Vol. 10, No. 1, 2024, pp. 87-94
DOI: https://doi.org/10.29210/020242384

Contents lists available at Journal IICET

JPPI (Jurnal Penelitian Pendidikan Indonesia)
ISSN: 2502-8103 (Print) ISSN: 2477-8524 (Electronic)

Journal homepage: https://jurnal.iicet.org/index.php/jppi

Validity and reliability tests on generalized anxiety disorder diagnostic scale

Valensia Putri Adhyartasari*), Christiana Hari Soetjiningsih, Susana Prapunoto
Psychology Faculty, Universitas Kristen Satya Wacana, Jl. Diponegoro No.52-60, Salatiga, Kec. Sidorejo, Kota Salatiga, Jawa Tengah 50711, Indoneia

Article Info

Article history:

Received Des 08th, 2022
Revised Mar 29th, 2023
Accepted Jan 21st, 2024

Keyword:

Construct validity,
Diagnostic scale,
Generalized anxiety disorder,
EFA,
CFA

ABSTRACT

Generalized anxiety disorder is one of the most common anxiety disorders in society. The purpose of this study is to develop a diagnostic scale for generalized anxiety disorder because there is no self-report diagnostic tool for generalized anxiety disorder based on DSM-V and ICD-10 criteria in Indonesia. The research uses quantitative research. Respondents in the field test consisted of 210 adult respondents aged 18-65 years, moderate or had experienced anxiety, and were Indonesian citizens. This study provides the results that the overall anxiety disorder diagnostic scale has gone through the internal validity and internal reliability test stages and obtained valid and reliable results so that this measuring instrument has described the suitability of the measuring instrument construct with the data. Based on testing the average processing time of 6 respondents, it was obtained an average of approximately 5 minutes (2 minutes 4 seconds to 5 minutes 21 seconds). Scoring for each item that supports or is in accordance with the symptoms (favorable), namely Yes = 1 and No = 0. The scoring for each item that does not support the symptoms is Yes = 0 and No = 1. This measuring instrument is declared to have good validity and reliability.

Corresponding Author:

Valensia Putri Adhyartasari
Universitas Kristen Satya Wacana,
Email: 832021801@student.uksw.edu

Introduction

Generalized anxiety disorder is one of the most common anxiety disorders (Barlow et al., 2016; Munir & Takov, 2021). Singapore as one of the countries in Southeast Asia has a percentage rate of 1.6% from 2010-2016 for generalized anxiety disorder (Statista Research Department, 2019). In addition, it is recorded that 6.8 million adults in America experience a generalized anxiety disorder (Anxiety and Depression Association of America, 2020). The percentage for generalized anxiety disorder in Indonesia does not clearly written, however, anxiety disorders are the second most common disorder experienced by Indonesian people (PDSKJI, 2020). The Indonesian Health Research and Development Agency states that 6.8% of the population experiencing anxiety disorders out of 2800 samples of the Indonesian population, while the PDSKJI data (2020) shows that 65% of the 4010 samples of the Indonesian population experience anxiety disorders.

Research conducted by Bruce et al in Olthuis et al.(2016) said that individuals who experience generalized anxiety disorder and 12 years after experiencing the first episode of GAD, only 58% will have the opportunity to experience a full recovery, but 47% who have experienced recovery are likely to experience relapse. return. In addition, someone who has a generalized anxiety disorder can experience other comorbidities such as

depression (Maslim, 2013), thus allowing errors in the determination of the diagnosis. Therefore, it is necessary to detect an early diagnosis of generalized anxiety disorder more precisely.

