Psychometric Properties of a Newly Online Work Aptitude Test: Faxtor's Endurance and Speed Test

Whisnu Yudiana
Faculty of Psychology
Universitas Padjadjaran
whisnu.yudiana@unpad.ac.id
(corresponding author)

Aryo Bimo Adjie
Faculty of Psychology
Universitas Padjadjaran
aryo18001@mail.unpad.ac.id

Nadim Alfan Assa
Faculty of Psychology
Universitas Padjadjaran
nadim18001@mail.unpad.ac.id

Anjar Karta Putra
Faculty of Psychology
Universitas Padjadjaran
anjar14001@mail.unpad.ac.id

Abstract

The study aimed to investigate the psychometric properties of a new computer-based work aptitude test, named Faxtor's Endurance and Speed Test (F-EAST). Two phases of study were conducted to examine the reliability of the test and collect the validity evidence based on the relation to other variables. This study used quantitative research with repeated measures design for a total of 116 participants who participated in two studies (1=63, 54.3% and 2=53, 45.7%). Reliability was evaluated through Cronbach's alpha and Intraclass Correlation Coefficient (ICC) with 95% CI, mean-rating (k=2) and absolute agreement, while validity evidence was collected through correlation with other tests that assess well-known work aptitude: Kraepelin Test and Pauli Test. The results shown that F-EAST has adequate internal consistency (α=0.751 - 0.987) and ICC (r=-0.220 - 0.925). The F-EAST also has satisfactory correlation with the Pauli test (r = 0.113 - 0.635) and the Kraepelin test (r=0.054 - 0.578) in similar or identical aspects of measurement. The current study concluded that F-EAST is a suitable test to measure work aptitude.

Keywords: *Work aptitude test, F-EAST, Reliability, Validity*

Received 1 December 2022/Accepted 27 February 2023

Introduction

Work aptitude is one of psychological aspects widely used by psychologist (Pradipta & Hidayat, 2019) to describe an individual personality in educational (Borualogo & Rabayani, 2009), occupational (Hidayat, 2016a, 2016b; Prasetiya et al., 2017) and clinical settings (Li et al., 2004; Nagarubini Paramasivam, 2021; Sugimoto et al., 2009). In a person, the work aptitude describes the energy, persistence, adaptation speed, consistency and work accuracy in the routine tasks (Arnold, 1975; Mikicin et al., 2015). Psychologists use information about work aptitude for vocational counseling and clinical diagnosis (Mikicin et al., 2015).

In Indonesia, the work aptitude is commonly measured by continuous work tests or also known as "Work Curve Test" namely Kraepelin Test (Kraepelin, 1922) and Pauli Test (Arnold, 1975). Those tests are designed to measure a person's task performance accuracy, speed and effectiveness performance in work situations, and character of mental stress (Arnold, 1975; Kraepelin, 1922; Reuning, 1983; Thornton III & Kedharnath, 2013). Originally, the tests were administered using paper and pencil based tests, which presented participants with numerous numbers in large paper. The assessment is conducted face to face both in an individual or group setting. The participants are required to perform additional calculations as fast and accurately as possible in several periods of time (Arnold, 1975; Kraepelin, 1922). Later, the administrator will score the test manually by calculating the number of complete and correct answers for every period of time representing the shape of the work curve or the work aptitude of the person.

Even though the Kraepelin and Pauli test is popular in Central Europe, Japan and Indonesia, the tests showed several limitations. Reuning (1983) explained that this test lacks validity evidence because appropriate criteria for validation are hard to find. He also explained the paper and pencil-based test requires a high amount of time for both administration and scoring process. Sometimes this process cannot be tolerated by the administrator due the limited time of assessment to assess a number of participants. That reason becomes one of the biggest disadvantages of using these tests. Recently, as technology develops, the obstacle can be overcome by applying computer based testing.

The development of computer technology provides a lot of possibilities for creating computer applications in psychological assessment (Naglieri et al., 2004). Furthermore, in the COVID-19 pandemic, many face to face interactions have migrated to online platforms and assessments (Adedoyin & Soykan, 2020; Javaid et al., 2020). Recently, in Indonesia several researchers, consisting of information and technology practitioners, have created computer based tests to measure work aptitude by converting the procedure Kraepelin's paper and pencil test into computer administration (Nada et al., 2022; Pane et al., 2020). The research reported that the development accelerated and increased the accuracy of scoring processes (Nada et al., 2022). Furthermore no different results in scoring were found between computer based assessment and the manual scoring process by administrators (Nada et al., 2022; Pane et al., 2020). However, that research did not report the

psychometric properties of the computer based test relating to reliability and validity evidence of the test mandatory for every computer based test (ITC, 2006; Lievens, 2006).

