Jurnal Psikologi Ulayat
Indonesian Journal of Indigenous Psychology / IJIP

Jurnal Psikologi Ulayat: Indonesian Journal of Indigenous Psychology (2026), 13(1), XX–XX
DOI: 10.24854/jpu1045

e-ISSN: 2580-1228
p-ISSN: 2088-4230

CONSTRUCT VALIDITY AND TEST-RETEST RELIABILITY OF THE SUPERVISED ONLINE-ADMINISTERED INDONESIAN EMOTION WORD FLUENCY TEST

Heni Gerda Pesau[1] & Gilles van Luijtelaar[2]

[1]*Department of Psychology, Faculty of Psychology, Atma Jaya University, Makassar, Indonesia*
[2]*Donders Centre for Cognition, Radboud University, Nijmegen, The Netherlands*
[2]*Bachelor Program, Faculty of Psychology, Maranatha Christian University, Bandung, Indonesia*

Correspondence: *heni_gerda@lecturer.uajm.ac.id*

VALIDITAS KONSTRUK DAN RELIABILITAS TES-ULANG DARI TES KELANCARAN KATA EMOSI BAHASA INDONESIA YANG DILAKSANAKAN SECARA DARING DENGAN SUPERVISI

Abstrak

Tes kefasihan kata emosi (EWFT), yang mengukur kemampuan menghasilkan kata-kata emosional, dapat mengukur aspek fleksibilitas verbal yang berbeda dan unik seperti tes kefasihan verbal fonemik (PVFT) dan semantik (SVFT) yang umum digunakan. Hal ini dihipotesiskan dan diuji melalui Analisis Komponen Utama (PCA) terhadap 150 subjek sehat dan tes reliabilitas ulang EWFT dua minggu setelahnya pada empat puluh subjek. Pengetesan dilakukan melalui internet dengan pengawasan dari penguji tes. Analisis korelasi menunjukkan bahwa EWFT berkorelasi moderat dengan PVFT dan SVFT. PCA menunjukkan bahwa model tiga faktor menjelaskan 68.5% dari total varian dan bahwa masing-masing dari tiga tes memuat faktor yang berbeda. Hal ini menunjukkan bahwa masing-masing dari tiga tes kelancaran bahasa ini mengukur konstruksi dasar yang berbeda. Studi *test-retest* menunjukkan reliabilitas pada taraf moderat atau sedang tanpa efek pengulangan/latihan yang jelas. Dapat disimpulkan bahwa EWFT versi Indonesia yang diadministrasikan secara daring melalui pengawasan secara potensial dapat mengukur keterampilan unik, dalam mendeteksi kemampuan pemrosesan emosi. Namun, hasil yang didapatkan masih membutuhkan validasi pada populasi klinis.

Manuscript type: Original Research

Article history:
Received 10 May 2024
Received in revised form 5 May 2025
Accepted 24 February 2026
Available online 18 May 2026

Kata Kunci:
emosi
fonemik
kefasihan verbal
neuropsikologi
semantik

Abstract

The emotion word fluency test (EWFT), measuring the ability to produce emotional words, may measure a different cognitive construct of verbal flexibility than the commonly used phonemic (PVFT) and semantic (SVFT) verbal fluency tests. This was hypothesized and tested through Principal Component Analyses (PCA) of the test scores of 150 healthy subjects. The test-retest reliability was additionally examined after two weeks in forty subjects. Data was collected through supervised internet-delivered testing. Correlation analyses for construct validity showed that the EWFT correlated only moderately with PVFT and SVFT. PCA showed that a three-factor model explained 68.5% of the total variance. The subscales of each fluency test loaded primarily on separate factors, indicating that the three tests measure different underlying constructs. The test-retest revealed moderate reliability without a clear repetition/practice effect. It can be concluded that the supervised Indonesian online-administered EWFT measures emotion processing abilities. However, it awaits clinical validation in various clinical populations.

Keywords: cognition; emotion; neuropsychology; phonemic; semantic; verbal fluency

Impacts and Implications in the Indigenous Context

The EWFT, a cognitive-semantic verbal fluency test, has been developed and adapted for Indonesia and contributes to the availability of neuropsychological tests, next to those covering the domains of learning and memory, attention, executive functions, and language. The EWFT is specifically used to measure the ability to produce emotion words and is intended for clinical purposes. Previous studies have shown that there is a possible influence of culture on an individual's emotions, so research related to the development of EWFT in Indonesia is expected to be able to represent the uniqueness of Indonesian culture in an emotional-cognitive context. Next, the huge variety of languages and ethnic groups is a challenge for a fair assessment. Here we present some psychometric analyses for this test for fluent Bahasa Indonesia-speaking persons.

Handling Editor: Ratih Aruum Listiyandini, Universitas YARSI, Jakarta, Indonesia

INTRODUCTION

Emotion is one of the important functions that can be measured in a neuropsychological assessment. Emotions are known to influence cognition, and cognition influences emotions. The development and adaptation of neuropsychological tests in Indonesia is currently carried out (e.g. Wahyuningrum et al., 2022), but this is not the case for how and in which cognitive factors are influenced by emotions. Therefore, it is necessary to develop tools at the interface between cognition and emotions, one of which is the emotion word fluency test (EWFT). EWFT is an instrument used to measure the respondent's ability to produce emotional words within a limited interval (Abeare et al., 2017). We introduced the test in Indonesia, determined some preliminary psychometric properties, and proposed a scoring system for positive and negative emotion words (Pesau et al., 2024). The EWFT may be useful in the assessment of affective language production in patient populations since the task of generating emotion words in EWFT demands the connection between the emotion lexicon and the concept of emotions and emotion categories, and the integration of semantic and emotional networks (Wauters & Marquardt, 2018).

