



**INSAN**

Jurnal Psikologi dan Kesehatan Mental

<http://e-journal.unair.ac.id/index.php/JPKM>

p-ISSN 2528-0104 | e-ISSN 2528-5181



RESEARCH ARTICLE / ARTIKEL PENELITIAN

## Psychometric Properties Analysis of Self-Reported Questionnaire (SRQ)-20 Instrument with Rasch Model

Nurul Fajriyah Prahastuti

Departemen Ilmu Kesejahteraan Sosial, Fakultas Dakwah dan Komunikasi, UIN Sunan Kalijaga  
Yogyakarta

### ABSTRACT

Screening instruments for mental disorders is useful for society in identifying the possibility of mental disorders in themselves or others. Screening instruments are tools used to examine early symptoms of mental disorders, early detection, intervention for mental disorders, as well as research. The Self-Reported Questionnaire-20 (SRQ-20) is an instrument for screening mental disorders through individual assessment. This study aims to determine the quality of SRQ-20 instrument using Rasch model analysis techniques with the assistance of Jamovi software version 2.3.18.0. The participants in this study were individuals aged 12 – 44 years. The number of participants was 165. The instrument used was SRQ-20 with a total of 20 items. The results of the study showed that all 20 items met the item-model fit. The reliability coefficient value was 0.823. Overall, SRQ-20 has good psychometric properties.

**Keywords:** *psychometric properties, rasch model, self-reported questionnaire-20, srq-20*

### ABSTRAK

Instrumen skrining gangguan mental berguna bagi masyarakat dalam mengidentifikasi kemungkinan adanya gangguan mental pada diri mereka sendiri atau orang lain. Instrumen skrining adalah alat yang digunakan untuk memeriksa gejala awal dari gangguan mental, deteksi dini, intervensi terhadap gangguan mental, maupun penelitian. *Self-Reported Questionnaire-20* (SRQ-20) adalah instrumen untuk melakukan skrining terhadap gangguan mental dengan melakukan asesmen individual. Penelitian ini bertujuan untuk mengetahui kualitas instrumen SRQ-20 dengan menggunakan teknik analisis model rasch dengan bantuan *software* Jamovi versi 2.3.18.0. Partisipan penelitian ini adalah individu berusia 12 – 44 tahun. Jumlah partisipan sebanyak 165. Instrumen yang digunakan adalah SRQ-20 dengan jumlah butir butir sebanyak 20. Hasil penelitian menunjukkan bahwa 20 butir memenuhi ketepatan butir-model. Nilai koefisien reliabilitas sebesar 0,823. Secara keseluruhan SRQ-20 memiliki kualitas properti psikometris yang baik.

**Kata kunci:** *model rasch, properti psikometris, self-reported questionnaire-20, srq-20*

INSAN Jurnal Psikologi dan Kesehatan Mental, 2023, Vol. 8(2), 148-173, doi: 10.20473/jpkm.v8i22023.148-173

Submitted: 30/05/2023 Accepted: 21/09/2023 Published: 11/12/2023

Editor: Rahkman Ardi

\*Corresponding Author's Email: nurulfajriyahprahastuti@gmail.com



This manuscript is under the open access policies and the Creative Common Attribution License (<http://creativecommons.org/licenses/by/4.0>); therefore, any use, distribution, and reproduction of this article, in any media, is not restricted as long as its original source is properly cited.

## INTRODUCTION

Mental health issues are a significant global challenge due to the high prevalence and high levels of suffering experienced by individuals, families, communities, and countries. It was found that there was a visible increase in cases of mental disorders across the globe (Handayani et al., 2020; Prasetyo et al., 2022). According to Dianovinina (2018), the prevalence of depression in adolescents was 8% and adolescents have the potential to experience mental disorders of 28%. The problem of mental disorders has worsened with the Covid-19 pandemic situation. Covid-19 survivors revealed that despite recovering from Covid-19, the mental recovery process did not happen instantly (Santoso et al., 2023). Mental disorders refer to behavioral or psychiatric patterns that cause considerable suffering or interfere with one's functioning (Xiao et al., 2023). According to Nurhaeni et al. (2022), anyone can experience mental or psychiatric disorders. Mental disorders are regardless of age, gender, social background, or other factors. This emphasizes the importance of understanding and support for mental health issues in society.

