
Indonesian Version of WHO-5 Well-being Index Amidst COVID-19 Pandemic Settings: Scale Validation and Standardisation

Rena Latifa¹, Kay Chang², Andrian Liem³, Natris Idriyani¹, Adiyo Roebiyanto⁴, Muthia Rahmah¹,
Valendra Granitha Shandika Puri¹

Faculty of Psychology, UIN Syarif Hidayatullah Jakarta, Indonesia¹

Psychology of Department, University of Macau, Macau (SAR) China²

Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Malaysia³

⁶Universitas Al-Azhar Indonesia, Indonesia⁴

rena.latifa@uinjkt.ac.id

Abstract

The 5-item World Health Organization Well-Being Index (WHO-5) is a frequently used brief standard measure in large-scale clinical studies. However, no research specifically on the validity test of WHO-5 was found in Indonesia before the pandemic. This study aims to test the validity of the Indonesian version of the 5-item WHO-5 in the COVID-19 pandemic setting. The online survey was used to collect data from February 2021 to September 2021. 1,084 Indonesians who were directly or indirectly affected by COVID-19 completed the survey. The scale was validated with Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT). The CFA and IRT analysis showed that the WHO-5 is valid. Thus, the WHO-5 is a short questionnaire that consists of five non-invasive questions adequate to measure the psychological well-being of the Indonesian sample, especially during the COVID-19 Pandemic, and could be used in other pandemics or crises.

Keywords: WHO-5, well-being scale, COVID-19 pandemic, scale validation, Indonesian version

Abstrak

Indeks Kesejahteraan World Health Organization (WHO-5) yang terdiri dari 5 item adalah ukuran standar singkat yang sering digunakan dalam studi klinis skala besar. Namun, belum ditemukan penelitian khusus uji validitas WHO-5 di Indonesia sebelum pandemi. Penelitian ini bertujuan untuk menguji validitas 5 item WHO-5 versi bahasa Indonesia pada situasi pandemi COVID-19. Survei online ini digunakan untuk mengumpulkan data pada bulan Februari-2021 hingga September 2021. Sebanyak 1.084 masyarakat Indonesia yang terdampak langsung atau tidak langsung oleh COVID-19 menyelesaikan survei tersebut. Skala tersebut divalidasi dengan Confirmatory Factor Analysis (CFA) dan Item Response Theory (IRT). Analisis CFA dan IRT menunjukkan WHO-5 valid. Oleh karena itu, WHO-5 merupakan kuesioner singkat yang terdiri dari lima pertanyaan non-invasif yang cukup untuk mengukur kesejahteraan psikologis sampel Indonesia, terutama selama Pandemi COVID-19, dan dapat digunakan pada pandemi atau krisis lainnya.

Kata kunci: WHO-5, skala kesejahteraan, pandemi COVID-19, validasi skala, versi Indonesia

Introduction

The increasing number of infections and death rates due to COVID-19 can cause collective anxiety about COVID-19 infection (Bento et al., 2020). Increased infection rates and their association with deteriorating mood and emotions result in decreased psychological well-being (Bathina et al., 2021). The World Health Organization (WHO) considers psychological well-being one of the essential aspects of health and an essential dimension of perceived quality of life (McDowell, 2010). Thus, it is also essential to understand how to measure psychological well-being using a valid and reliable instrument in the population to plan an effective intervention. One of the most used instruments to measure psychological well-being is the 5-item WHO Well-Being Index (WHO-5).

The WHO-5 was first introduced at the WHO meeting in Stockholm in February 1998 for the project on well-being measurement, and the WHO Regional Office initiated the translation into several languages (Staehr, 1998). The WHO-5 is derived from the WHO-10, first derived from a 28-item rating scale used in a WHO study in eight European countries (Warr et al., 1985; Staehr, 1989). The 28 items in the original scale are adapted from the Zung scales (for depression, distress, and anxiety), the General Health Questionnaire, and the Psychological General Well-Being Scale (Paisley, 2000). Thus, both the 28-item scale and the WHO-10 consists of item that measures distress both in favourable and unfavourable forms. In contrast, since the WHO now considers well-being in a positive light and sense, the WHO-5 only includes positively phrased items (Bech, 1999).

This WHO-5 well-being index is translated into at least 30 languages, for example, Arabic, Chinese, Filipino, etc. (Topp et al., 2015). Despite being translated into multiple languages, not all of these translation versions have been validated, including the Indonesian version with a large sample of Indonesian. This study aims to provide a validity test of WHO-5 from the Indonesian language and its psychometric properties.

Psychological Well-being Amidst COVID-19

The devastation caused by the COVID-19 pandemic affects the growth rate of mortality, rate of infection, and other health hazard issues and the decline of psychological well-being among people all around the world, especially Indonesian people (Bathina et al., 2021). Psychological well-being is related to physical, mental, and socio-cultural aspects and spirituality (Božek et al., 2020; Kumar, 2020). The pandemic might not have a significant impact if every individual has immunity both in physical and mental power. Therefore, it is vital to create an order in which society can achieve good psychological well-being to be resilient in the face of a pandemic.

Psychological well-being emphasizes how and why a person lives life in positive ways, including cognitive judgments and affective reactions. It includes studies that have used various aspects such as happiness, satisfaction, morale, and positive influence (Diener, 2009). The main goals of the state, society, and people are to understand and accept that human well-being is a fundamental, foundational, basic, and indispensable condition of a healthy society and its successful development and prosperity.