The way to establish a diagnosis is by conducting assessments such as interviews, observations, and psychological tests to find out the symptoms felt. One way of assessment in determining the diagnosis is a psychological test. Psychological tests in the clinical field are used for the purpose of determining the diagnosis, prognosis, and therapy of a psychiatric disorder (Bhugra et al., 2017). A psychological test is expected to be able to provide information that can guarantee the quality of diagnostic decision making by practitioners (psychologists or psychiatrists) (Evans et al., 2015). So far, in Indonesia, the diagnosis of psychiatric disorders is carried out through open clinical interviews by professional practitioners (psychologists or psychiatrists) based on the interpretation of the results of the exploration of criteria or the qualitative content of the results of clinical interviews. Measuring instruments in the form of interviews have a weakness, namely for interview data, reliability is centered on an agreement between two or more interviewers (interrater agreement or inter-interviewer agreement), so the reliability coefficient for agreement between interviewers depends on who the rater or rater agrees with (Kaplan & Saccuzzo, 2012). One example of a clinical diagnosis instrument that has been adapted and in the form of an interview is the DSM-IV version of the Mini International Neuropsychiatric Interview (MINI).

Mini International Neuropsychiatric Interview (MINI) is one type of diagnostic measurement tool in the form of a structured diagnostic interview consisting of 17 modules of Axis I psychiatric disorders which was originally developed in 1990 by Sheehan and Lecruiber, two psychiatrists from the United States. The initial theoretical basis for MINI came from the DSM-III-R and ICD-10 at that time (Harm Research Institute, n.d.)Its validity and reliability have also been tested using CIDI (Composite International Diagnosis Interview) and SCID (Structured Clinical Interview for DSM) as the gold standard. In the research of Sheehan et al (1997) the validity and reliability of the MINI has also been tested with the results of validity with concordance (kappa) = 0.36 – 0.82; sensitivity = 0.46 – 0.94; specificity = 0.72 – 0.97; PPV = 0.34 – 0.91; NPV = 0.88 – 0.99. The inter-rater reliability is 0.88 – 1.00 (kappa) and the test-retest reliability is 0.76 – 0.93 (kappa). The DSM-IV version of the MINI validation study was carried out by the Research and Development Center of the Ministry of Health of the Republic of Indonesia (2017) with the results of the validity values of the MINI instrument for depression being Sensitivity 60.68%, Specificity 80.8%; for anxiety is Sensitivity 68.94%, Specificity 80.36%; and for psychotic is Sensitivity 79.28%, Specificity 82.15%. The reliability value of the MINI instrument for depression is 0.472; for anxious is 0.399; and for psychotic is 0.577. These results indicate that MINI has the ability to distinguish individuals with and without disorders.

Psychological measuring tools in the clinical field is a measuring tool that is expected to provide information that can guarantee the quality of decision making for diagnosis and treatment that will be taken next (Eabon & Abrahamson, 2013). Although there is a structured interview instrument that has been adapted by the Research and Development Center of the Ministry of Health of the Republic of Indonesia, namely the DSM-IV version of the MINI, clinical instruments in the form of a scale such as the GAD-7 are still chosen by professionals as a screening test tool for generalized anxiety disorder (Budikayanti et al., 2019; Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI), 2020). In Indonesia, the Indonesian version of GAD-7 has been tested for epilepsy patients. The research method used is internal validity and internal reliability with the results of GAD-7 having good validity and reliability because the Spearman correlation coefficient is in the range of 0.648 to 0.800 ($p < 0.01$) while Cronbach's alpha value is 0.867 ($p > 0.70$) (Budikayanti et al., 2019). The conclusion of this research is the Indonesian version of GAD-7 was a valuable screening tool to detect GAD in patients with epilepsy (Budikayanti et al., 2019). However, this study has not tested yet with construct validity to show the extent to which the results of the instrument are able to reveal a trait or a theoretical construct that it wants to measure especially in Indonesian people, which is will find the same or different construct results (Azwar, 2018).