Mental health is a key element in the conception of general health (Ayuningtyas et al., 2018). Mental health, according to Law No. 18/2014 on Mental Health, includes conditions in which individuals can experience balanced development physically, mentally, spiritually, and socially (Kemenkes RI, 2014). This law underscores the importance of maintaining balance and well-being in various aspects of life, including physical, mental, spiritual, and social aspects. Mental health is recognized as a serious issue which includes not only patients, but also general population (Lee et al., 2020). Mentally healthy individuals have an awareness of their own potential, are able to cope with stress, productive, and are able to contribute to society. According the law, mental health is emphasized as an important component in maintaining balance and quality of life. Through this understanding, the government and citizens are expected to pay more attention to mental health and ensure adequate support and access for individuals in maintaining their mental balance and well-being.

Based on previous explanation, it is necessary to do mental health screening. Mental health screening aims to identify the possibility of mental disorders, prevention, provision of appropriate referrals, and also stress management. By recognizing the importance of mental health status screening, it is hoped that people can be more proactive in maintaining their mental health and providing support to those around them who may need it. An instrument used for screening purposes is expected to have good quality, both overall and per items. This quality can be tested by conducting item analysis, an analysis to determine the items' quality of an instrument, one of which is by analyzing the difficulty level or difficulty index (Suryana, 2015). According to Embretson & Reise (2000), the ideal difficulty level in screening or diagnostic instruments depends on various factors, such as the characteristics of the population being tested, the purpose of the test usage, and the type of construct to be measured. Ideal statement items have balanced difficulty levels, not too easy, nor difficult (Amalia & Widayati, 2012). An ideal statement item is a type of item that is carefully designed so that it has a balanced level of difficulty, thus neither too easy nor too difficult for participants. With a balanced difficulty level, the test can

provide more accurate information about the level of understanding or ability of the test taker. If an item is too easy, there is no significant difference in the performance of each participants, making it difficult to distinguish between their ability levels. On the other hand, if an item is too difficult, many participants may fail to answer and test administrators will have difficulty in distinguishing ability among participants. With a balanced level of difficulty, the test can measure more precisely the range of abilities of the participants, helping to identify those who excel, those who need attention, or those who need further help in understanding the material being tested. A balanced difficulty level also provides a fairer experience for all participants, ensuring that they are tested proportionally according to their ability. Therefore, in the test designing process, it is necessary to consider and ensure the right difficulty level in order to achieve better evaluation and more valuable information about participants.

Self-Reported Questionnaire (SRQ-20) is an instrument used for screening for mental disorders, consisted of 20 statements (Netsereab et al., 2018). SRQ-20 is widely used due to its convenience and time efficiency (Sunjaya et al., 2022). SRQ-20 was developed for screening non-psychotic disorders as part of a WHO collaborative study on strategies for mental health care expansion, particularly in developing countries (Husain et al., 2016). SRQ-20 is generally completed by individuals themselves where they respond to questions with reference to their own experiences and circumstances. The purpose of designing SRQ-20 is to identify the possible presence of general and specific mental disorders, such as depression, anxiety, and other emotional problems (Chua et al., 2022; Dharmayanti et al., 2018). SRQ-20 has been used in various research and clinical contexts as an initial tool to identify individuals who may require further evaluation related to their mental health.

SRQ-20 has five factors: energy, cognitive, depression, physiological, and anxiety (Prasetio et al., 2022). The energy factor reflects some of the characteristics of people who have decreased physical and mental energy. Cognitive factors describe some signs of cognitive problems, such as decreased quality of work output, loss of interest, and difficulty in making decisions. The depression factor describes some of the emotional and cognitive symptoms associated with depressive conditions, such as feelings of unhappiness and worthlessness, as well as suicidal thoughts. The physiological factor is characterized by the effect of psychological disorders on the digestive system. The anxiety factor describes several physical and psychological symptoms, such as feelings of tension or fear, and trembling hands. Given the high prevalence of mental disorders, it is important to have accurate and effective measurement instruments to identify symptoms of mental disorders. The use of these instruments is beneficial in improving the prevention and/or immediate treatment of individuals with mental problems, so that they can develop their potential.