The aim of this study is to develop a computer based assessment to measure work aptitude, which will be called Faxtor's Endurance and Speed Test (F-EAST). This article consists of two studies. The first study aims to report the item development processes and the reliability of F-EAST. Reliability index of F-EAST tested by internal consistency and intraclass correlation coefficient (ICC), while the second study reported the validity evidence based on the relation with other variable by the conducting the correlation analysis F-EAST with other test measuring same constructs namely Kraepelin and Pauli test as the standardized test to measure work attitude.

Method

Study 1

Participants

The subjects of first study were 63 students, who participated voluntarily, from universities in Bandung Indonesia. The majority of participants were females (n=54, 85%), and the age ranged from 19 – 25 years (M = 21.43 years, SD =1.32).

Instrument

Faxtor Endurance and Speed Test (F-EAST) measures one's work aptitude by performing continuous tasks requiring participants to evaluate simple addition of one-digit numbers in a particular time period as fast and accurately as possible. Every question is delivered in a multiple choice format requiring participants to determine whether the result of the addition is an odd, represented by '1', or even, represented by '0' number (see figure 1). The test contains 20 blocks with 100 questions in each block. The participants are given one minute to answer questions on every block, making the total time needed to complete this test is 20 minutes. In Addition, the test was developed and administered in a computer-based format that uses computers to score items automatically. Every correct answer scores as '1' while incorrect answer scores as '0'. The total responses will be evaluated as work performance of the participants.

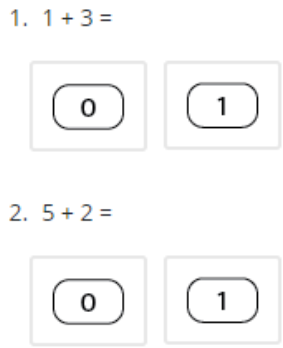

Figure 1. F-EAST Appearance on participant's view

F-EAST intended to portray how one's work in a real work situation. Thus, with simulating monotonous tasks, F-EAST measures six constructs presented in table I.

Procedure

Researchers contacted the potential participants via students' social media groups to share the link. This study was voluntary participation; therefore, before commencing the study, researchers explained the procedures and processes of the study for gaining consent from every participant. Every participant received Rp.50.000 as a reward after they completed the process. F-EAST was administered online through Faxtor's test platform in which the participants took the test two times within a 30 days interval.

Tabel 1

Operationalization Aspects of F-EAST

Aspects	Operationalization
Workload Capacity (Total Worked Out)	Workload capacity was measured by the number of simple additions that individuals can be answered in 20 blocks.
Thoroughness (Error rate %)	Thoroughness was measured by the number of incorrect responses given over a period of 20 blocks.
Emotional Stability (Deviation rate %)	Emotional stability was measured by the percentage of the overall difference between the workmanship average graph and the graph created from the work completed on 20 blocks.
Achievement Motive (Range)	Achievement driving was measured by the difference between the highest and lowest score in 20 blocks.
Ambition and Planning (Peak performance)	We gauge ambition and planning by examining the block with the highest score.
Working Style (Graph Types)	An average workmanship graph for 20 blocks divided into 6 categories was used to gauge working style.

Analyses

The F-EAST scores analyzed in this study were the total number of calculations, error rate, and percentage of deviation rate. Every score of the F-EAST is categorized into three levels: Poor when the score is lower than one standard deviation of the mean; Good when the score is higher than one standard deviation of the mean; Fair when the score is between minus and one standard deviation of the mean. For the graph of F-EAST was categorized using standardization of graphs in the Pauli Test (Arnold, 1975).

