed-accuracy trade-off is present. *Journal of Cognition*, *4*(1), 22. https://doi.org/10.5334/joc.154

Vicente, S. G., Rivera, D., Barbosa, F., Gaspar, N., Dores, A. R., Mascialino, G., & Arango-Lasprilla, J. C. (2021). Normative data for tests of attention and executive functions in a sample of European Portuguese adult population. *Aging, Neuropsychology, and Cognition*, *28*(3), 418–437. https://doi.org/10.1080/13825585.2020.1781768

Wahyuningrum, S. E., Sulastri, A., Hendriks, M. P. H., Indonesian NP Consortium, & van Luijtelaar, G. (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. https://doi.org/10.5604/01.3001.0016.1339

Wahyuningrum, S. E., van Luijtelaar, G., & Sulastri, A. (2023). An online platform and a dynamic database for neuropsychological assessment in Indonesia. *Applied Neuropsychology: Adult*, *30*(3), 330–339. https://doi.org/10.1080/23279095.2021.1943397

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Widhianingtanti, L. T., Luijtelaar, G. Van, Suryani, A. O., Hestyanti, Y. R., & Sulastri, A. (2022). Indonesian Trail Making Test: Analysis of psychometric properties, effects of demographic variables, and norms for Javanese adults. *Jurnal Psikologi*, *49*(2), 104–122. https://doi.org/10.22146/jpsi.68953

Zelazo, P. D. (2020). Executive function and psychopathology: A neurodevelopmental perspective. *Annual Review of Clinical Psychology*, *16*(1), 431–454. https://doi.org/10.1146/annurev-clinpsy-072319-024242

This page has been intentionally left blank.