Ryff and Keyes (1995) define psychological well-being as an encouragement to explore the individual's potential as a whole. This can cause a person to become resigned to a situation that decreases the psychological well-being of individuals or try to improve living conditions that will increase the psychological well-being of the individual. Individuals who have high psychological well-being are individuals who are satisfied with their lives, are in positive emotional states, can go through bad experiences that can cause negative emotions, have positive relationships with others and can determine their destiny without depending on others, control environmental conditions, have a clear purpose in life and can self-develop (Ryff, 1989).

The decline of psychological well-being that is impacted mainly by the COVID-19 pandemic has become one of the priorities of the issues that need solving. The first step is to understand the accurate level of psychological well-being by measuring it using specific psychological tools. One of the most accessible screening tools used to measure psychological well-being using simple, non-invasive items is the 5-item World Health Organization Well-Being Index, usually known as WHO-5.

WHO-5, Adaptation Indonesian Version

Among various assessments of psychological well-being, the 5-item WHO-5 is one of the most employed measuring instruments of psychological well-being across the world. This unidimensional WHO-5 allows respondents to complete the questionnaire in under one minute. This instrument measures the psychological well-being of an individual in the past 14 days or two weeks.

The WHO-5 is generally used to assess clinical outcomes in clinical trials and adequately measure responsiveness/sensitivity to treatment (Topp et al., 2015). The WHO-5 was initially developed to measure well-being. Still, several studies also use it as a screening tool for depression because the items represent aspects closely related to depression, such as lack of positive emotions, interests, and energy (Krieger et al., 2014). The WHO-5 is also used mainly in multinational studies in various study fields, such as health fields related to diabetes (Nicolucci et al., 2013) and quality of life in the real coronavirus crisis (Ahrendt et al., 2020).

Prior Validity and Reliability Test of WHO-5

The WHO-5 was generally used in public health studies, such as research on diabetes, heart diseases, and depression before the COVID-19 pandemic (Hindoro et al., 2018; Larasati & Kristina, 2020; Soewondo et al., 2010). However, no research specifically on the validity test of WHO-5 was found in Indonesia before the pandemic. The Indonesian researchers conducted studies using the WHO-5 as one of the scales that measured well-being with other psychological and non-psychological variables. Most prior studies using WHO-5 as a measuring instrument only examine the instrument's reliability by calculating Cronbach's alpha and one-dimensionality using exploratory factor analysis or principal component analysis. Others also test the external criterion. However, the methods mentioned above are not sufficient to test the validity and reliability of WHO-5. Validity is achieved when the instrument truly measures the intended construct (Comrey, 1973; Jöreskog, 1969). One of the best, more informative construct validity tests is Confirmatory Factor Analysis (CFA); thus, in this study, CFA is used as a construct validity test of WHO-5. Furthermore, the Item Response Theory is used to calibrate and evaluate items in the questionnaire.

Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis represents the relationship between observed variables and the latent variables (Brown, 2015). Based on Brown (2015), CFA can be used to analyse the construct validity of a variable. It provides evidence of the validity of theoretical constructs.

Item Response Theory Model

The Item Response Theory (IRT) is an essential psychometric property for item analysis (Ostini et al., 2015). The WHO-5 employs scoring from 0 (absence of well-being) to 25 (maximal well-being) (WHO, 1998). The scoring requires an equal discrimination parameter (IRT 1 parameter). However, the total score may not be enough to show the extent of one's psychological well-being. Thus, the IRT model may be more sufficient.

Standard Setting with Rasch Model

The standard setting uses a systematic approach that identifies a common agreement among experts or raters to cut score(s) or threshold(s) for a certain level of proficiency and/or traits (MacCann & Gordon, 2019). The author uses item mapping to decide the cut scores according to the item difficulties or item locations (Wang, 2003; MacCann & Gordon, 2019). The next step is the author uses a convenient method (World Health Organization, 2009; Bonacchi et al., 2021) to classify people into two categories as follows: if the logit scale falls less than 0.00, then it would indicate an absence of well-being ($<.00$ logit scale “absence of well-being”); and if the logit scale falls more than equal .00, it indicates high well-being ($\geq .00$ logit scale “high well-being”).

Methods

Survey Design and Participants

This study was part of a more extensive study on screening instrument development, for which ethics approval was provided by the University of Macau (Reference numbers: SSHRE20-APP020-FSS, EA210291). The online survey was distributed from February to September 2021, when the COVID-19 pandemic was at its peak using the Survey Monkey platform. The completion time for the WHO-5 part was approximately one minute. Non-purposive sampling method was used by distributing the link to the survey to the Indonesian population who were at least 18 years old and affected directly or indirectly by COVID-19. The survey assessed and quantified the psychological well-being of participants, characterised by (1) being exposed to the negative impact of COVID-19 directly and (2) being exposed to the negative impact of COVID-19 indirectly through family, friends, and the surrounding environment. A total of 1,084 participants completed the WHO-5 scale. Thus, no data was excluded for being incomplete or missing. These 1,084 respondents were predominantly females (72.1%) with ages ranging from 18 to 64 years ($M = 24.1$, $SD = 8.4$) and located in 26 of 34 provinces in Indonesia (see Table 2 for further sample characteristics detail).