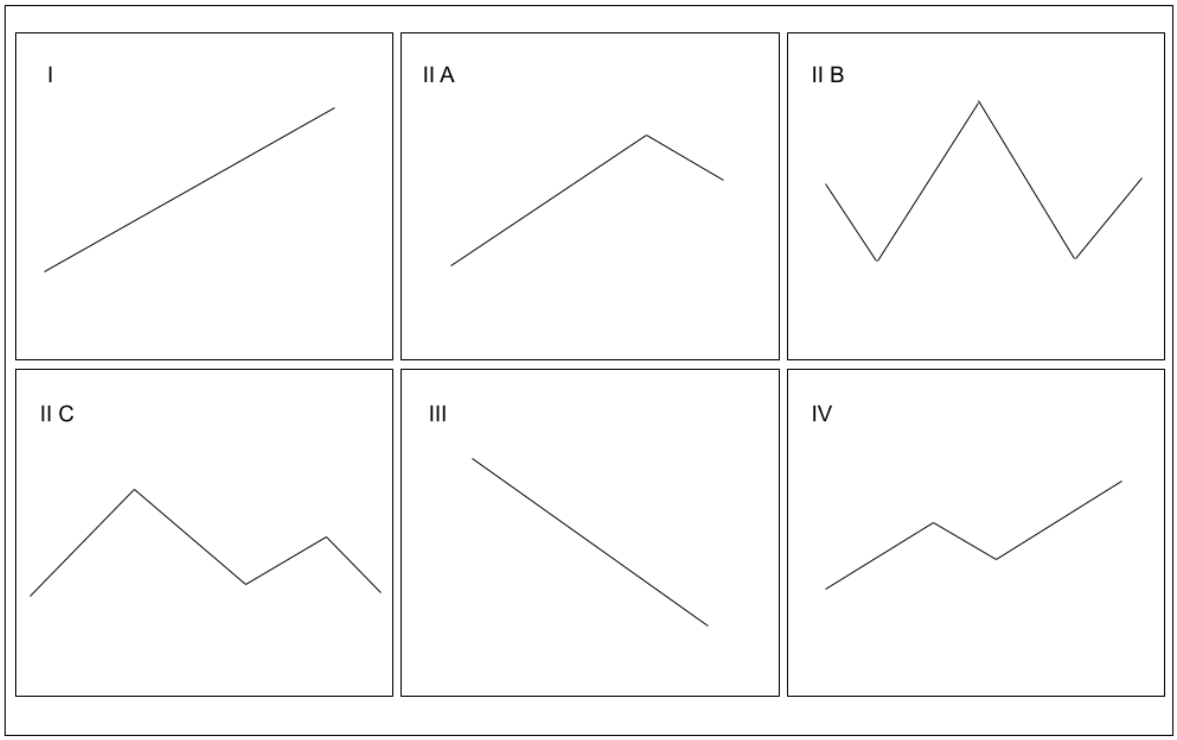

Figure 1. Categorization Graph Types of Pauli Test (Arnold, 1975)

Cronbach's Alpha reliability was conducted to analyze the internal reliability aspects of workload capacity, error rate, and deviation rate based on the score on the 20 block of the F-EAST. Kaplan & Saccuzo (2017) suggest that a reliability value higher than 0.70 indicates good reliability. Due to the variability of population and the two-time points data collection, the test-retest reliability was also calculated using the Intraclass Correlation Coefficient (ICC). The estimation used a 95% confidence interval based on mean-rating (k=2) and absolute agreement (Koo & Li, 2016). Koo & LI (2016) also described that an ICC's value lower than 0.5 indicates poor reliability; a value between 0.5 and 0.75 indicates moderate reliability, a value between 0.75 and 0.9 indicates good reliability, and values greater than 0.90 indicate excellent reliability. The reasons for a low ICC can reflect the low degree of measurement agreement, the lack of variability among the sampled subjects, and the small number of participants (Portney & Watkins, 2009; Lee et al., 2012). Characteristics of the sample were analyzed using descriptive statistics and frequency distribution. The analyses were performed using the statistical software program, SPSS version 25.

Study 2

Participants

The subjects used in this study 2 were fresh graduated students (n = 32, 60.4%) and workers (n = 21, 39.6%) in Bandung, Indonesia. Before commencing the study II, informed consent was gained from all the participants, either autonomously or through their legal representatives. The sample consisted of 53 students (35 females and 18 males). The mean age of the participants in this study was 22.72 (1.42).

Instrument

Another work aptitude test was measured in this study 2 using the "Work Curve Test" namely Kraepelin Test and Pauli Test. The Pauli test was created to measure the speed, effectiveness and accuracy of work (Arnold, 1975). The purpose of this test is to perform the calculation task as quickly and accurately as possible, within one hour. participant is asked to perform the task of adding two digits in adjacent columns and writing the results obtained on the right of the column.