Measuring scales such as the GAD-7 are more time efficient where patients or clients can directly fill out instrument questions (self-report) and professional practitioners immediately get screening results or a picture of the psychological condition of their patients or clients through the interpretation of scores that have been prepared. available. In Indonesia, there is no self-report diagnostic measuring tool in the form of a questionnaire with closed questions based on the diagnostic criteria of the DSM-5 and ICD-10. Coupled with the need for more efficient psychological tools in Indonesia. The initial blueprint of the overall anxiety disorder scale was designed based on the theory of the DSM-V and ICD-10 consisting of 38 items, then the 38 items were tested for validity and reliability tests. Therefore, the purpose of this study is to develop a comprehensive anxiety disorder diagnostic measuring instrument accompanied by a validity test and reliability test on a comprehensive anxiety disorder diagnostic scale.

Method

The research uses quantitative research. Respondents in the field test consisted of 210 adult respondents aged 18-65 years, moderate or had experienced anxiety, and were Indonesian citizens. This type of data collection technique uses purposive sampling, which is a sample collection technique based on the criteria required by researchers (Sugiyono, 2019). Thus, not all members of the population have the same opportunity or opportunity to be selected as samples (non-probability sampling) (Sugiyono, 2019). Sampling of respondents is done by distributing posters. Respondents who are willing can open a google form link containing self-identity, informed consent, work instructions, and 38 questions on the overall anxiety disorder scale. The scale development stage until the validity and reliability tests are carried out are as follows:

Scale Theoretical Construction

The stages of scale development are based on the theory of psychological scale development from Azwar (2018). The initial stage is choosing a definition, recognizing, and understanding carefully the theory that underlies the attribute construct to be measured. The definition of generalized anxiety disorder is one type of anxiety disorder characterized by feelings of excessive anxiety and worry that occur all the time where it seems as if something bad will happen even though there is actually no worrying situation, but the anxiety is not specific or focused on a particular situation (DSM-5, 2013; ICD-10, 2019). Furthermore, the theory that has been owned needs to be limited according to the measurement objectives that have been set. Territory (domain) delimitation is carried out by outlining the theoretical construct of the measured attribute into several formulations of behavioral aspects more clearly. The dimensions or aspects are still conceptual so they need to be operationalized into a more concrete form of behavior so that the item writer or researcher correctly understands the direction of the response that must be revealed from the respondent. This operationalization is in the form of behavioral indicators. The indicators in this study are symptoms of generalized anxiety disorder in DSM-5 and ICD-10.

The next stage is item writing. The researcher or item writer conducts an overall review to re-examine each item written and check whether it is in accordance with the indicators to be revealed and whether it is also not out of the item writing guidelines. The scale used in this study is a 2-point Likert scale (Taherdoost, 2019). The items on this scale consist of favorable items (which contain behavioral concepts that are appropriate or support the measured attributes) with a scoring on the answers Yes = 1 and No = 0. However, the unfavorable items (which contain behavioral concepts that contradict or do not support the characteristics behavior desired by the behavioral indicators) the Yes answer will get a score of 0 and No will get a score of 1. The weighting is done by comparing the number of indicators in each aspect (Azwar, 2018). In the initial formation, this scale consists of 19 indicators so that the items formed are 38 items with 19 favorable items and 19 unfavorable items.

Pre-Tryout (Readability Test)

Pre-tryout is done with the aim of seeing whether the items, instrument responses, and instrument instructions can be understood by the target population or can be said to carry out a readability test. The readability test method used is cognitive debriefing, which is a pretest method in the form of one-on-one interviews with the main objective of determining whether the instrument can be understood by the respondent as understood by the instrument designer regarding the interpretation of items, instructions, or response options that will be chosen by the respondent (Tse et al., 2020). Minimum respondents are 5 adult respondents (Peterson et al., 2017; Willis, 2015). Respondents were obtained from the results of the scale distribution using the link provided on the poster. Five respondents have answered the scale and are willing to perform a readability test on the scale. The scale link is then closed for further legibility testing.