The demographic information presented in this study is gender and age. The gender is coded into three codes: 0 = self-described, 1 = female, and 2 = male. The age is coded into three codes: 1 = early adulthood (18-34), 2 = mid-adulthood (35-44), and 3 = late adulthood (>45). The characteristics of the sample are shown in Table 2 below. The majority of respondents are female (72.1%), and the minority chose to self-describe (1.9%), while males made up 25.9% of the sample. The majority of respondents are people in their early adulthood (88.5%), and a minority of respondents are in late adulthood (3.5%), while people in mid-adulthood made up 8% of the sample.

Table 1. Participants

		Frequency	Per cent	Valid Percent	Cumulative Percent
Gender	Self-described	21	1.9	1.9	1.9
	Female	782	72.1	72.1	74.1
	Male	281	25.9	25.9	100.0
	Total	1,084	100.0	100.0	
Age	Early Adulthood (18-34)	959	88.5	88.5	88.5
	Mid Adulthood (35-44)	87	8.0	8.0	96.5
	Late adulthood (> 45)	38	3.5	3.5	100.0
	Total	1,084	100.0	100.0	

Source: Authors' primary data (2021)

Furthermore, the raw data is analysed to obtain the mean and standard deviation for age, as shown in Table 2 below:

Table 2. Mean and Standard Deviation

	N	Minimum	Maximum	Mean	Std. Deviation
Age	1,084	16	71	24.1448	8.370
Total	1,084				

The minimum age of respondents is 16 years old, and the maximum age is 71 years old. The mean age shows a value of 24.15 with a standard deviation of 8.37. In other studies, during the COVID-19 pandemic, the mean age varied. The well-being index among nurses in Spain, Chile, and Norway ranges from 36 to 48 years old, with a mean age of 39.3 years and an SD of 12.1 (Lara-Cabrera et al., 2022). A Nigerian study showed that doctors and nurses with a mean age of 39.85 and SD 8.49 have ages ranging from 21 to 63 years (Seb-Akahomen et al., 2021). Moreover, in the Austrian survey, the mean age is 40.22 years with SD 11.60, and the range age is from 18 to 79 years old (Simon et al., 2021).

Instruments

The English version of WHO-5 was translated into the Indonesian language (Appendix A) using forward and backward translation techniques to ensure the clarity and consistency of each item's meaning (World Health Organization, 2009). The participants were asked to rate how often the statement applies to them considering the last 14 days by choosing a Likert scale ranging including 1 (at no time), 2 (some of the time), 3 (less than half the time), 4 (more than half the time), 5 (most of the time), and 6 (all of the time).

Procedure

The researchers translated the WHO-5 into the Indonesian language as this study is aimed at the Indonesian sample. The process of translation was conducted through forward translation and backward translation to ensure the accuracy of the translated items. First, translation is done from the English version to the Indonesian version by an expert in psychology who has the English language ability to translate. Then, the Indonesian draft scale is translated into English by an expert with the same criteria. A back-translation draft is discussed on the equality of meaning per item by foreign experts in psychology and the use of English as a daily language. The Indonesian draft is checked again for its appropriateness in meaning.

Data Analysis

The collected data were analysed with Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT). A CFA was conducted to examine the uni-dimensionality of the WHO-5 using Lisrel software. A CFA was conducted through several steps (Comrey, 1973; Jöreskog, 1969):

1. Determining the operational definition of the measured construct. To measure the construct, an item (stimulus) is needed as the indicator;
2. Formulating a hypothesis where all items are valid in measuring the construct. In other words, building Ha's hypothesis based on uni-dimensionality where there is only one factor that is measured by all items;
3. The correlation matrix between items, known as the S matrix, is calculated, which is used to estimate the correlation matrix;
4. The construct validity is tested through a hypothesis test by checking the Chi-Square test and RMSEA. The Chi-Square showing an insignificant value ($p > .05$) means that the null hypothesis (H_0) is not rejected. This indicates that the theory that says all items only measure one construct alone is proved to be in accordance with the data. Meanwhile, RMSEA showing a significant value ($< .05$) indicates

that the null hypothesis (H0) is not rejected. This indicates that the theory that says all items only measure one construct alone is proved to be in accordance with the data;

5. If a unidimensional (one-factor) model is proven to fit the data, then the selection or evaluation of items can be made using three criteria:
 - a. Items with insignificant factor loading ($>.05$) are dropped as they do not provide statistically meaningful information;

Items with a negative factor loading coefficient are also dropped as they measure the opposite of the defined construct. Researchers should check which item statement is unfavourable; the score is reversed. This applies especially to items with no right or wrong answer (e.g., personality, motivation, perception, etc.);
 - b. Items can also be eliminated if the residual (measurement error) correlates with more than three items; this indicates that the item measures construct other than the construct to be measured.

Uni-dimensionality indicates that the measuring instrument only measures one construct, in this case, psychological well-being. After the construct validity test shows a fit unidimensional model by checking the RMSEA as an index of fit, IRT is conducted as the construction of measurement instruments, linking and equating measurements, and evaluation of test bias and differential item functioning. IRT is an essential aspect of instrument measurement. Therefore, this study employed IRT on WHO-5. The IRT is done using MPlus software.