The Kraepelin test is a serial addition test, which requires takers to perform calculations as fast and accurately as possible within 45 min. This was achieved using printed paper containing 45 lines of random, single-digit, vertically aligned numbers. For each minute of the test, the subject was instructed to begin a new line regardless of their position on the current line. Each line contained an excess of calculations such that the subjects were not able to finish any line for a particular minute before being prompted to move on to the start of the next minute by the examiner's prompting.

Procedure

This study was voluntary participation; therefore, before commencing the study, researchers explained the procedures and processes of the study for gaining consent from every participant. Every participant received Rp100.000 as a reward after they completed the process. F-EAST, Pauli, and Kraepelin were administered to each participant in the same day test. Quantitative research with repeated measure design was used in this study with counterbalancing techniques to reduce bias effect. The participants were divided into three groups with different order of testing in each group. The tests' sequence of the Group I were Pauli, Kraepelin, and F-EAST. While the tests'

sequence of the Group 2 were Kraepelin, F-EAST, and Pauli. Last, The tests' sequence of the Group 3 were F-EAST, Pauli, and Kraepelin.

Analyses

The correlation coefficient was used to determine the validity based on correlation with other variables of each aspect F-EAST, Pauli, and Kraepelin. Higher correlations were observed between similar or identical aspects, illustrating validity based on correlation with other variables for the F-EAST, Pauli, and Kraepelin aspects. Correlation coefficient is a number between −1 and 1 that expresses the degree of linear dependence between two quantitative variables. If negative, it indicates that one variable decreases as the other increases; if positive, it indicates that one variable increases as the other increases (Johnson, 2007). The r values are distributed as follows: r = 0 – 0.25, very low correlation; r = 0.26 – 0.49, low correlation; r = 0.5 – 0.69, moderate correlation; r = 0.7 – 0.89, high or strong correlation; r = 0.9 – 1.0, very high or very strong correlation (Kozak, 2009). The statistical tool SPSS version 25 was used to conduct the analysis.

Tabel 2
Operationalization Aspects of Pauli and Kraepelin

Aspects	Pauli	Kraepelin
Workload Capacity	The number of simple addition that individuals answered in 20 blocks.	The average of calculations done in each line.
Error Rate	The percentage of erroneous answers given from time 13 to 15.	The average of the error calculation in each line.
Deviation rate (%) / Working Curve	The percentage of the overall difference between the workmanship average graph and the graph created from the work completed on 20 blocks.	The score difference between the highest and lowest score.
Range	The score difference between the highest and lowest score in 20 blocks.	-
Peak performance	Peak performance can be determined by examining the period with the highest score.	-
Graph Types[a] / Working Style	An average workmanship graph for 20 blocks divided into 6 categories.	Workload capacity score divided by the working curve score.

Result

Study 1

A total of 63 students completed F-EAST twice in this study within a four week time interval. Table 3 presents the F-EAST score categorization of every aspect from the first and second tests. Reliability of the F-EAST was calculated using the internal consistency coefficient and test-retest reliability. As can be seen in table 4, the internal consistency value measured by Cronbach's alpha for three aspects of F-EAST show good results as they are higher than the standard of .70 (Kaplan & Saccuzo, 2017). Specifically, the values were from 0.751 to 0.987 and from 0.768 to 0.989 in the first and second test, respectively.

Table 5 shows the intraclass correlation for several aspects of the F-EAST. Due to interpretation concern, test-retest reliability F-EAST values were measured by using Intraclass correlation coefficients (ICCs) estimation calculated based on two types of data, namely raw and categorization data. Based on the raw data, the F-EAST has good reliability on the workload and error rate aspects since the ICC's values are higher than 0.30, Specifically, the ICC values were between -0.220 (95% CI = -0.997 – 0.262) and 0.925 (95% CI = 0.876 – 0.955). While based on the categorization, the FEAST has also good reliability on workload, error rate and deviation rate aspects since the ICC's values are also higher than 0.30, between -0.145 (95% CI = -0.892 – 0.307) and 0.835 (95% CI = 0.727 – 0.900).

Study 2

In this study, a total of 53 participants completed the F-EAST, Pauli and Kraepelin tests to obtain validity evidence based on correlation with other variables. Generally, this study expects high correlations between similar or identical constructs measured by F-EAST, Pauli, and Kraepelin. The results of Spearman's rho correlations, descriptive statistics for aspects F-EAST, Pauli, and Krapelein are presented in Table 6.