Table 1. Readability Test Results – Cognitive Debriefing

Aspect of Readability Test	Result
Determine respondents' rough understanding, comments on items, response options from items, and general instrument design.	Overall the items are easy to understand and understand by the respondents, relevant to their lives, the responses and instructions from the measuring instrument can be understood by the respondents well and are easy to do without help from others.
Determine if response items or options are confusing or problematic.	Items and response options are unobtrusive and easy to understand.
Determine whether the respondent considers the item the	Five respondents have understood the items they mentioned in accordance with the intent of the researcher. For example, in the question "have you felt anxious or depressed during the last 6 months?", according

same or similar to the intent of the instrument maker.	to the respondent, the researcher wants to ask about the state of anxiety of someone who answers the question, emphasizes whether or not it is true that someone is experiencing anxiety, so the nature of the question is more certain whether really restless suppose or not.
Identify confusing words or phrases.	The language used in the instrument is also easy for respondents to understand and it is easy to understand the purpose of the measuring instrument, namely the respondent captures will measure anxiety, but so far respondents do not know the type of anxiety to be measured.
Determine whether the items are relevant to the respondent.	Five respondents who were willing to do a readability test on this scale said that this scale could be understood well, the words given were correct, there were even three respondents who said that the questions on the scale were relevant to their condition.
Determine whether respondents can easily select a response option that best suits their conditions or situation.	The number of instrument questions is considered not too much because the questions given are quite short for each question and the responses are easy to choose. Suggestions from the respondents regarding language and randomization of items have also been improved by the researchers according to the suggestions from the respondents.

Note. The aspect of the readability test was adapted from Lam et.al (2020)

Content validity

Content validity is the extent to which the elements in a measuring instrument are really relevant and represent a construct that is in accordance with the measurement objectives obtained from the agreement between the assessor panels regarding the relevance of the scale items to the construct being measured (Azwar, 2018; Supratiknya, 2016). The method used to obtain content validity is the content validity index (CVI) from. Item-CVI score of at least 0.78 to be accepted that the item is valid and measures the construct to be measured (Polit & Beck, 2006). Scale-CVI score of at least 0.90 which indicates a good content validity (Supratiknya, 2016). In this study, the panel of judges consisted of 9 people. With a background of 8 people as Clinical Psychologists - Adults and 1 person is a Masters Student of Professional - Clinical Psychology who has taken the HIMPSI exam. The results of the content validity of the overall anxiety disorder diagnostic scale got an S-CVI score of 0.904. These results mean that this measuring instrument is relevant or in accordance with the theory used and has measured what you want to measure between the existing items and the theory used.

Field Test

Field testing is needed as a step to conduct a quantitative evaluation of a number of participants (Azwar, 2018).

Validity and Reliability Test

The research data analysis was carried out through the process of item discrimination power, reliability estimation, and construct validity. In the analysis of the discrimination power of an item, it is said to have item discrimination power if the item has a coefficient of 0.30. If <0.30 in this study the item has the potential to be aborted. Construct validity with the EFA method needs to meet several assumptions and stages such as Kaiser Meyer-Olkin (KMO) with a KMO value of 0.5 which indicates that the number of respondents is suitable for the factor analysis process (Hair et al., 2019; Iedliany et al., 2018) and Barttelet's test of <0.05 which indicates that a measuring instrument that is prepared is feasible to go through the factor analysis process because the correlation between items is significant (Iedliany, Fahmi, & Kusrini, 2018; Hair, et al, 2014).

The factor formed with an eigenvalue > 1 is the standard that will be used to determine the number of factors (Hardiansyah et al., 2020). The eigenvalue > 1 indicates that the formed factor deserves to be called a factor (Hardiansyah et al., 2020). The extraction technique used is the principal axis factoring and the rotation technique is oblique - oblimin (Hair et al., 2019). Initially there were 3 factors consisting of 38 items, but because in the rotation process there were still loading factors of < 0.40 which were finally eliminated so that 2 factors were formed consisting of 18 items that had loading factors of 0.40 and above also there was no cross-loading. Items that have a loading factor with this weight can state that the item is considered capable of representing the factor well, especially with a minimum of 200 respondents (Hair et al., 2019).