Results and Discussion

Results

Researchers present an additional means of comparison for both demographic variables, namely gender and age. Mean comparison is intended for checking the level of psychological well-being across gender and age. This data comparison is additional information on the difference in psychological well-being between female, male, and self-described individuals and people in their early adulthood, mid-adulthood, and late adulthood. The means comparison is shown in Table 3 below:

Table 3. Compared Means

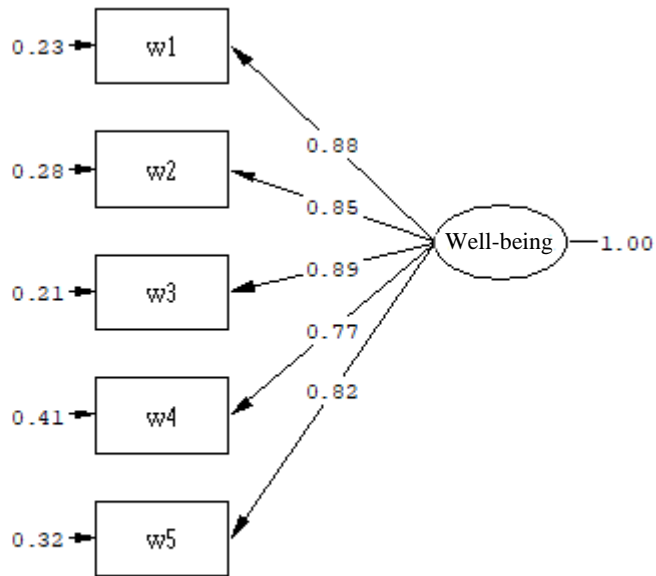
	Mean	N	Std. Deviation
Gender			
Others	15.48	21	5.81
Female	15.82	782	4.93
Male	18.03	281	5.64
Total	16.39	1084	5.22
Age			
Early Adulthood (18-34)	15.97	959	5.05
Mid Adulthood (35-44)	19.61	87	5.06
Late adulthood (> 45)	20.89	38	5.71
Total	16.40	1084	5.21

Table 4. Participants' Demographic of Provinces

Province	Frequency	Percent
Bali	6	0.55%
Banten	306	28.23%
Bengkulu	7	0.65%
DKI Jakarta	300	27.68%
Jambi	1	0.09%
West Java	274	25.28%
Central Java	34	3.14%
East Java	42	3.87%
West Kalimantan	2	0.18%
South Kalimantan	3	0.28%
Central Kalimantan	1	0.09%
East Kalimantan	4	0.37%
Bangka Belitung	4	0.37%
Lampung	3	0.28%
West Nusa Tenggara	4	0.37%
East Nusa Tenggara	4	0.37%
Riau	6	0.55%
Riau Islands	4	0.37%
West Sulawesi	2	0.18%
South Sulawesi	8	0.74%
Central Sulawesi	3	0.28%
North Sulawesi	2	0.18%
West Sumatera	13	1.20%
South Sumatera	11	1.01%
North Sumatera	9	0.83%
Yogyakarta	27	2.49%
Outside Indonesia	4	0.37%
Total	1,084	100%

The first model of Confirmatory Factor Analysis (CFA) of the 5-item World Health Organization Well-Being Index (WHO-5) is shown in Figure 1. The results in Figure 1 shows the Chi-Square = 91.58, $df = 5$, $p < .001$, and RMSEA = 0.126. The fit indexes indicate that the model does not fit with the data (Jöreskog et al., 2016). Therefore, the model is modified by allowing the item's measurement errors to correlate or the theta-delta to correlate in order to reach the fit model. Based on the modification indices in Lisrel, the modification made is as follows

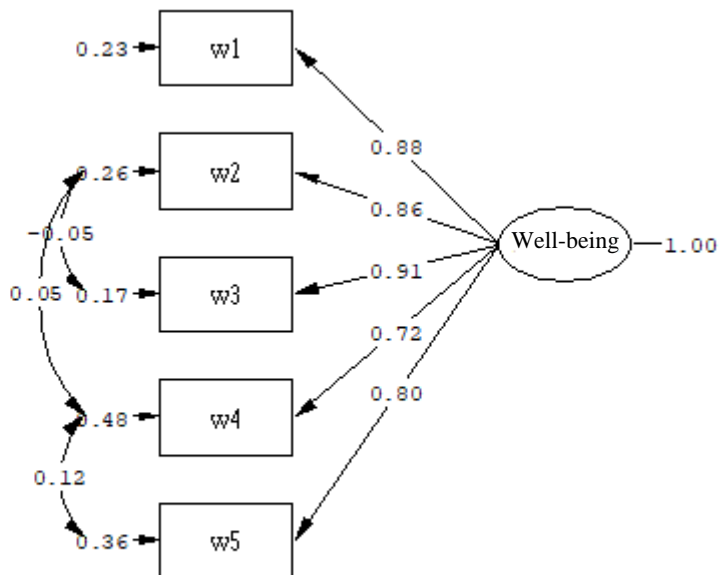
Figure 2.



Chi-Square=91.58, df=5, P-value=0.00000, RMSEA=0.126

Figure 1. Path Diagram of WHO-5 without Modification

Figure 2 shows the Chi-Square = 0.32, df = 2, *p*-value = 0.85218, and RMSEA = 0.000. The fit indexes indicate that the model is fitted with the data (*p*-value of Chi-Square > 0.05 and RMSEA < 0.05) (Jöreskog et al., 2016).