This might be different in and between Indonesia's ethnic groups, considering that differences in emotional intelligence competencies were found across Minangkabau, Javanese, and Dajak persons (Zahrasari et al., 2018). Shameful and happy experiences occurred in different situations in Indonesia’s ethnic groups (Djiwatampu, 2009). Next, different preferences for negative and positive emotions between Javanese and non-Javanese people were reported (Masturah, 2018), as differences in the concept of angry emotions among Javanese, Sundanese, Madurese, Bugis-Makassar, Batak, and Minang ethnic groups (Zuhdi & Nuqul, 2022). Other authors also proposed that the emotions felt by an individual in each situation are shaped by culture, just as the expression of emotions is molded by culture-specific display rules (Röttger-Rössler & Markowitsch, 2009). Therefore, the integration of semantic and emotional networks might be different in different cultures but also in various patient populations (Shao et al., 2014; Abeare et al., 2017).

Another reason why the EWFT is interesting in the Indonesian context is that it is well accepted that although there exists a high degree of similarity in the emotion concepts of different cultures and languages, labeling emotions is often language-specific, which implies that it is sometimes difficult to translate them into a single word or a group of words in another distinct language. Next, different languages recognize different emotions, and there are emotion words lacking an equivalent in other languages (Frijda et al., 1995; Altarriba, 2003). A comparative study between Indonesian and Dutch students on the cognitive structure of emotional terms found that many

emotional words showed an equivalent factor structure; on the other hand, the social emotions “shame” and “guilt” were closer to "fear" and somewhat further away from "anger" in Indonesia than in the Netherlands (Fontaine et al., 2002). The results of these previous studies show that emotional terms have a cognitive structure, and even if equivalent, emotional terms are diverse and influenced by cultural and linguistic factors.

EWFT is a new assessment tool; it belongs to the category of semantic verbal fluency tests (VFT), introduced by Abeare et al. (2017). VFT have been widely used both for research and in the clinic to measure verbal skills such as word production ability and executive control in patients with cerebrovascular damage (Shao et al., 2014). There are two different verbal fluency tests: phonemic (letter) verbal fluency (PVFT) and semantic (category) verbal fluency (SVFT) (Lezak et al., 2012; Abeare et al., 2017). Previous studies indicate that the PVFT and SVFT assess partially distinct constructs. Performance on the PVFT relies more heavily on executive control processes, particularly the ability to retrieve words based on phonemic cues, inhibit semantically or associatively related but task-irrelevant responses, and generate novel retrieval strategies. In contrast, the SVFT requires the retrieval of words from a single semantic category within semantic memory and is more strongly dependent on lexical access efficiency and vocabulary knowledge (Shao et al., 2014; Li et al., 2017; Aita et al., 2018). The EWFT involves functions that partly overlap with these two verbal word production tests, and therefore, the scores of the EWFT might show a positive correlation with the PVFT and SVFT. But since the EWFT involves the production of words from a semantic category, it should be considered as an SVFT, and the correlations between the EWFT and the SVFT might be higher than between the EWFT and PVFT.

These three word fluency tests involve different parts of the brain while retrieving and producing words. Phonological word search in the PVFT was associated with decreased beta and increased theta activity in the left frontal cortex, whereas word search in SVFT specifically involved the temporal (Mousavi et al., 2020) and occipital cortex (Li et al., 2017). fMRI studies showed that word production in the EWFT involves the prefrontal area, amygdala, and hippocampus (Bevilaqua et al., 2016), dorsolateral prefrontal cortex (Auerbach et al., 2015), parietal and insular cortex (Schreiter et al., 2019).

In a country like Indonesia, with its large distances, many people living in remote areas, and a huge number of islands, testing people online via the internet may have clear advantages, such as cost efficiency, accessibility, and convenience, presenting a promising alternative to traditional assessments (Pesau & van Luijtelaar, 2021; Pesau et al., 2025). However, the psychometric properties of online tests need to be established. Here, we will evaluate the convergent and discriminant validity

and test-retest reliability of the online version of the EWFT (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Earlier studies revealed this in offline studies, also in Indonesia (Camodeca et al., 2021; Pesau et al., 2024; Hegefeld et al., 2023)

Based on differences in the required cognitive abilities, the neural substrates involved, and the international literature concerning the underlying different cognitive constructs assessed by various word fluency tests, it is hypothesized that the EWFT is moderately related to two other word fluency measures. Moderate but statistically significant positive correlations between EWFT scores and scores on the two other measures of verbal fluency (semantic and phonemic tasks) would provide evidence of convergent validity. These two fluency tasks are known to engage cognitive control, executive functioning, inhibition of irrelevant responses, and lexical–semantic access, and this is shared with the EWFT.

Next, it is expected that the EWFT assesses a construct that is distinct from those measured by the semantic and phonemic word fluency tasks. This should demonstrate the uniqueness of the EWFT—particularly with respect to its emotional component. This distinction will be examined using Principal Component Analysis (PCA). The identification of separable components would provide evidence of discriminant validity and support the construct validity of the EWFT as well.

More precisely, it is expected that the factor loadings of the three phonemic, three semantic, and two EWFT variables will reveal that the three word fluency tasks measure three different constructs and that there are no or few cross-loadings (typically > .30 or .40) on two or more factors instead of loading primarily on one. We used PCA because we wanted to explain the highest amount of variance in our data set with its eight variables without any assumptions on the number of latent variables. This data reduction method and analysis, together with the correlation part of our study, will give some evidence for and insight into the construct validity of the EWFT when administered through supervised-internet testing.

The test-retest reliability for the online EWFT needs to be established as well. Earlier, the reliability of the off-line English EWFT version was investigated in a young adult healthy population (Abeare et al.,2017). We have evaluated the test-retest reliability and the reliability of the scoring system in the Indonesian context for the offline face-to-face version in a similar young population (Pesau et al., 2023). Based on these explanations, the study aims to evaluate psychometric properties of the online version of the EWFT, its convergent and discriminant validity and the test-retest reliability.

METHODS

Participants

Research participants were between 17 and 34 years old with minimal senior high school education or a bachelor's degree (S1). They were relatively young, given the necessity of assessing them online via video calls. It was anticipated that younger individuals would be more familiar and proficient with the required devices, software, and computer systems (Olson et al., 2011; Capizzi et al., 2021). In addition, participants were required to be fluent in Bahasa Indonesia, as the tests were administered in that language. Exclusion criteria comprised a current or prior neurological or psychiatric disorder (e.g., traumatic brain injury, cerebrovascular accidents, epilepsy, Parkinson's disease, depression, or other psychiatric conditions), a history of alcoholism or illicit drug use, and hearing or visual impairments, as determined by a medical history questionnaire administered through an online form.