Next, a Confirmatory Factor Analysis (CFA) test was conducted to show that the model was in accordance with the data or that it was fit for the data according to measurement standards such as Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), dan Goodness of Fit Index (GFI) (Hair et al., 2019; Mindrila, 2010). The instrument reliability test is said to be reliable if the alpha-cronbach reliability coefficient 0.70 (Supratiknya, 2014). Processing of data along with the sociodemographic characteristics of

respondents (age, gender, education, occupation) is displayed in the form of a simple frequency processed with the help of JAMOVI software. This research has obtained a research ethics permit from the Psychological Ethics Commission, Faculty of Psychology, Universitas Gadjah Mada (Nomor: 2129/UN1/FPsi.1.3/SD/PT.01.04/2021).

Results and Discussion

Based on the results of the validity test, the number of respondents has met the requirements of factor analysis (KMO = 0.876) and the correlation between items is significant (Bartlett's Test of Sphericity = <0.001) with the measure of sampling adequacy value on each item at a level high enough to high so that all items passed for the factor analysis test (MSA = 0,614 – 0,945) (Hair et al., 2019).

Tabel 2. *Factor Loading* (EFA)

Item-Item	Sosio-Emotional Symptoms	Physiological Symptoms
Item2		0,654
Item3	0,470	
Item4		0,643
Item7		0,796
Item11	0,522	
Item13	0,628	
Item15	0,538	
Item16	0,558	
Item17	0,754	
Item21_UF	0,463	
Item25_UF		0,500
Item28_UF	0,633	
Item29_UF	0,497	
Item30_UF	0,860	
Item31_UF	0,708	
Item32_UF	0,508	
Item33_UF	0,862	
Item34_UF	0,665	

Note. The rotation method is *Oblimin.*

The final factors obtained are 2 factors compared to the initial blueprint designing 4 aspects with 18 indicators. From the table, it is known that there were 20 items that were dropped because they had loadings < 0.40 and crossloading. Items that have a loading factor value < 0.40 include item 5, item 6, item 8, item 9, item 10, item 12, item 14, item 18, item 19, item 22, item 27, item 35, item 37 and item 38. Items that have crossloading with a distance difference between the square of the largest factor load and the least squared factor load < 2.0 are item 20 and item 1. In item 23, item 24, item 26, and item 36 make up factor 3 disqualified by considering that item 23 and item 24 have a low index of discrimination power, which is <0.30 so that these items are considered to lack the ability to distinguish individuals with symptoms of generalized anxiety disorder and those who do not have it.

When item 23 and item 24 are dropped, item 26 and item 36 have a loading size below 0.4 and even under 0.3. On the other hand, a stable factor is a factor that has a minimum number of 4 items (Fabrigar, et al, 1999, in Watkins, 2018), so the stable factor obtained is 2 factors. The results of the correlation between factor 1 and factor 2 show that the two factors have a moderate correlation of 0.436 (Fabrigar & Wegener, 2018). Thus the results of construct validity get 2 factors with 18 items that make up the construct of the overall anxiety disorder instrument. Factor 1 has a factor load range of 0.463 – 0.862 and factor 2 has a factor load range of 0.500 – 0.796. Factor 1 is named socio-emotional symptoms and Factor 2 is named physiological symptoms. The alpha-cronbach reliability of this measuring instrument is 0.911 which indicates this measuring instrument has good consistency because the minimum reliability coefficient is 0.70 (Supratiknya, 2014).

The results of the confirmatory factor analysis on the overall anxiety disorder measurement tool have a fit model. With the estimated index obtained according to table 2. The construct model that is formed shows that this measuring instrument actually has a unidimensional construct that has two aspects in it. The unidimensional referred to in this measuring instrument is a generalized anxiety disorder which consists of two aspects, namely sosio-emotional symptoms and physiological symptoms.