Chi-Square=0.32, df=2, P-value=0.85218, RMSEA=0.000

Figure 2. Path diagram of WHO-5 with Modification

The comparison between the first model before modification and the model after modification is shown in Table 5 below:

Table 5. Model Comparison

Fit Index	Before	After	Notes
Chi-Square	91.58, df = 5, p-value = 0.000	0.32, df = 2, p-value = 0.85218	Modifications = 3 times
RMSEA	0.126	0.000	
		p-value (RMSEA < 0.05) = 0.99	

The results of Confirmatory Factor Analysis obtained model fit after three modifications by allowing measurement error between items to correlate. Furthermore, researchers used Phi standardisation for standardising techniques. After obtaining the fit model for WHO-5, the items are tested to determine the significance and validity to check whether some items must be dropped or not.

Table 6. Significance Test

Item	Factor Loading	t-value	p-value	Validity
w1	0.88	35.54	0.05	Valid
w2	0.86	33.71	0.05	Valid
w3	0.91	37.50	0.05	Valid
w4	0.72	26.52	0.05	Valid
w5	0.80	30.91	0.05	Valid

Table 6 shows the result of the significance test to consider item dropping. There are three main requirements to declare if an item is valid: (1) favourable factor loading, (2) t-value > 1.96, and (3) less than three modifications for an item. According to that table, the significance test results of all items of WHO-5 show favourable factor loading or lambda for all items; this indicates that all items measure the intended construct. Furthermore, all t-values are more than 1.96; all p-values are 0.05, meaning that all items are significant in measuring psychological well-being. The final verdict is that all items are valid based on evaluating factor loading, t-value, and p-value of each item constructing the 5-item World Health Organization Well-Being Index (WHO-5). In conclusion, the construct validity test using Confirmatory Factor Analysis shows a fit unidimensional model, meaning that all items of WHO-5 only measure one construct: psychological well-being.

Item Response Theory

Item Response Theory analysis shows the result of the p-value of the Chi-Square test of the model for the binary and ordered categorical (ordinal) outcomes to be significant (p-value > 0.05), in which case the p-value of Chi-Square is 1.000. Thus, based on the p-value of the Chi-Square test of the model for the binary and ordinal outcomes, the model is fit. The Loglikelihood for the model is -6,770.245, which shows a fit model because the Loglikelihood is less than 10,000.

Furthermore, the IRT parameterisation results show two pieces of information, namely item discriminations and item locations. Table 7 shows the results of item discriminations, and Table 8 shows the item locations.

Table 7. Item Discriminations

Item	Estimate	SE.	Est./SE.	Two-Tailed p-value
W1	2.547	0.222	11.456	0.000
W2	2.165	0.199	10.860	0.000
W3	2.675	0.231	11.566	0.000
W4	1.313	0.105	12.543	0.000
W5	1.680	0.139	12.055	0.000

Item discrimination is the ability of each item to discriminate or differentiate among respondents based on how high or low they are in the context of the construct measured, in this case, psychological well-being. Table 7 shows that the item discrimination results are all significant (p -value < 0.05), with all estimates being optimistic, further supporting the validity of each item. Furthermore, all the discrimination indexes are above .60. In conclusion, all items have good item discrimination.

Table 8. Item Location

Item	Estimate	SE.	Est./SE.	Two-Tailed p-value
W1	-0.072	0.043	-1.659	0.097
W2	0.063	0.044	1.424	0.155
W3	0.187	0.040	4.673	0.000
W4	0.272	0.048	5.713	0.000
W5	0.162	0.043	3.728	0.000

Item location is also known as item difficulty, which estimates how high they must possess a particular ability or psychological construct, in this case, psychological well-being, to pick the correct answer. Table 8 shows that all estimates of item difficulty are typical in which all values are ideal as they should range from -2 to 2. In this study, all range from -0.072 to 0.272. In conclusion, item W1 is the most accessible item with an estimate of -0.072, and W4 is the most complex item with an estimate of 0.272.

To further analyse items using Item Response Theory, researchers present the Item Characteristic Curve (ICC) for all items with the range of -3 SD to +3 SD. The item characteristic curve (ICC) is shown in Figure 3. According to the diagram, the further to the left the items are, the easier they are, and the further to the right, the more difficult they are (Muthén & Muthén, n.d.). According to Figure 3 The items in WHO-5 tend to be difficult as all of the curves are on the right side of the diagram. This further shows that most respondents have low psychological well-being amidst the pandemic.