Participants were recruited through a Google form link distributed by the research assistants via social media; those who were willing to take part in the research registered themselves online and provided online informed consent. They were contacted by phone to arrange a time for data collection via a video internet meeting. The informed consent contained an explanation of the research to be carried out, guarantees of participant confidentiality (identity and any information provided by participants is only used for research purposes), benefits and potential risks, voluntary and non-coercive participation, and that their answers during data collection were recorded.

Procedure

The testers were previously trained and provided test instructions. Before the test started, participants were asked to prepare a laptop/gadget with a good camera, microphone, and loudspeaker, and install a working video meeting application, to ensure that they had a good internet connection and a quiet place at home where they were not disturbed. This allowed us to do the assessments online via video meetings, with active cameras so the tester could monitor the participants throughout the data collection. The assessments started after checking the quality of the video connection and the position of the camera. This procedure was developed, implemented, evaluated earlier, and equivalence was found between traditional (face-to-face) and online testing for the EWFT (Pesau & van Luijtelaar, 2021). It allows participants to take the test anywhere as long as the test settings are appropriate (Pesau et al., 2025). The order of administration of the three tests was controlled by counterbalancing, as can be seen in Table 1.

Data collection was carried out in two batches: the first involved 150 participants, and this data was used for correlation and factor analyses. The data of forty participants was used for the test-retest study and collected in the second batch, two weeks later. Previous research recommends that intervals of around two weeks showed the highest reliability coefficient, while longer time intervals may indicate changes in participant condition, whereas very short intervals may increase the likelihood that participants will still remember their initial responses (Wyse, 2021; Poder et al., 2022). They were selected randomly from the first group of participants. Before joining the research, the participants were fully informed about the goal of the study, their privacy was respected, their data were anonymously stored, that they could withdraw from the study any time, and their data could be used for scientific and educational purposes. They all declared to accept these conditions.

Instruments

Phonemic and Semantic Verbal Fluency. In the phonemic verbal fluency test (PVFT), participants were asked to generate as many words as possible in one minute (Shao et al., 2014), starting with the letter S, next with the letter K, and next with the letter T. Participants were instructed to exclude trade names, brand names, and personal names. The number of correctly generated words for each of the three categories was the score for the three subscales of the PVFT (Hendrawan & Hatta, 2010; Pesau et al., 2023). For the semantic word fluency test (SVFT), the words to be generated must fall into a semantic category; we used the more often used categories of 'animals', 'fruits', and 'furniture'; the time allotted to each semantic category was one minute. The number of correct words per category was used as the score for the three semantic subtests (Villalobos et al., 2023).

Emotion Word Fluency Test. Participants were asked to name as many words related to emotion as possible in one minute (Abeare et al., 2017). The number of positive and negative emotion words (PE and NE, respectively) was the dependent variable. Emotion words were classified as having either a positive (PE) or negative (NE) annotation (Pesau et al., 2023).

After the instructions were read, a one-minute countdown began. All words produced were recorded and then scored (correct or not) by HPG or by a trained research assistant. Only the number of correct words was used for the statistical analyses.

Table 1.
Test Order of Emotion and Verbal Fluency Tests

Test Order	Group 1	Group 2	Group 3
1	EWFT	PVFT	SVFT
2	PVFT	SVFT	EWFT
3	SVFT	EWFT	PVFT

Note: EWFT = Emotion Word Fluency Test; PVFT = Phonemic Verbal Fluency Test; SVFT = Semantic Verbal Fluency Test.

Analysis Strategies

The convergent element of construct validity of the EWFT was determined through correlation analysis (Pearson's product-moment correlation), and discriminant validity as part of construct validity by PCA. It was used to identify the factor structure of the combination of the three verbal fluency tests with their eight subscales. It was expected that each of the word fluency tests, even though they are all word fluency tests, would measure a different underlying component/construct. High and unique loadings of the subscales on each of the three factors representing the three verbal fluency tests, together with no or low cross-loadings (significant loading on more than one verbal fluency test) support the construct (discriminant) validity. The reliability of the tests was determined via the test-retest procedure with an interval of two weeks, and Pearson product-moment correlation of the scores of the first and second assessment sessions was used as well as paired sample t-tests to evaluate learning or practice effects. Considering that the results of the EWFT will be compared with those of PVFT and SVFT, it was checked whether the order of test administration affected the performance of the tests. Test order is a putative confounding factor.

RESULTS

Data collection was conducted by online video meetings previously developed and evaluated (Pesau & van Luijtelaar, 2021). The data collection was carried out on 150 respondents who had met the inclusion criteria with demographic characteristics as shown in Table 2. Based on Table 2, the number of male and female respondents is almost equal, and this is also the case in the education levels of senior high school and bachelor. Most of the respondents were between 20 and 25 years.

Table 2.
Demographic Characteristics (N = 150)

	M (SD)	**Group**	***N***	**Percentage (%)**
Age	23.52 (4.01)	17–19	22	14.67
		20–22	49	32.67
		23–25	37	24.67
		26–28	27	18.0
		29–31	6	4.0
		32–34	9	6.0
Education (in years)	13.92 (2.01)	Senior High School	78	52.0
		Bachelor	72	48.0
Sex		Male	72	48.0
		Female	78	52.0

Analysis 1: Validity

The convergent validity of the EWFT as a type of construct-related validity was measured through Pearson's product-moment correlation with the two other fluency tests; the results can be

seen in Table 3. There were significant positive correlations ($p < .01$) between the EWFT on the one side and the three subtests of PVFT (S, K, T), and Total PVFT (correlation coefficients were between .27 and .38) on the other. The EWFT also showed moderate correlations with two of the three categories of the SVFT ($p < .01$) (correlation coefficients were .27 (fruit) and .40 (animals), whereas no significant correlation was found with SVFT Furniture ($p > .05$), and again with Total SVFT (.38).