Table 3. Measurement of Fit Index

Index	Value
Comparative Fit Index (CFI)	0.983
Root mean square error of approximation (RMSEA)	0.059
Goodness of fit index (GFI)	1,000

Table 4. Loading Score CFA

Factor	Indicator	Loading Score	Description
General Anxiety Disorder	Sosio-Emosional	0.753	Valid
	Physiological Symptoms	0.753	Valid

Reliability results were carried out before and after the discriminatory power and construct validity tests. From 38 items, the reliability obtained is 0.914. After performing the discriminatory power and construct validity, 18 items were found that had good discriminatory power and were constructively valid. Reliability was also tested to see the consistency of these 18 items using the alpha-cronbach coefficient. The coefficient obtained states that the reliability coefficient obtained is 0.911, exceeding the standard set as good reliability, which is 0.70. The reliability of each factor formed was also analyzed. The results obtained in factor 1 have an alpha-cronbach reliability coefficient of 0.912 and factor 2 has a reliability coefficient of 0.775. Both factors also show good reliability because both are at a reliability coefficient of 0.70. This indicates that the overall anxiety disorder measuring instrument has good consistency in measuring the symptoms of generalized anxiety disorder.

This measuring tool is based on theory from DSM-5 (2013) and ICD-10 (2019) to form aspects and indicators. This study uses adult respondents aged 18-65 years (Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI), 2020) with a total of 210 respondents who are willing to come from various cities. In the item discrimination power, a good coefficient of 0.30, if the discrimination power index <0.30 or closer to zero or even negative, the item should not be used because it is considered not to have the ability to distinguish individuals with symptoms of generalized anxiety disorder and those who do not have it (Azwar, 2018). The items are item 22, item 23, item 24, item 35, and item 37 which have a low-value item discrimination index (<0.30) and have a negative value so they should be aborted. The low discrimination power of items may occur because of the writing of items that do not meet the rules of item writing, such as the existence of multiple meanings (Azwar, 2018). For example, item 35 and item 37 have a double meaning, namely giving quite a lot of physical complaints in one sentence so that it can cause confusion for respondents in giving answers that are in accordance with their conditions.

The results of the factor analysis are as follows: item 5, item 6, item 8, item 9, item 10, item 12, item 14, item 18, item 19, item 22, item 27, item 35, item 37 and item 38 which have loadings below 0.40 then it is decided to be aborted. The low loading factor in an item or variable can occur because the scores obtained from the answers of each respondent have a short or less varied range of scores, so the contribution of these items or variables is less able to explain the latent construct (Bafadal, 2012). Item 20 and item 1 experienced crossloading with a significant difference in the square of the load value or loading of less than 2.0. Crossloading is a condition when a variable or item is found to have more than one significant loading (usually significant loading starts at >0.30 or >0.40), which makes it difficult to name the factors that share the same item. and thus can make it difficult to differentiate the factors involved and represent separate concepts (Dang, 2020; Hair et al., 2019). Crossloading occurs because in social research items it is rare to find "pure" items that do not share relationships with other constructs (Li et al., 2020).

In item 23, item 24, item 26, item 36 which make up a factor of 3 are discarded considering that item 23 and item 24 have low item discrimination power. When item 23 and item 24 are aborted, it is known that item 26 and item 36 have loading below the specified standard, so these two items automatically fall. When comparing the reliability when these four items are dropped, the coefficient increases to 0.911. On the other hand, the characteristics of factor 3 which consists of unfavorable items which actually form a new factor with the name of a construct that has a negative direction or the so-called method effect factor (Widhiarso, 2013). The four items form a factor that actually has the name factor of the absence of physiological symptoms where there are other factors consisting of items with physiological symptoms. Therefore, the effect factor of this method results in a factor structure that is difficult to explain (Widhiarso, 2013). We recommend that the indicators measured by unfavorable items must truly represent the low level of the measuring attribute, not measure other attributes (Widhiarso, 2013).