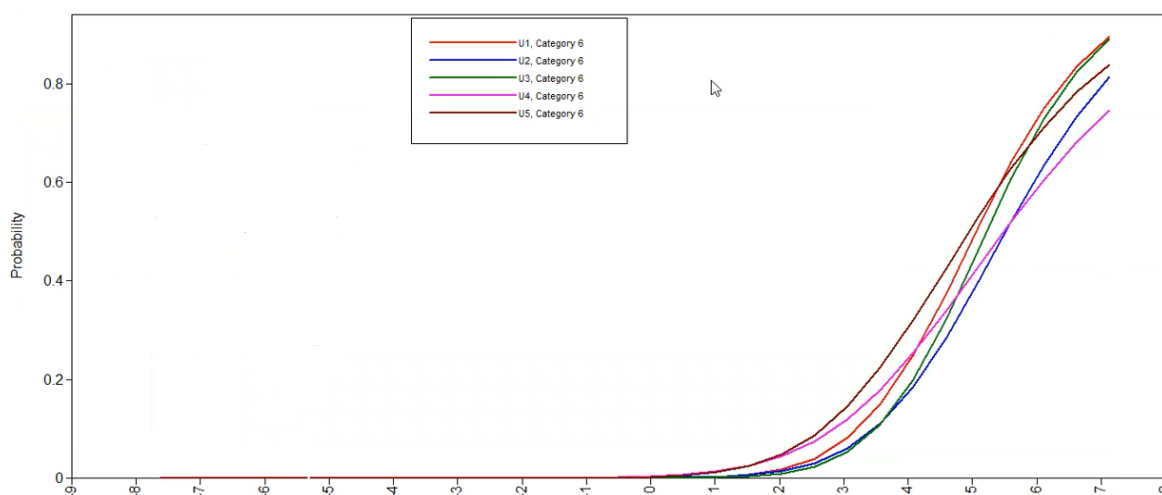


Figure 3. Item Characteristic Curve (ICC)

Table 3 shows that males have the highest mean of psychological well-being, meaning that, on average, males have higher psychological well-being compared to females and self-described people. Self-described people have the lowest mean of psychological well-being, indicating that, on average, they have lower psychological well-being compared to females and males.

Late adulthood has the highest mean of psychological well-being in the age group, meaning that, on average, people in late adulthood have higher psychological well-being than people in their early and middle adulthood (Table 3). Early adulthood has the lowest mean of psychological well-being. On average, people in early adulthood have lower psychological well-being than those in mid and late adulthood.

Standard Setting

As mentioned earlier, the author uses Rasch's standard setting to classify people on the scale of well-being. The results of the standard setting can be seen in Table 8 below.

Table 9. Rasch Standard Setting

Score	Measure (logit scale)	Frequency	Percentage (%)	Category	Frequency	Percentage (%)			
0	-8.3	4	0.4						
1	-6.93	2	0.2						
2	-5.95	13	1.2						
3	-5.17	14	1.3						
4	-4.37	37	3.4						
5	-3.46	93	8.6						
6	-2.65	64	5.9	Absence of Well-Being	620	57.3			
7	-2.05	70	6.5						
8	-1.6	61	5.6						
9	-1.22	81	7.5						
10	-0.86	75	6.9						
11	-0.52	67	6.2						
12	-0.17	39	3.6						
13	0.22	86	7.9						
14	0.65	66	6.1						
15	1.14	89	8.2						
16	1.66	41	3.8						
17	2.17	36	3.3						
18	2.66	41	3.8						
19	3.14	26	2.4	High Well-Being	464	42.7			
20	3.61	25	2.3						
21	4.1	13	1.2						
22	4.63	13	1.2						
23	5.25	9	0.8						
24	6.13	3	0.3						
25	7.46	16	1.5						

According to Rasch's analysis, the logit scale of 0.00 was approximately equal to 13 of the total score. Therefore, those obtained scores below 13 can be as absence of well-being and for those more than equal to 13 belongs to high well-being. Using these categorizations, there are 620 (57.3%) participants belongs to absence of well-being, and 464 (42.7%) participants fall into high well-being.

Discussion

Psychological well-being emphasizes how and why a person lives life in positive ways, including cognitive judgments and affective reactions. It includes studies that have used various aspects such as happiness, satisfaction, morale, and positive influence (Diener, 2009). The main goals of the state, society, and people are to understand and accept that human well-being is fundamental, foundations, basic premises, and indispensable conditions of a healthy society and its successful development and prosperity (Alatartseva & Barysheva, 2015).

The importance of psychological well-being, especially amidst the COVID-19 pandemic, encourages studies on developing the measuring instrument of psychological well-being. Thus, this study aims to validate and test the WHO-5 using Item Response Theory (IRT) as the construction of measurement instruments, linking and equating measurements, and evaluation of test bias and differential item functioning.

First, the Confirmatory Factor Analysis (CFA) was administered. The results fit the model after three modifications by allowing the measurement error of items to correlate with each other (allowing for theta-delta to correlate using LISREL) (Chi-Square = 0.32, $df = 2$, $p = .85$ and RMSEA = 0.000). The significance test of each item also shows that all items are valid in measuring the construct of psychological well-being. The validity is shown through positive factor loadings and t -values > 1.96 which consequently shows p -value = 0.05. The results concluded that the WHO-5, a unidimensional instrument, consists of five useful items measuring psychological well-being. Prior studies among hypertension patients and diabetes mellitus patients also found similar validity results stating that the WHO-5 is a valid unidimensional scale (Hindoro et al., 2018; Larasati & Kristina, 2020; Soewondo et al., 2010).