Significant ($p < .01$) and strong correlations were found between the three subscales of the PVFT (.61–.68) and between these subscales and the total PVFT score (.87–.89). The correlation among the subscales of the SVFT ranged between .25 and .44, the correlation between the subscales of the SVFT and its total score ranged from .68 to .82. The correlation between the two EWFT subscales (PE and NE) was .36. Then, the correlation of total score of the EWFT with the number of positive words were .77 and for the number of negative words was .87.

Table 3.
Correlations Between the Number of Correct EWFT, PVFT, and SVFT Words

Variable	Mean (SD)	PVFT (S)	PVFT (K)	PVFT (T)	Total PVFT	SVFT (A)	SVFT (Fr)	SVFT (Fu)	Total SVFT	EWFT	PE	NE
PVFT (S)	13.06 (5.3)	-	.66**	.61**	.87**	.34**	.26**	.31**	.40**	.27**	.23**	.21**
PVFT (K)	13.98 (5.4)			.68**	.89**	.40**	.31**	.26**	.44**	.38**	.34**	.28**
PVFT (T)	12.96 (5.2)				.87**	.28**	.34**	.31**	.41**	.34**	.21**	.33**
Total PVFT	40.00 (13.9)					.39**	.34**	.33**	.48**	.38**	.30**	.31**
SVFT(A)	19.97 (6.3)						.44**	.25**	.82**	.40**	.29**	.37**
SVFT(Fr)	14.60 (4.0)							.37**	.75**	.27**	.14	.28**
SVFT(Fu)	12.16 (4.7)								.68**	.14	.15	.11
Total SVFT	46.73 (11.4)									.38**	.27**	.35**
EWFT	7.87 (3.56)										.77**	.87**
PE	2.67 (1.83)											.36**
NE	5.26 (2.46)											-

Note: ** = Significant at .001 level; All variables stand for the number of correct words; PVFT = Phonemic Verbal Fluency; SVFT (A) = Semantic Verbal Fluency Animals; SVFT (Fr) = SVFT Fruit; SVFT (Fu) = SVFT Furniture; EWFT = Emotion Word Fluency Test; PE = Positive Emotions; NE = Negative Emotions.

The correlation coefficients indicate that the scores of the three different word fluency tasks are positive, as expected, and significant. However, considering that the correlation coefficients between the EWFT total score and the total scores of the two other fluency tests were not that high (both were .38), it can be concluded that the three tests are only moderately correlated.

Beside based on the correlation, the results of the PCA revealed the underlying factor structure of these three verbal fluency tests. The outcomes of the PCA are presented in Table 4. The preliminary tests regarding the adequacy of the sample showed a KMO value of .785 (> .50) and Barlett's Test of Sphericity of .000 ($p < .05$), Anti-image correlation value between variables of > .05, and a subsequent test showed that communalities were between .59–.80. Both statistics imply that further analysis could be carried out and might reveal interpretable outcomes.

Based on the further analysis, three factors had Eigenvalues of more than 1, and the total amount of variance explained by the three-factor model was 68.5%. Table 4 shows the three factors after Oblimin rotation with Kaiser Normalization; we chose Oblimin rotation since the underlying factors are expected to correlate (all factors represent word fluency tests).

The results clearly showed that each of the three verbal fluency tests has unique and high loadings from a single test only, implying that the three verbal tests measure different components. Only the two variables of the EWFT, the number of positive and negative emotion words loaded high (>.757) on factor 2; the scores of the three subscales of the PVFT (S, K, and T) loaded higher than .83 on factor 1, while the three subscales of the SVFT (Animals .492, Fruit .824, and Furniture .740) loaded on factor 3. There was only a single cross-loading larger than 0.30, and this regarded the semantic subscale Animals, which loaded .469 on the same factor as the two emotion subscales.

In addition, the data as presented in Table 5 showed that the number of correctly produced words was 12 – 14 for the PVFT (S, K, T), 12 – 19 for the SVFT (animals, fruit, furniture), and about 8 for the EWFT. The outcomes of the PCA indicate that each of the three fluency tests measures a different aspect of verbal fluency. In other words, the EWFT has both some discriminant and convergent validity, both aspects of construct validity. Discriminant validity is based on the PCA, and convergent validity is based on the correlation analyses.

Table 4.
Rotated Component Matrix (Converged in 6 Iterations)

	Component/Construct		
	1	**2**	**3**
PVFT (S)	.892	-.055	-.001
PVFT (K)	.855	.130	-.029
PVFT (T)	.832	.007	.063
SVFT (Animal)	.033	.469	.492
SVFT (Fruit)	-.046	.133	.824
SVFT (Furniture)	.158	-.204	.740
Positive emotion words (PE)	.124	.757	-.136
Negative emotion words (NE)	-.009	.782	.096

Note: Extraction Method: Principal Component Analysis; Rotation Method: Oblimin with Kaiser Normalization.

Analysis 2: Reliability

A test-retest was conducted to measure the reliability of the EWFT. The test-retest involved forty respondents randomly selected from the first batch. The retest was conducted two weeks after the initial assessment.

The test-retest analysis was established with Pearson's product-moment correlation, while Student t-tests were used to establish whether there were differences between the first and second test sessions (Session 1 vs. Session 2). The results are presented in Tables 5 and 6. The test-retest

correlations were significant ($p < .05$), and the correlation coefficients for the most important variable, the number of correct words was .63, indicating moderate test-retest reliability (Schober & Boer, 2018). The reliability for other variables, such as the number of incorrect words and the number of positive emotion words, was lower. Although the number of generated words was generally a bit higher in the second session, there were no significant differences between the results of the first and the second session ($p > .05$).