This phenomenon occurs as research conducted by Widhiarso (2013) where there is a collection of unfavorable items that form new factors that actually bias the scale. This event can be caused by a large proportion of unfavorable items (Widhiarso, 2013). In addition, it can also be caused by the respondent's lack of ability to understand unfavorable items or the respondent's carelessness in understanding unfavorable items (Widhiarso, 2013). Looking at the psychometric properties of the low discrimination power of items and the low content validity of items on item 23 and item 24, in addition, the loading value of item 26 and item 36 is low, which is below 0.4 and even below 0.3, as well as looking at the reliability value. increases, the factor 3 is removed. Thus, based on the results of construct validity, the final blueprint is obtained that is in accordance with the constructs that make up this measuring instrument, which has a unidimensional construct that has two aspects in it. With a construct model that has been fit. The unidimensional referred to in this measuring instrument is a generalized anxiety disorder which consists of two aspects, namely socio-emotional symptoms and physiological symptoms. With 9 favorable items and 9 unfavorable items that make up 11 indicators in it.

This study provides the results that the overall anxiety disorder diagnostic scale has gone through the internal validity and internal reliability test stages and obtained valid and reliable results so that this measuring instrument has described the suitability of the measuring instrument construct with the data. On the other hand, this research still needs to be developed such as the need for a Receiver Operating Characteristic Curve (ROC) test, which is to assess the accuracy of the relationship between sensitivity and specificity, and based on positive predictive value (PPV), which is a value that leads to how much positive results are given by truly positive instrument; and negative predictive value (PPV), which is how much the negative result given by the instrument is really negative. The four assessment indicators are usually used as a standard for testing clinical instruments so that they can be used by clinicians such as psychologists or psychiatrists. For further researchers, standardization and validation of criteria can be carried out by comparing this scale with the gold standard that is already available so that it can then be used in clinical practice.

Conclusion

The GAD-N scale is presented individually and classically considering that this scale is self-reporting and objective. Based on testing the average processing time of 6 respondents, it was obtained an average of approximately 5 minutes (2 minutes 4 seconds to 5 minutes 21 seconds). Scoring for each item that supports or is in accordance with the symptoms (favorable), namely Yes = 1 and No = 0. The scoring for each item that does not support the symptoms (unfavorable) is Yes = 0 and No = 1. Total score the result of this measurement tool is a rough score. This measuring instrument is declared to have good validity and reliability. Therefore, for further researchers, standardization and validation of criteria can be carried out by comparing this scale with the gold standard that is already available so that it can then be used in clinical practice.

References

Azwar, S. (2018). Penyusunan Skala Psikologi (2nd ed.). Pustaka Pelajar.

Bafadal, A. . (2012). Interpretasi hubungan nilai loading faktor model pengukuran dan nilai rata-rata. Diakses November 2021. https://arifkamarbafadal.wordpress.com/2012/05/26/interpretasi-hubungan-nilai-loading-faktor-model-pengukuran-dan-nilai-rata-rata/

Barlow, D. H., Durand, V. M., & Hofmann, S. G. (2016). Abnormal psychology: An integrative approach. Cengage learning.

Bhugra, D., Tasman, A., Pathare, S., Priebe, S., Smith, S., Torous, J., Arbuckle, M. R., Langford, A., Alarcón, R. D., & Chiu, H. F. K. (2017). The WPA-lancet psychiatry commission on the future of psychiatry. The Lancet Psychiatry, 4(10), 775–818. https://doi.org/https://doi.org/10.1016/S2215-0366(17)30333-4

Budikayanti, A., Larasari, A., Malik, K., Syeban, Z., Indrawati, L. A., & Octaviana, F. (2019). Screening of generalized anxiety disorder in patients with epilepsy: Using a valid and reliable Indonesian version of generalized anxiety disorder-7 (GAD-7). Neurology Research International, 2019.