Second, the Item Response Theory (IRT) analysis shows that item parameterization resulted in a fit model for item discrimination and difficulty. All estimates for item discriminations are optimistic. All items can differentiate among respondents based on how high or low they are in the context of the construct measured, in this case, psychological well-being. Item W3 has the strongest discrimination among the other items, with the highest estimate of 2.68. Item W1 is the most accessible item with an estimate of -0.07, and W4 is the most difficult to answer item with an estimate of 0.272, meaning the easier the items are, the more likely respondents will choose the correct answer and vice versa. There is no right or wrong answer in the WHO-5 as the scale itself does not measure ability; thus, it is a non-ability scale measuring well-being. All items have a standard range of estimates. Furthermore, Item Characteristic Curve (ICC) shows that all items of WHO-5 tend to be further on the right side of the diagram, meaning that all items tend to be complicated. This further shows that most respondents have low psychological well-being amidst the pandemic, as shown in previous studies on the respondents of research on public health before COVID-19 (Hindoro et al., 2018; Larasati & Kristina, 2020; Soewondo et al., 2010).

Findings from this validation study might be different when compared with previous validation studies conducted before the COVID-19 pandemic settings, for example, among elderly (Heun et al., 2001), employee in the 6th European Working Condition survey in 2015 (Sischka et al., 2020), and university students in China in 2018-2019 (Fung et al., 2022)

Conclusion

In conclusion, WHO-5 is a psychometrically fit and valid brief measure that consists of simple and non-invasive questions measuring the psychological well-being of respondents. The present study suggests that future studies implement advanced psychometric and statistical measures to validate the instrument. The study's strength is the large sample size throughout Indonesia, meaning that the sample may represent the actual condition amidst the COVID-19 pandemic. The limitations of this study are that since the number of items on the WHO-5 scale is not enough for Rasch Model analysis, a further advanced analysis could not be conducted, and the majority of early adulthood participants might be because of the online method used, which could increase the self-selection of the bias.

Conflict of Interest

Authors disclose any conflicts of interest related to the work in a manuscript.

Authors Contributions

RL, KC, AL, NI develop theoretical formalism and collected the data. RL, MR, AR, writing the draft of the manuscript. AR, MR, VGSP performed computations analysis and contributed to the revised version of the manuscript. All authors discussed the results and contributed to the refinement of final manuscript.

References

- Ahrendt, D., Cabrita, J., Clerici, E., Hurley, J., Leončikas, T., Mascherini, M., Riso, S., & Sándor, E. (2020). Living, working and COVID-19. In *European Foundation for the Improvement of Living and Working Conditions*.
- Alatartseva, E., & Barysheva, G. (2015). Well-being: Subjective and objective aspects. *Procedia - Social and Behavioral Sciences*, 166, 36–42. <https://doi.org/10.1016/j.sbspro.2014.12.479>
- Bathina, K. C., ten Thij, M., Valdez, D., Rutter, L. A., & Bollen, J. (2021). Declining well-being during the COVID-19 pandemic reveals US social inequities. *PLOS ONE*, 16(7), e0254114. <https://doi.org/10.1371/journal.pone.0254114>
- Bech, P. (1999). Health-related quality of life measurements in the assessment of pain clinic results. *Acta Anaesthesiologica Scandinavica*, 43(9), 893–896. <https://doi.org/10.1034/j.1399-6576.1999.430906.x>
- Bento, A. I., Nguyen, T., Wing, C., Lozano-Rojas, F., Ahn, Y.-Y., & Simon, K. (2020). Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proceedings of the National Academy of Sciences*, 117(21), 11220–11222. <https://doi.org/10.1073/pnas.2005335117>
- Božek, A., Nowak, P. F., & Blukacz, M. (2020). The relationship between spirituality, health-related behavior, and psychological well-being. *Frontiers in Psychology*, 11(August). <https://doi.org/10.3389/fpsyg.2020.01997>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd editio). Guilford Publications.
- Comrey, A. L. (1973). *A first course in factor analysis*. Academic Press.
- Diener, E. (2009). Assessing subjective well-being: Progress and opportunities. In *Social Indicators Research* (Vol. 31, Issue 2, pp. 25–65). https://doi.org/10.1007/978-90-481-2354-4_3
- Fung, S., Kong, C. Y. W., Liu, Y., Huang, Q., Xiong, Z., Jiang, Z., Zhu, F., Chen, Z., Sun, K., Zhao, H., & Yu, P. (2022). Validity and psychometric evaluation of the Chinese version of the 5-item WHO Well-Being Index. *Frontiers in Public Health*, 10(March), 1–8. <https://doi.org/10.3389/fpubh.2022.872436>
- Heun, R., Bonsignore, M., Barkow, K., & Jessen, F. (2001). Validity of the five-item WHO Well-Being Index (WHO-5) in an elderly population. *European Archives of Psychiatry and Clinical Neuroscience*, 251(S2), 27–31. <https://doi.org/10.1007/BF03035123>
- Hindoro, E., Prapto, D., Andri, F., Tanoyo, G., Suryawijaya, N., Hermawan, P., & Taner, S. (2018). 22 quality of life and functional capacity improvement following medication adherence in long-standing hypertension. *Journal of Hypertension*, 36(Supplement 2), e6–e7. <https://doi.org/10.1097/01.hjh.0000544387.28830.c7>
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183–202. <https://doi.org/10.1007/BF02289343>
- Jöreskog, K. G., Olsson, U. H., & Y. Wallentin, F. (2016). Multivariate analysis with LISREL. In *Springer Series in Statistics*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-33153-9>
- Krieger, T., Zimmermann, J., Huffziger, S., Ubl, B., Diener, C., Kuehner, C., & Grosse Holtforth, M. (2014). Measuring depression with a well-being index: Further evidence for the validity of the WHO Well-Being Index (WHO-5) as a measure of the severity of depression. *Journal of Affective Disorders*, 156, 240–244. <https://doi.org/10.1016/j.jad.2013.12.015>