Table 5.
Mean and SD Variables on the Emotion Word Fluency Test

	Total words		**Incorrect**		**Repetition**		**Correct words**		**Positive emotion**		**Negative emotion**	
	Mean	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***
Session 1	11.48	4.8	2.10	2.2	.53	1.3	8.85	3.3	2.93	1.5	5.98	2.5
Session 2	12.63	4.1	2.65	2.8	.60	1.1	9.38	3.2	3.25	1.3	6.25	2.7

Table 6.
The Results of The Correlation Test and The Difference in The Emotion Word Fluency Test Between the First and Second Session

	Corr.	**Sig.**	**Mean (SD)**	**95% Confidence Interval of the Difference**		***t* (*df* = 39)**	**Sig. (2-tailed)**
				Lower	**Upper**		
Total words session 1 & 2	.63	.000	1.15 (3.9)	-2.399	.099	-1.862	.070
Incorrect words session 1 & 2	.30	.059	-.55 (3.0)	-1.505	.405	-1.165	.251
Repetitions session 1 & 2	.46	.003	-.08 (1.3)	-.481	.331	-.374	.711
Correct words session 1 & 2	.63	.000	-.53 (2.8)	-1.409	.359	-1.201	.237
Positive emotion session 1 & 2	.28	.071	-.33 (1.7)	-.859	.209	-1.231	.226
Negative emotion session 1 & 2	.60	.000	-.28 (2.3)	-1.014	.464	-.753	.456

Analysis 3: Test-Order Effects

This analysis was conducted to measure whether the order of the administration (see Table 1) affected the scores of the three-word fluency tests. One-way ANOVA with order as a between-subjects factor, the outcomes can be seen in Table 7, showed no significant order effects ($p > .05$) in test scores between groups based on the order of test administration. An exception was PVF (S) ($p < .05$). Bonferroni post-hoc results showed a mean difference between groups 2 and 3 ($p < .05$), a lower score in Group 2 where PVF (S) was administered first. It is assumed that clear instructions regarding what to expect and how the task is administered might play a role. A major conclusion, however, is that there were no general order effects for the scores of the emotion word and verbal fluency (phonemic, semantic) test.

Table 7.
ANOVA With Order Test (Three Groups) As Between Subject Factors

		Sum of Squares	df	Mean Square	F	Sig.
PVFT (S)	Between Groups	173.560	2	86.780	3.184	.044
	Within Groups	4006.900	147	27.258		
	Total	4180.460	149			
PVFT (K)	Between Groups	62.440	2	31.220	1.083	.341
	Within Groups	4236.500	147	28.820		
	Total	4298.940	149			
PVFT (T)	Between Groups	60.840	2	30.420	1.124	.328
	Within Groups	3978.920	147	27.067		
	Total	4039.760	149			
SVFT A	Between Groups	.373	2	.187	.005	.995
	Within Groups	5942.460	147	40.425		
	Total	5942.833	149			
SVFT Fr	Between Groups	12.640	2	6.320	.389	.678
	Within Groups	2385.360	147	16.227		
	Total	2398.000	149			
SVFT Fu	Between Groups	16.480	2	8.240	.371	.690
	Within Groups	3261.680	147	22.188		
	Total	3278.160	149			
EWFT	Between Groups	45.453	2	22.727	1.809	.168
	Within Groups	1847.140	147	12.566		
	Total	1892.593	149			

Note: PVFT = Phonemic Verbal Fluency Test; SVFT (A) = Semantic Verbal Fluency Test - Animal; SVFT (Fr) = SVFT - Fruit; SVFT (Fu) = SVF - Furniture; EWFT = Emotion Word Fluency Test; PE = Positive Emotion; NE = Negative Emotion.

DISCUSSION

This study aimed to examine the construct validity and test-retest reliability of the Indonesian version of the EWFT administered in Bahasa Indonesia and online. The results of the construct validity test through correlation analyses showed that EWFT has a significant positive correlation with the PVFT and a slightly higher positive correlation with the SVFT. This pattern was observed across all three subscales of the PVFT, the SVFT, and the total scores of both tests. The stronger correlations between the EWFT and SVFT than between the EWFT and PVFT are noteworthy. The correlation results also showed differences in the strength of correlation of the different tests: the three subtests and total score of the phonemic tests (S, K, T, total) correlated strongly with each other, while their correlation with EWFT was weak. Similarly, the correlation of the three semantic tests with each other and the total score was high, but their correlations with the EWFT were low. The correlation between the total EWFT score and the number of positive and negative emotions is significant and strong, showing that both subscales contribute to the total score. The findings are broadly consistent with the validity measurement by Abeare et al. (2017), with a replication study by Camodeca et al. (2021). They found that EWFT correlated with the number of words produced on

the total score of PVFT (r = .46) and SVFT Animals (r = .50). Therefore, it can be concluded that the EWFT has convergent validity with other verbal fluency tests and resembles semantic verbal fluency tasks slightly more closely than phonemic verbal fluency tasks.

Overall, the findings suggest that the EWFT, SVFT, and PVFT are moderately related verbal fluency measures, although each appears to assess a distinct construct. The lower correlation of the semantic category furniture with the two other semantic categories animal names, and fruit is a point of attention for further research. Earlier research (Ardila et al., 2019) showed that the semantic category *animals* is particularly suitable cross-linguistically. Our data likewise point to the usefulness of animal names as an appropriate semantic category in Indonesian-speaking individuals.

PCA was subsequently used to test whether the three tests would indeed measure different factors. The PCA revealed a three-factor model with a high amount of variance explained. Only the number of correct words on the three phonemic subscales loaded highly *(*> .83) on Factor 1, whereas the three semantic subscales loaded moderately to highly on Factor 3 (.492–.824). The two EWFT subscales loaded highly (> .756) on the third factor; interestingly, the semantic subscale *Animals* also showed a moderate loading (.469) on this factor. This finding may reflect the semantic nature shared by both the EWFT and the Animals fluency task, and that animal names constitute a particularly suitable semantic category compared with other semantic categories.

The combination of the correlation and PCA analyses showed that the three verbal tests are positively, but only moderately, correlated with each other. They therefore share some common variance, as expected, given that they all involve processing speed, lexical access, word retrieval, and elements of executive functioning such as initiation, monitoring, and inhibition (i.e., task-related similarities). However, the PCA results indicate that the Indonesian version of the EWFT should be considered a distinct semantic measurement tool, clearly different from both other semantic word fluency tasks and from phonemic word fluency tasks. Similar results were reported by Camodeca et al. (2021) and Abeare et al. (2017).