There are several previous studies to identify the quality of SRQ-20. Prasetio et al. (2022) investigated the factor structure of SRQ-20 and collect its psychometric properties. The one-factor SRQ-20 model stated that the variable measured is general mental disorders, while the five-factor model indicated the presence of specific factors that help professionals on more targeted actions. Both models can be used as a reference for interpretation. For five-factor model, further analysis and the use of a wider sample population are needed. Stratton et al. (2013) have evaluated the psychometric properties of the 20-item WHO Self-Reporting Questionnaire (SRQ-20) in an epidemiological survey of adults in Vietnam. The results of Stratton et al. (2013) mentioned that the scree plot and EFA results, two latent structures were considered feasible and then subjected to further modeling. The bi-factor model (BFM) and the correlated three-factor model solution (Negative Affect, Somatic Complaints, and Hopelessness) provided adequate results. The BFM determined a single Generalized Distress Dominant factor (all SRQ-20 items) with orthogonal group factors for a subset of items: Negative Affect (9 items), Somatic Complaints (8 items), and Hopelessness (3 items). This model fits the data as well or even better than the three-factor model. Netsereab et al. (2018) mentioned that SRQ-20 has good internal reliability

( $\alpha=0.78$ ). Factor analysis yielded two factors that explained 31.2% of the total variance. The instrument performed well in detecting common mental disorders, with an area under the curve (AUC) of 0.879 (SE=0.23, 95% CI 0.83-0.92) in the overall sample, with an optimal cut-off score of 5/6 with a sensitivity of 78.6% and specificity of 81.5%. The cut-off score differed between females (5/6) and males (4/5). For male participants, the AUC statistics were 0.877 (SE=0.04, 95% CI 0.79-0.96) and 0.871 (SE = 0.02, 95% CI 0.81-0.92) for female participants.

Of the few studies that have been conducted to test the quality of SRQ-20 psychometric properties, research using the Rasch model has not been widely conducted. The purpose of Rasch analysis is not to identify dimensions that appear in the data as a result of shared variance between indicators, but to test and measure the psychometric characteristics of measurement instruments that rely on the assumption that a person's probability of successfully answering an item is related to individual ability and the difficulty level of the item (Stolt et al., 2022; Widhiarso, 2018). In the context of research using Rasch model analysis on SRQ-20, the aim is to test the validity and psychometric quality of SRQ-20 using Rasch approach that produces item difficulty levels in logit units (Urbayatun & Widhiarso, 2012). As part of Item Response Theory, Rasch model analysis has several advantages when compared to classical model analysis. First, it is dependent on group characteristics. According to Magno (2009) and Steyer (2001), reliance on group characteristics in classical test theory means that it is assumed that the group being tested is a reflection of the entire population to be measured. The generalizability of the results can be improved because the results obtained are more specific to each individual, not just limited to a particular test group (Steyer, 2001). Second, there are difficulties in estimating item parameters. Classical test theory faces the problem of precisely estimating the difficulty of an item. This estimate relies on group performance, which does not always accurately reflect the true difficulty of the item (Magno, 2009).

Based on previous statements, our purpose for conducting this study is to determine the quality of Self-Reported Questionnaire (SRQ-20) developed by WHO by analyzing it using the Rasch approach. Although considered a rigorous approach, Rasch model can still be influenced by the interaction of individuals working with the scale, as well as the heterogeneity of individual characteristics (Preinerstorfer & Formann, 2011; Widhiarso, 2013). There has not been much research on analyzing the quality of SRQ-20 with Rasch approach on adolescent subject groups in Indonesia, thus we considered the test is necessary.

## METHOD

### *Research Design*

This study used a quantitative approach, namely Rasch model analysis. Data collection was carried out using the help of Google Form. Data collection took place from August 30, 2022 to November 5, 2022.

### *Participants*

The total number of participants was 165 Indonesian citizens; 113 women and 52 men within the age range of 12 – 44 years. I used convenience sampling method, a non-probability sampling method where participants were willing to work on their own decisions after getting access to filling out this scale (Stratton, 2021; Firmansyah & Dede, 2022).

On Google Form, I put informed consent on the first page, preceding the identity form. Participants were allowed to write their full names or initials, and all data collected are confidential and only used for research purposes. I provided our contacts in case they had further questions. Participants were allowed to withdraw themselves in case they felt uncomfortable.