Table 3
Demographic Categorization of Aspects F-EAST

Aspects	Categorization	%	Score Categorization Poor n (%)	Fair n (%)	Good n (%)
Test I					
Workload Capacity			26 (41.27%)	31 (49.21%)	6 (9.52 %)
Error Rate			34 (54.00%)	19 (30.20%)	10 (15.90%)
Deviation Rate			24 (38.10%)	30 (47.62%)	9 (14.29%)
Range			23 (36.50%)	25 (39.70%)	15 (23.80%)
Peak Performance			24 (38.10%)	20 (31.75%)	19 (30.16%)
Graph Type					
I	4	6.35%			
II A	8	12.7%			
II B	12	19.05%			
II C	7	11.11%			
III	1	1.59%			
IV	31	49.21%			
Test II					
Workload Capacity			17 (27.00%)	30 (47.60%)	16 (25.40%)
Error Rate			23 (36.50%)	28 (44.50%)	12 (19.00%)
Deviation Rate			32 (50.80%)	23 (36.50%)	8 (12.70%)
Range			25 (39.70%)	15 (23.80%)	23 (36.50%)
Peak Performance			12 (19.00%)	28 (44.40%)	23 (36.50%)
Graph Type					
I	5	7.95%			
II A	8	12.70%			
II B	14	22.20%			
II C	6	9.50%			
III	5	7.95%			
IV	25	39.70%			

Table 4
Internal Consistency Cronbach's Alpha aspects F-EAST

F-EAST	Ɑ 1 (n = 63)	Ɑ 2 (n = 63)	SEM 1	SEM 2
Workload Capacity	0.987	0.989	25.768	27.059
Error rate	0.806	0.984	1.189	1.834
Deviation rate	0.751	0.768	2.359	3.179

Table 5
Intraclass correlation coefficients aspects F-EAST

F-EAST		Test 1 M (SD)	Test 2 M (SD)	Mean Diff (SD Diff), 95% CI	ICC (95% CI)
Raw Data	Workload Capacity	653 (226)	729 (258)	76 (32), 68 – 84	0.925 (0.876 – 0.955)
	Error Rate (%)	3.4 (2.7)	7.2 (14.5)	3.8 (11.8), 0.8 - 6.7	0.345 (-0.082 – 0.604)
	Deviation rate (%)	9.9 (4.8)	10.3 (6.6)	4.0 (1.8), -0.1 - 0.8	0.270 (-0.207 – 0.558)
	Range	19.56 (12.43)	21.52 (12.30)	1.97 (-0.14), 2.00 - 1.93	0.086 (-0.511 – 0.447)
	Peak performance	15.05 (3.946)	13.32 (5.88)	-1.73 (1.93), -2.22 - -1.24	-0.220 (-0.997 – 0.262)
Category (1 - 3) Poor-Good	Workload Capacity	1.68 (0.64)	1.98 (0.73)	0.3 (0.09), 0.28 - 0.32	0.778 (0.534 – 0.886)
	Error Rate (%)	1.76 (0.69)	1.83 (0.73)	0.06 (0.04), 0.05 - 0.07	0.835 (0.727 – 0.900)
	Deviation rate (%)	1.62 (0.75)	1.62 (0.71)	0 (-0.04), 0.01 - -0.01	0.425 (0.049 – 0.652)
	Range	2.1 (0.76)	1.97 (0.88)	-0.13 (0.12), -0.16 - -0.1	-0.145 (-0.892 – 0.307)
	Peak performance	1.92 (0.83)	2.17 (0.73)	0.25 (-0.1), 0.28 - 0.23	0.094 (-0.497 – 0.452)

The results showed that four aspects of the FEAST, namely: workload capacity (r = 0.635), error rate (r = 0.301), deviation rate (r = 0.339), and Graph Types (r = 0.557), correlated positively with Pauli. While, the other aspects, specifically: range (r = 0.156) and peak performance (r = 0.113), were not statistically significant. In addition, based on the graph type, F-EAST and Pauli were not statistically different (U = 1305.50, z = -0.643, p>0.05, d=0.116) indicating the similar pattern between those tests. Thus, the results indicate that the F-EAST has good validity evidence based on correlation with Pauli, specifically on Total worked out, error rate, deviation rate, and graph type of work aspects.