EWFT involves semantic functions similar to classical semantic fluency tasks. Participants, when working on a semantic task, need a strategy to connect the concept of emotion with the resulting words (Shao et al., 2014). The lower number of words generated in the EWFT compared to the SVFT, see Table 3, suggests that this is more difficult than generating words within semantic categories such as animals, fruit, or furniture. In addition, the performance of the EWFT test is more influenced by the ability to process emotions than language skills (Wauters & Marquardt, 2018), so EWFT may be useful in identifying emotional processing difficulties. Furthermore, in EWFT, participants are asked

to perform a recall task, where the recalled emotions can also be influenced by various types of emotions stored in memory as a result of daily experiences (Planck et al., 2020).

The EWFT may involve different parts of the brain, considering the emotional content of the words, compared to the non-emotional content of the other fluency tasks. Earlier fMRI and EEG studies have already shown this: the SVFT involves the temporal lobe, the PVFT predominantly the frontal lobe, and EWFT the prefrontal area, amygdala, and hippocampus, dorsolateral prefrontal cortex, and parietal and insular cortex, the part of the brain that plays a role in processing emotions (Mousavi et al., 2020; Li et al., 2017; Bevilaqua et al., 2016; Auerbach et al., 2015; Schreiter et al., 2019).

Differences among the EWFT, PVFT, and SVFT were also reflected in the average number of correctly generated words. It is the lowest for the EWFT. This could explain the difficulty and different task demands when generating neutral and emotional words. These findings suggest that generating emotion-related word lists may impose greater cognitive demands compared to the generation of neutral category words. This might lead to a lower number of emotion words generated (Wauters & Marquardt, 2018). EWFT can be regarded as a more difficult word fluency test than the two other verbal fluency tests (Abeare et al., 2021).

EWFT reliability was measured through test-retest reliability with an interval of two weeks, both times via supervised-internet delivered testing (online). The results of the analysis showed that there was a significant correlation between sessions 1 and 2 with Pearson correlation scores for the number of correctly generated emotion words (.63.). This coefficient indicates moderate test-retest reliability (Schober & Boer, 2018). This moderate reliability can be due to the instability of the construct that is measured. The type and number of emotion words given by the participants can be shaped by recently experienced emotions, and the type of emotions that are occurring may vary over time. Such an assumption can be based on the “Mood Congruent Memory” theory (for review see Faul and LaBar, 2023), stating that transient emotional feelings and experiences and longer-term mood states may affect the type of memories and perhaps also the type of emotion words that are generated by the participants assessed with the EWFT. Besides, a longer test-retest interval may enhance the experience of more and different emotions, enhance the intra-subject variation, and this negatively influences the test-retest reliability. Abeare et al. (2017) used an interval of one week and found a coefficient of .68, our coefficient was .63 with an interval of two weeks. The results in the two studies are broadly similar and may suggest that a longer test-retest interval decreases the test-retest reliability. There was no significant difference between the number of correct words of the first and second sessions. Thus, it can be said that there is no practice or learning effect on the EWFT,

suggesting that the test may be suitable for repeated administration in longitudinal assessments to monitor a client's ability. The results obtained also show that the EWFT is sufficiently reliable in measuring the semantic category of emotion words, although some caution is warranted for the interpretation of the test scores.

The design of our study, with three groups of test sequences, allowed us to test whether the order in which the tests were administered affected the participants' test performance. The results showed that only group two, in which PVFT (S) was administered first, had lower scores than the other sequences. Therefore, the administration of the test sequence also needs to be considered or studied further, especially in the implementation of the PVF test, perhaps with more instructions and a probe trial. On the other hand, the EWFT presented as the first test in group 1 did not show a significant difference compared to when it was administered as the second or third. The results obtained are the same as in previous research, where the order of test administration has no large effect or influence on the test scores (Ryan et al., 2019).

Our results (Table 5) with the Indonesian EWFT showed more negative than positive emotion words. The results are consistent with what is internationally reported (Abeare et al., 2017) and with our results in our previous study (Pesau et al., 2023). The number of correct words reflects language proficiency and is influenced by the number of emotion words in a given language, as well as the ethnic and cultural background, and the appropriateness of expressing a certain type of emotion. The recognition of facial expressions of emotions seems universal, the description of what is felt in different situations and what is expressed might vary between different ethnic groups. We did not ask for the ethnic background of our participants, nor their proficiency in their daily spoken local language, if any. We only knew that our participants were fluent in Bahasa Indonesia; it was an inclusion criterion. It is common to assess people in their most proficient language. It might be interesting, given Indonesia's linguistic and ethnic diversity, to investigate whether the same results will be obtained for a more diverse population regarding spoken language(s) and ethnic groups (Pesau et al., 2023). Possible differences between Indonesian ethnic groups on performance on language tests, including the EWFT are feasible, considering that differences in word production tasks were found between Balinese and Banjarnese (Pesau et al., 2022).

Cross-cultural variations in the cognitive representation of emotions may regard linguistics and internal emotional processes (Fontaine et al., 2002). Regarding linguistic processes, differences in cultural influences on the term emotion were found in a semantic analysis of Indonesian and Malay Asahan (Mulyadi et al., 2012). Among these groups, there are differences in the meaning of the word emotion, in both the categories of static emotion verbs and active verbs, as well as differences in

thematic relations in emotion verbs. Likewise, Fontaine et al. (2002) found differences in the position of the emotional words "shame" and "guilt" in cognitive maps between Indonesian and Dutch students. Cross-cultural research was also carried out in Indonesia (e.g., Dewi et al., 2018 regarding emotional intelligence), and another comparison between six ethnic groups showed differences in the concept of anger, verbal expressions, and terms of anger between tribes. These authors (Zuhdi & Nuqul, 2022) mentioned that these differences cannot be separated from the cultural context of the respondents, in this case, the Indonesian people, who are in a collective culture that influences the categorization or labeling by society of a culture. Dasen (2022) explained that cognitive processes are universal, but there are cultural differences in cognitive style and development. Therefore, from a cross-cultural perspective, the Indonesian EWFT, as an instrument designed to assess emotion words modulated by brain function, cognitive style, and developmental processes, may yield outcomes that differ from scores obtained in other countries. It may also vary across ethnic groups, particularly when the test is administered in a local language. A final remark is that all three word fluency tests can be administered online, which might be advantageous within the archipelago, with its many remote islands and regions.