Dang, H. (2020). Re: What do do with cases of cross-loading on Factor Analysis? https://www.researchgate.net/post/What-do-do-with-cases-of-cross-loading-on-Factor-Analysis/5efd7bbd6b9aab36ac600a0c/citation/download.

Eabon, M., & Abrahamson, D. (2013). Understanding psychological testing and assessment. American Psychological Association.

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., Stough, C. O., Canter, K. S., Robles, R., & Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. International Journal of Clinical and Health Psychology, 15(2), 160–170. https://doi.org/https://doi.org/10.1016/j.ijchp.2014.12.001

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis . United Kingdom: Cengage Learning, EMEA.

Hardiansyah, H., Putri, A. P., Wibisono, M. D., Utami, D. S., & Diana, D. (2020). Penyusunan alat ukur resiliensi akademik. Psikostudia: Jurnal Psikologi, 9(3), 185–194.

Harm Research Institute. (n.d.). No Title. 2019. https://harmresearch.org/index.php/mini-international-neuropsychiatric-interview-mini/#Adult Versions of the MINI

Iedliany, F., Fahmie, A., & Kusrini, E. (2018). Pengembangan dan validasi instrumen pengukuran efektivitas tim di usaha mikro kecil dan menengah (UMKM). Psikohumaniora: Jurnal Penelitian Psikologi, 3(2), 177–196. https://doi.org/https://doi.org/10.21580/pjpp.v3i2.3014

Kaplan, R. M., & Saccuzzo, D. P. (2012). Pengukuran Psikologi: Prinsip, penerapan, dan isu. Jakarta: Salemba Humanika.

Li, Y., Wen, Z., Hau, K.-T., Yuan, K.-H., & Peng, Y. (2020). Effects of cross-loadings on determining the number of factors to retain. Structural Equation Modeling: A Multidisciplinary Journal, 27(6), 841–863. https://doi.org/10.1080/10705511.2020.1745075

Maslim, R. (2013). Diagnosis Gangguan Jiwa, Rujukan Ringkas PPDGJ-III dan DSM-V (Cetakan 2). PT. Nuh Jaya.

Mindrila, D. (2010). Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. International Journal of Digital Society, 1(1), 60–66.

Munir, S., & Takov, V. (2021). StatPearls. StatPearls Publishing.

Olthuis, J. V, Watt, M. C., Bailey, K., Hayden, J. A., & Stewart, S. H. (2016). Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database of Systematic Reviews, 3. https://doi.org/https://doi.org/10.1002/14651858.CD011565.pub2

Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI). (2020). 5 Bulan Pandemi Covid di Indonesia. http://pdskji.org/home

Peterson, C. H., Peterson, N. A., & Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response processes. Measurement and Evaluation in Counseling and Development, 50(4), 217–223.

Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Alfabeta.

Supratiknya, A. (2014). Pengukuran psikologis. Yogyakarta: Universitas Sanata Dharma.

Supratiknya, A. (2016). Kuantifikasi validitas isi dalam asesmen psikologis. Lustrum Ke-4 Fakultas Psikologi Universitas Sanata Dharma.

Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. Hamed Taherdoost, 1–10.

Tse, E. T. Y., Lam, C. L. K., Wong, C. K. H., Chin, W. Y., Etz, R. S., Zyzanski, S. J., & Stange, K. C. (2020). Cultural adaptation and content validity of a Chinese translation of the 'Person-Centered Primary Care Measure': findings from cognitive debriefing. Family Medicine and Community Health, 8(4). https://doi.org/10.1101/2020.07.15.20154179

Widhiarso, W. (2013). Estimasi reliabilitas pengukuran dalam pendekatan model persamaan struktural. Buletin Psikologi, 17(1).

Willis, G. B. (2015). Analysis of the cognitive interview in questionnaire design. Oxford University Press.