Workload capacity F-EAST were statistically significant correlations with workload capacity in Kraepelin (r = 0.576). Error rate with wrong answer (r = 0.578). And graph type of work in F-EAST with working style in Kraepelin (r = 0.310). Deviation rate F-EAST and working curve Kraepelin theoretically have similar or identical aspects, but not statistically significant correlation. Working

curves were statistically significant with range in F-EAST (r = 0.369), this can be because the way to measure range in F-EAST is the same as the way to measure working curve in Kraepelin. Thus, the results indicate that the F-EAST has good validity, evidence based on correlation with Kraepelin on workload capacity, error rate, and graph type of work.

Pauli and Kraepelin are two speed tests that have been widely used around the world. There are several aspects that are measured in Pauli and Kraepelin theoretically: total worked out with workload capacity, error rate with wrong answer, deviation rate with working curve, and graph type with working style. Correlations coefficient between similar or identical aspects describe convergent validity for the Pauli and Kraepelin aspects. Total worked out Pauli were statistically significant correlations with workload capacity in Kraepelin (r = 0.820). Deviation rate Pauli and working curve Kraepelin also have theoretically similar or identical aspects, but not statistically significant correlation (r = 232). Working curves in Kraepelin were statistically significant with range in Pauli (r = 0.412) possibly because the way to measure range in Pauli is the same as the way to measure working curve in Kraepelin.

Discussion

Study 1 aims to examine the reliability of F-EAST using internal consistency and intraclass correlation coefficient (ICC). The results show that the F-EAST, based on the quality standards (Kaplan & Saccuzzo, 2017), has good internal reliability, especially on the aspects of total worked out, error rate, and deviation rate. The internal consistency results indicate that all the items in the F-EAST measure the same constructs. A good internal consistency also means that the inter-relatedness is present within test items (Tavakol & Dennick, 2011). In addition, the internal consistency values of the F-EAST were higher than the 0.7, reflecting low standard error of measurements which is important in development of instruments.

The ICC analyses were performed to evaluate the consistency of participants' test scores between first and second tests within a month interval based on the raw scores and categorized scores. The results varied from poor to excellent test-retest reliability indexes based on total workout score,

error rate, deviation rate, range, peak performance score, and the graph types. Excellent reliability values were found on the total workout and error rate scores especially when the raw scores were grouped into three categories, poor, average, and good. The result indicated that the F-EAST is consistent to measure a person's work capacity and thoroughness over time. Meanwhile, based on the emotional aspect scores (deviation rate, range and peak performance scores), the degree of ICC index ranged from medium to low, indicating that F-EAST is slightly less consistent in measuring a person's emotional stability, achievement motive, and ambition. The finding is consistent with Peretti, P. O., & O'Connor's (1989) study, which found that a person's level of emotional stability tends to fluctuate and is influenced by his circumstances. In addition, the F-EAST's graph type showed good reliability indicating that the working style of a persons' are relatively stable over time.

Study 2 reported the validity evidence based on the correlation with other variables namely Kraepelin and Pauli test as the standardized test to measure work attitude. Overall, the results demonstrated that the F-EAST has good validity based on the degree of correlations between similar or identical aspects measured by the Pauli test and the Kraepelin test. First, the correlation between the F-EAST and Pauli test ranged between 0.113 and 0.635. Specifically, the moderate positive correlations were found on total workout score, error rate deviation rate and graph types indicating that the F-EAST can measure the same aspects as Pauli's in workload capacity, thoroughness, emotional stability, and working style. The F-EAST is less effective in measuring achievement driving, and ambition planning in Pauli. The ineffectiveness might be due to the processing times for F-EAST and Pauli differing by 20 minutes and 60 minutes, respectively. Additionally, F-EAST and Pauli have a stronger correlation coefficient in regards to workload capacity and error rate than than other tests in another study with the same objective, namely Qutest (Febriawan et al., 2022).

Second, the F-EAST also has good validity evidence based on correlation with Kraepelin. The correlation between F-EAST and Kraepelin was between 0.054 and 0.578 in similar or identical aspects. The F-EAST also showed moderate positive correlation workload capacity, wrong answer, and working style indicating the same construct with Kraepelin test. Meanwhile, Emotional stability F-EAST is less effective in measuring kraepelin's working curve even though it measures the same about stability. The ineffectiveness in the emotional stability measurement might be due to the

differences of evaluation between Kraepelin's working curve and the F-EAST deviation rate. The F-EAST workmanship average graph based on the participants who are overly steady will be labeled as poor; however, this is not the case with Kraepelin.