CONCLUSION AND RECOMMENDATIONS

Construct validity analyses using comparison with two other word fluency tests (SVFT and PVFT) showed that the EWFT involves task demands and cognitive strategies that are unique and different from the commonly used phonemic and semantic verbal fluency tasks. EWFT specifically measures the ability to generate emotion words. More negative than positive emotion words were generated. The test-retest reliability is moderate, which implies that its interpretation should be done with care. Test sequence effects are small, but they might be present, and therefore, they should not be completely ignored. Perhaps, practice trials should be considered for future study.

Based on these results, it is suggested that the online-administered EWFT may have potential to be a screening test for a neuropsychological evaluation among healthy Indonesian individuals related to their emotional processing skills. The test can be administered online for those who have the skills for online video calls. The Indonesian online-administered EWFT can be repeatedly assessed with no significant differences between the first and second assessments. However, studies involving Indonesian patient populations are still needed to test its usability and feasibility in various settings before it can be utilized for clinical purposes.

ACKNOWLEDGEMENT

The authors would like to thank research assistants and all participants who participated in this study.

FUNDING

This work was supported by the Directorate of Higher Education General of Indonesia under grant number: 6570/LL9/KU.03.00/2021.

COMPLIANCE WITH ETHICAL STANDARD

Ethical Statement

All procedures performed in this study were in accordance with the 1964 Helsinki Declaration and its amendments or with comparable ethical standards. The ethical aspect of the study has been institutionally reviewed. Informed consent has been obtained from all participants in this study.

Conflict of Interest

All authors declare no conflict of interest.

Data Availability

The datasets associated in this study are not publicly available due to concerns about privacy (participant consent)

USE OF AI SERVICES

The authors declare they have used AI services, specifically for grammar correction and minor style refinements. The authors carefully reviewed all suggestions from these services to ensure the original meaning and factual accuracy were preserved.

AUTHOR CONTRIBUTIONS

HGP and GvL conceptualized the study. HGP conducted data collection. HGP and GvL analyzed the data. HGP and GvL jointly wrote and revised the manuscript. Both authors approved the final manuscript.

REFERENCES

Abeare, C. A., Freund, S., Kaploun, K., McAuley, T., & Dumitrescu, C. (2017). The Emotion Word Fluency Test (EWFT): Initial psychometric, validation, and physiological evidence in young adults. *Journal of Clinical and Experimental Neuropsychology, 39*(8), 738–752. https://doi.org/10.1080/13803395.2016.1259396

Abeare, C. A., An, K., Tyson, B., Holcomb, M., May, N., & Erdodi, L. A. (2021). The emotion word fluency test as an embedded performance validity indicator – Alone and in a multivariate

validity composite. *Applied Neuropsychology: Child, 11*(4), 713–724. https://doi.org/10.1080/21622965.2021.1939027

Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2018). 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

Altarriba, J. (2003). Does cariño equal "liking"? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism*, *7*(3), 305–322. https://doi.org/10.1177/13670069030070030501

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf

Auerbach, R. P., Stewart, J. G., Stanton, C. H., Mueller, E. M., Pizzagalli, D. A. (2015). Emotion-processing biases and resting EEG activity in depressed adolescents. *Depression and Anxiety, 32*(9), 693–701. https://doi.org/10.1002/da.22381

Bevilaqua, C., Carvalho, M. R. De, Ribeiro, P., Palmer, S., Nardi, A. E., & Dias, G. P. (2016). Electroencephalographic findings in patients with major depressive disorder during cognitive or emotional tasks: A systematic review. *Revista Brasileira de Psiquiatria 38*(4), 338–346. https://doi.org/10.1590/1516-4446-2015-1834

Camodeca, A., Walcott, K., Hosack, A., & Todd, K. Q. (2021). Preliminary Evidence for the Emotion Word Fluency Test as a unique semantic fluency measure. *Psychological Assessment, 33*(2), 195–200. https://doi.org/10.1037/pas0000965

Capizzi, R., Fisher, M., Biagianti, B., Ghiasi, N., Currie, A., Fitzpatrick, K., Albertini, N., Vinogradov, S. (2021). Testing a novel web-based neurocognitive battery in the general community: Validation and usability study. *Journal of Medical Internet Research, 23*(5), Artikel e25082. https://doi.org/10.2196/25082

Dasen, P. R. (2022). Culture and cognitive development. *Journal of Cross-Cultural Psychology*, 53(7–8), 789–816. https://doi.org/10.1177/00220221221092409

Dewi, Z. L., Halim, M. S., & Derksen, J. (2018). Emotional intelligence competencies of three different ethnic groups in Indonesia. *Asian Ethnicity*, *19*(1), 36–58, https://doi.org/10.1080/14631369.2017.1310615

Djiwatampu, L. W. (2009). Happy and shameful experiences among ethnic groups in Indonesia: Cognitive, affective and behavioral dimensions. *Psychology: The Journal of Hellenic Psychological Society, 16*(2), 175–184. https://doi.org/10.12681/psy_hps.23812

Faul, L., & LaBar, K. S. (2023). Mood-congruent memory revisited. *Psychological Review*, *130*(6), 1421–1456. https://doi.org/10.1037/rev0000394

Frijda, N. H., Markam, S., Sato, K., & Wiers, R. (1995). Emotions and emotion words. In Russell, J. A., Fernández-Dols, J. M., Manstead, A. S. R., Wellenkamp, J. C. (Eds.), *Everyday conceptions of emotion*. Springer. https://doi.org/10.1007/978-94-015-8484-5_7