- Lara-Cabrera, M., Betancort, M., Muñoz-Rubilar, A., Rodríguez-Novo, N., Bjerkeset, O., & De las Cuevas, C. (2022). Psychometric properties of the WHO-5 Well-Being Index among nurses during the COVID-19 pandemic: A cross-sectional study in three countries. *International Journal of Environmental Research and Public Health*, 19(16), 10106. <https://doi.org/10.3390/ijerph191610106>
- Larasati, L. A., & Kristina, S. A. (2020). Prevalence of depression symptoms among diabetes mellitus patients: A survey from primary care setting in Indonesia. *International Medical Journal*, 25(5), 2273–2280.
- McDowell, I. (2010). Measures of self-perceived well-being. *Journal of Psychosomatic Research*, 69(1), 69–79. <https://doi.org/10.1016/j.jpsychores.2009.07.002>
- Muthén, L. K., & Muthén, B. O. (n.d.). *Mplus user's guide. Seventh Edition*. Muthén & Muthén.
- Nicolucci, A., Kovacs Burns, K., Holt, R. I. G., Comaschi, M., Hermanns, N., Ishii, H., Kokoszka, A., Pouwer, F., Skovlund, S. E., Stuckey, H., Tarkun, I., Vallis, M., Wens, J., & Peyrot, M. (2013). Diabetes Attitudes, Wishes and Needs second study (DAWN2TM): Cross-national benchmarking of diabetes-related psychosocial outcomes for people with diabetes. *Diabetic Medicine*, 30(7), 767–777. <https://doi.org/10.1111/dme.12245>
- Ostini, R., Finkelman, M., & Nering, M. (2015). Selecting among polytomous IRT models. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 285–304). Routledge/Taylor & Francis Group.
- Paisley, S. (2000). Book review. *Health Libraries Review*, 17(1), 61–61. <https://doi.org/10.1046/j.1365-2532.2000.00264.x>
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069–1081. <https://doi.org/10.1037/0022-3514.57.6.1069>
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69(4), 719–727. <https://doi.org/10.1037/0022-3514.69.4.719>
- Seb-Akahomen, O. J., Okogbenin, E. O., Obagaye, O. M., Erohubie, P. O., & Aweh, B. E. (2021). The 5-Item WHO Well-Being Index as a screening tool for depression in a population of doctors and nurses in Nigeria during the COVID-19 Pandemic. *Open Journal of Depression*, 10(02), 43–53. <https://doi.org/10.4236/ojd.2021.102004>
- Simon, J., Helter, T. M., White, R. G., van der Boor, C., & Łaszewska, A. (2021). Impacts of the Covid-19 lockdown and relevant vulnerabilities on capability well-being, mental health and social support: An Austrian survey study. *BMC Public Health*, 21(1), 314. <https://doi.org/10.1186/s12889-021-10351-5>
- Sischka, P. E., Costa, A. P., Steffgen, G., & Schmidt, A. F. (2020). The WHO-5 well-being index – validation based on item response theory and the analysis of measurement invariance across 35 countries. *Journal of Affective Disorders Reports*, 1, 100020. <https://doi.org/10.1016/j.jadr.2020.100020>
- Soewondo, P., Soegondo, S., Suastika, K., Pranoto, A., Soeatmadji, D. W., & Tjokroprawiro, A. (2010). The DiabCare Asia 2008 study – Outcomes on control and complications of type 2 diabetic patients in Indonesia. *Medical Journal of Indonesia*, 19(4), 235–244.
- Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84(3), 167–176. <https://doi.org/10.1159/000376585>

Warr, P., Banks, M., & Ullah, P. (1985). The experience of unemployment among black and white urban teenagers. *British Journal of Psychology*, 76(1), 75–87. <https://doi.org/10.1111/j.2044-8295.1985.tb01932.x>

World Health Organization. (2009). *Research Tools*.

Appendix A

WHO-5 Indonesian Version

Dalam dua minggu terakhir, seberapa sesuai setiap pertanyaan menggambarkan Anda?	Setiap saat	Sering sekali	Sering	Kurang sering	Terkadang	Tidak pernah
1 Saya merasa ceria dan bersemangat	6	5	4	3	2	1
2 Saya merasa tenang dan rileks	6	5	4	3	2	1
3 Saya merasa aktif dan penuh semangat	6	5	4	3	2	1
4 Saya bangun dengan perasaan segar dan cukup istirahat	6	5	4	3	2	1
5 Kehidupan sehari-hari saya dipenuhi dengan hal-hal yang menarik minat saya	6	5	4	3	2	1

Appendix B
The Original WHO-5

Over the past two weeks, how well did each statement describe you?	All of the time	Most of the time	More than half of the time	Less than half of the time	Some of the time	At no time
1 I have felt cheerful and in good spirits	6	5	4	3	2	1
2 I have felt calm and relaxed	6	5	4	3	2	1
3 I have felt active and vigorous	6	5	4	3	2	1
4 I woke up feeling fresh and rested	6	5	4	3	2	1
5 My daily life has been filled with things that interest me	6	5	4	3	2	1