Tabel 6
Intercorrelation matrix aspects F-EAST, Pauli, and Kraepelin

Aspects	F-EAST						Pauli						Kraepelin			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
F-EAST																
1. Workload capacity	1															
2. Error rate (%)	-0.039	1														
3. Deviation rate (%)	-0.627**	0.108	1													
4. Range	0.068	-0.026	0.338	1												
5. Peak performance	0.107	0.041	-0.273	-0.041	1											
6. Graph Type	-0.064	-0.218	-0.072	0.098	0.192	1										
Pauli																
7. Workload capacity	0.635**	-0.179	-0.628**	-0.109	0.121	0.034	1									
8. Error rate (%)	0.011	0.301*	-0.011	-0.039	0.114	-0.076	-0.076	1								
9. Deviation rate (%)	-0.159	0.175	0.330*	0.342*	-0.182	-0.367*	-0.367*	0.065	1							
10. Range	0.209	-0.064	-0.187	0.156	-0.130	0.222	0.222	0.121	0.468**	1						
11. Peak performance	-0.070	0.057	-0.017	0.073	0.113	-0.100	-0.100	-0.075	-0.111	-0.018	1					
12. Graph Type	-0.104	0.009	0.021	0.203	0.108	0.557**	-0.028	0.003	0.012	0.171	0.367*	1				
Kraepelin																
13. Workload capacity	0.576**	-0.049	-0.556**	0.019	-0.140	0.047	0.820**	0.083	-0.173	0.414**	0.051	0.041	1			
14. Wrong Answer	-0.098	0.578**	0.037	-0.134	-0.325	-0.043	0.020	0.188	-0.072	0.025	-0.063	-0.143	0.041	1		
15. Working Curve	0.229	-0.087	0.054	0.369*	-0.232	-0.041	0.085	-0.010	0.232	0.412**	0.091	0.066	0.197	-0.010	1	

Aspects	F-EAST						Pauli						Kraepelin			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16. Working Style	0.000	-0.344*	0.024	0.094	0.240	0.310*	-0.014	-0.077	0.112	-0.100	-0.003	-0.017	-0.167	-0.300*	-0.015	1
Mean	884	2.00	6.3	1.92	1.60	3.34	2716	0.20	4.50	1.53	1.57	3.51	22.75	14.24	12.06	-1.49
SD	205	1.40	2.3	0.27	0.82	1.36	650	0.40	1.70	0.77	0.82	1.42	5.21	16.63	2.88	27.87

*p<0.05, **p<0.01

Regarding the limitation, this study reveals that there are two similar aspects between F-EAST and Pauli that are not correlated, namely Range and Peak Performance. The length of the two test administrations is their primary distinction (20 minutes for F-EAST and one hour for Pauli). This enables Range and Peak Performance of each test to address different constructs. In contrast, neither of these aspects are measured by the Kraeplin test, which has the same administration length as F-EAST. Another potential limitation is the small number of samples which may affect the generalization of this study.

Conclusion

To conclude, the current study provides that the F-EAST are valid and reliable for computer based tests as general and to measure work aptitude by converting the procedure Pauli's and Kraepelin's paper and pencil test into computer administration. F-EAST also has good report psychometric properties of the computer-based test relating to reliability and validity evidence.

Acknowledgment

The study was funded by Faxtor Indonesia in 2022 as a research project.

Conflict of Interest

All authors declare that they have no conflicts of interest.

References

Adedoyin, O. B., & Soykan, E. (2020). Covid-19 pandemic and online learning: the challenges and opportunities. *Interactive Learning Environments*, 1–13. https://doi.org/10.1080/10494820.2020.1813180.

Arnold, W. (2013). *Der Pauli-Test: Anweisung zur sachgemäßen Durchführung, Auswertung und Anwendung des Kraepelinschen Arbeitsversuches*. Springer-Verlag.

Borualogo, I. S., & Rabayani, W. M. (2009). Studi perbandingan profil pauli antara mahasiswa berprestasi tinggi dan mahasiswa berprestasi rendah angkatan 2004 jurusan teknik pengecoran logam di Politeknik Manufaktur Negeri Bandung. *Schema: Journal of Psychological Research*, 1(2), 28-37.