Fontaine, J. R. J., Poortinga, Y. H., Setiadi, B., & Markam, S. S. (2002). Cognitive structure of emotion terms in Indonesia and The Netherlands. *Cognition and Emotion, 16*(1), 61–86. https://doi.org/10.1080/02699933014000130

Hegefeld, H. M., Satpute, A. B., Ochsner, K. N., Davidow, J. Y., & Nook, E. C. (2023). Fluency generating emotion words correlates with verbal measures but not emotion regulation, alexithymia, or depressive symptoms. *Emotion*, *23*(8), 2259–2269. https://doi.org/10.1037/emo0001229

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

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

Li, Y., Li, P., Yang, Q. X., Eslinger, P. J., & Sica, C. T. (2017). Lexical-semantic search under different covert verbal fluency tasks: An fMRI study. *Frontiers in Behavioral Neuroscience, 11*, Artikel 131. https://doi.org/10.3389/fnbeh.2017.00131

Masturah, A. M. (2019). *Which emotion is preferred by Indonesian People?* [Conference paper]. 4th ASEAN conference on psychology, counseling, and humanities (ACPCH 2018), Indonesia. https://doi.org/10.2991/acpch-18.2019.10

Mulyadi, Beratha, N. L. S., Oktavianus, Sudipa, I. N. (2012). Emotion verbs in bahasa Indonesia and Asahan Malay language: Cross-language semantics analysis. *e-Journal of Linguistics, 6*(1), 1–19. https://ojs.unud.ac.id/index.php/eol/article/view/4584

Mousavi, N., Ali, M., Jalil, N., & Ali, B. (2020). Electroencephalographic characteristics of word finding during phonological and semantic verbal fluency tasks. *Neuropsychopharmacology Reports, 40*(3), 254–261. https://doi.org/10.1002/npr2.12129

Olson, K. E., O'Brien, M. A., Rogers, W. A., Charness, N. (2011). Diffusion of technology: Frequency of use for younger and older adults. *Ageing International, 36*(1), 123–145. https://doi.org/10.1007/s12126-010-9077-9

Pesau, H. G., Immanuel, A. S., Sulastri, A., Wulanyani, N. M. S., & van Luijtelaar, G. (2022). *The influence of ethnicity on language tests: A comparison between Balinese and Banjarese*. [Conference paper]. The 2nd International Conference on Biopsychosocial Issues, Semarang, Indonesia. https://conference.unika.ac.id/index.php/ssic/iconbi/paper/view/524

Pesau, H. G., Immanuel, A. S., Sulastri, A., & van Luijtelaar, G. (2023). The role of daily spoken language on the performance of language tests: The Indonesian experience. *Bilingualism: Language and Cognition*, *26*(3), 538–549. https://doi.org/10.1017/S136672892200075X

Pesau, H. G., Utami, M. S. S., Sulastri, A., Suryani, A., Sanjaya, R., Kiling, I. Y., & Damayanti, Y. (2025). Comparison of traditional and tele-neuropsychological testing. *Acta Neuropsychologica*, *23*(2), 189–198. https://doi.org/10.5604/01.3001.0055.1278

Pesau, H. G., & van Luijtelaar, G. (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

Pesau, H. G., Sulastri, A., & van Luijtelaar, G. (2024). Evaluation of Indonesian Version of the Emotion Word Fluency Test. *Jurnal Psikologi*, *22*(2), 157–176. https://doi.org/10.14710/jp.22.2.157-176

Planck, M., Development, H., Masitah, A., & Hills, T. T. (2020). The emotional recall task: Juxtaposing recall and recognition-based affect scales. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(9), 1782–1794. https://doi.org/10.1037/xlm0000841

Poder, T. G., Coulibaly, L. P., Hassan, A. I., Conombo, B., Laberge, M. (2022). Test–retest reliability of the Cost for Patients Questionnaire. *International Journal of Technology Assessment in Health Care, 38(1),* Artikel e65. https://doi.org/10.1017/S0266462322000460.

Röttger-Rössler, B., & Markowitsch, H. J. (2009). Introduction. In B. Röttger-Rössler & H. J. Markowitsch (Eds.). *Emotions as bio-cultural processes* (pp. 1–10). Springer.

Ryan, J. J., Glass, L. A., Hinds, R. M., & Brown, C. N. (2019). Administration order effects on the test of memory malingering. *Applied Neuropsychology, 17*(4), 246–250. https://doi.org/10.1080/09084282.2010.499802

Schober, P., & Boer, C. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia, 126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schreiter, M. L., Chmielewski, W. X., Mückschel, M., Ziemssen, T., & Beste, C. (2019). How the depth of processing modulates emotional interference – evidence from EEG and pupil diameter data. *Cognitive, Affective, & Behavioral Neuroscience, 19,* 1231–1246. https://doi.org/10.3758/s13415-019-00732-0

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology, 5*, Artikel 772. https://doi.org/10.3389/fpsyg.2014.00772

Villalobos, D., Torres-Simón, L., Pacios, J., Paúl, N., & Del Río, D. (2023). A systematic review of normative data for verbal fluency test in different languages. *Neuropsychology Review, 33*(4), 733–764. https://doi.org/10.1007/s11065-022-09549-0

Wahyuningrum, S. E., Sulastri, A., Hendriks, M. P. H., & 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

Wauters, L., & Marquardt, T. P. (2018). Category, letter, and emotional verbal fluency in Spanish-English bilingual speakers: A preliminary report. *Archives of Clinical Neuropsychology, 33*(4), 444–457. https://doi.org/10.1093/arclin/acx063

Wyse, A. E. (2021). How days between tests impacts alternate forms reliability in computerized adaptive tests. *Educational and Psychological Measurement, 81*(4), 644–667. https://doi.org/10.1177/0013164420979656

Zuhdi, M. S., & Nuqul, F. L. (2022). Konsepsi emosi marah dalam perspektif budaya di Indonesia: sebuah pendekatan indigenous psychology. *Jurnal Psikologi, 18*(1), 51–62. https://doi.org/10.24014/jp.v18i1.14680