Febriawan, I. M., Fauzan, A. A., Adelina, C., Afifah, H. N., Pari, R., & Januarsjaf, A. (2022). Qutest construction and psychometric evaluation as test of attention and willpower for employee selection screening. *JP3I (Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia)*, 11(1), 55–63.

https://doi.org/10.15408/jp3i.v11i1.22352
Hidayat, R. (2016). Menentukan promosi jabatan karyawan dengan menggunakan metode profile matching dan metode promethee. *IJSE-Indones. J. Softw. Eng*, 2(1), 57-65.
ITC. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6(2), 143–171. https://doi.org/10.1207/s15327574ijt0602_4
Javaid, M., Haleem, A., Vaishya, R., Bahl, S., Suman, R., & Vaish, A. (2020). Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4), 419–422. https://doi.org/https://doi.org/10.1016/j.dsx.2020.04.032
Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological testing: Principles, applications, and issues*. Cengage Learning.
Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012
Kraepelin, E. (1922). Gedanken über die arbeitskurve. *Psychologische Arbeiten*, 7, 535–547.
Kozak, M. (2009). What is strong correlation?. *Teaching Statistics*, 31, 85–86. https://doi.org/10.1111/j.1467-9639.2009.00387.x
Lee, K. M., Lee, J., Chung, C. Y., Ahn, S., Sung, K. H., Kim, T. W., ... & Park, M. S. (2012). Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. *Clinics in orthopedic surgery*, 4(2), 149-155.
Li, G. Y., Ueki, H., Kawashima, T., Sugataka, K., Muraoka, T., & Yamada, S. (2004). Involvement of the noradrenergic system in performance on a continuous task requiring effortful attention. *Neuropsychobiology*, 50(4), 336–340.
Lievens, F. (2006). The ITC guidelines on computer-based and internet-delivered testing: where do we go from here?. *International Journal of Testing*, 6(2), 189–194.
Mikicin, M., Orzechowski, G., Jurewicz, K., Paluch, K., Kowalczyk, M., & Wróbel, A. (2015). Brain-training for physical performance: a study of EEG-neurofeedback and alpha relaxation training in athletes. *Acta Neurobiologiae Experimentalis*, 75(4), 434-445.
Nada, Q., Arhami, M., & Simbolon, Z. K. (2022). Pengukuran aptitude dengan uji kraepelin menggunakan metode Linear Congruential Method (LCM). *Jurnal Teknologi*, 22(1), 1–9.
Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: new problems, old issues. *American Psychologist*, 59(3), 150.
Paramasivam, N. (2021). Mass hysteria phenomenon in high schools: A cross-sectional study of mental stress among Malay students by using the Kraepelin Test: Fenomena histeria massa di sekolah menengah: kajian lintas bahagian mengenai tekanan mental di kalangan pelajar Melayu dengan menggunakan ujian kraepelin. *The Sultan Alauddin Sulaiman Shah Journal (JSASS)*, 8(1), 84-89.
Pane, H., Fauziah, & Nurhayati. (2020). rancang bangun aplikasi kraepelin test berbasis web menggunakan metode bubble sort. *JOINTECS (Journal of Information Technology and Computer Science)*, 5(1), 41–48.
Peretti, P. O., & O'Connor, P. (1989). Effects of incongruence between the perceived self and the ideal self on emotional stability of stripteasers. *Social Behavior and Personality: an international journal*, 17(1), 81-92.
Pradipta, R. T. D., & Hidayat, R. (2019). *Konvergensi dan divergensi tes kraepelin dan tes pauli pada komunitas psikolog* (Doctoral dissertation, Universitas Gadjah Mada).
Prasetiya, K. A., Witanti, W., Ilyas, R., Militer, P., & Jenderal, A. (2017). Perancangan sistem pendukung

keputusan penempatan kecabangan TNI AD menggunakan metode analytic hierarchy process dan simple additive weighting. *No. September*, 1, 21–26.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: applications to practice (Vol. 892)*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Reuning, H. (1983). Continuous work tests: Their scope in cross-cultural contexts. In S. H. Irvine & J. W. Berry (Eds.), *Human Assessment and Cultural Factors* (pp. 303–318). Springer US. https://doi.org/10.1007/978-1-4899-2151-2_21

Sugimoto, K., Kanai, A., & Shoji, N. (2009). The effectiveness of the Uchida-Kraepelin test for psychological stress: an analysis of plasma and salivary stress substances. *BioPsychoSocial Medicine*, 3(1), 5. https://doi.org/10.1186/1751-0759-3-5

Thornton, G. C. III, & Kedharnath, U. (2013). Work sample tests. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 533–550). American Psychological Association. https://doi.org/10.1037/14047-029