### Measurement

The measurement instrument is Self-Reported Questionnaire (SRQ-20) developed by WHO and consists of 20 close-ended questions (WHO, 1994). Correct answers get a score of 1 and a score of 0 for incorrect answers. Theoretically, the maximum score of SRQ-20 is 20 points and the minimum score is 0. The higher the score, the higher the number of symptoms experienced (Fatahya & Abidin, 2017).

### Data Analysis

Data analysis was carried out using Rasch analysis to obtain reliability values, item fit, Wright Map, and scale dimensionality. Data analysis was carried out with the help of Jamovi software version 2.3.18.

## RESULTS

The Rasch analysis output from Jamovi shows the item statistics for proportion, measure, SE Measure, infit, and outfit. Details of the output can be seen in the following table.

Table 1. Statistics Item

	Proportion	Measure	S.E. Measure	Infit	Outfit
1	0.448	0.352	0.200	1.502	1.941
2	0.327	1.185	0.210	1.212	1.421
3	0.406	0.635	0.202	1.281	1.469
4	0.418	0.554	0.202	0.998	0.991
5	0.164	2.587	0.256	1.200	1.287
6	0.533	-0.208	0.200	0.924	0.940
7	0.236	1.895	0.228	1.042	1.050
8	0.376	0.842	0.205	0.907	0.808
9	0.364	0.926	0.206	0.843	0.665
10	0.212	2.109	0.235	1.213	1.307
11	0.358	0.968	0.206	0.921	0.781
12	0.497	0.032	0.200	0.854	0.708
13	0.291	1.455	0.215	0.848	0.760
14	0.309	1.318	0.212	0.705	0.477
15	0.291	1.455	0.215	0.747	0.523
16	0.358	0.968	0.206	0.835	0.697
17	0.176	2.460	0.249	0.859	0.706
18	0.376	0.842	0.205	0.899	0.697
19	0.364	0.926	0.206	1.184	1.258
20	0.600	-0.656	0.204	0.976	0.942

*Note. Infit: Information-weighted mean square statistic; Outfit: Outlier-sensitive means square statistic.*

Based on the results of the study, it was found that SRQ-20 has a good reliability (0.823). Meanwhile, the analysis results also showed that the person separation reliability is 0.837. According to Widhiarso (2013), the reliability value is acceptable because it is in the range of  $0.81 \leq p < 0.90$  (good).

Participants received scores ranging from 2 to 20, with a mean of 7.1 and SD of 5.83. Infit values ranged from 0.705 to 1.502. All items were fit as there were no INFIT MNSQ values  $>2$ . The results of the OUTFIT MNSQ analysis were 0.477 to 1.941 and 2 invalid statement items were found because they had an OUTFIT MNSQ value  $>1.5$ , namely item number 1 with a value of 1.941 and item number 14 with a value of 0.477. Item fit refers to the extent to which individual items in a test or instrument are in line with the Rasch model used to measure individual abilities. Item fit relates to the residual, which is the difference between the actual response and the expected response. The expected response is calculated based on the probability of a particular response based on individual ability and item difficulty. According to Müller (2020), item fit statistics present a summary of the standardized residuals for all items in a test or instrument. These statistics provide a measure of how well the observed response matches the expected response based on the Rasch model and can be used to measure the ability of the subject. If the items in the test have a good item fit, it indicates that the test as a whole fit the Rasch model and can be relied upon to accurately measure individual ability (Linda Crocker, 1986; Müller, 2020). Conversely, if the observed response pattern is not as expected, it indicates that the item is problematic, does not fit the Rasch model and needs to be revised or deleted (Kurniawan & Andriyani, 2018; Linacre, 2011).

The proportion value ranges from 0.164 to 0.6. The proportion of this value serves to assess the extent to which the data obtained is in accordance with the assumptions of the Rasch model being used. According to Veas et al. (2016), the measure value and proportion in Rasch model analysis were interconnected and both had an important role in ensuring the accuracy and validity of measurement results. A good measure will produce a good proportion as well. A good measure value indicates that respondents have a good ability to answer questions, so they tend to give answers that are in accordance with the Rasch model. This will increase the proportion of values that fit the Rasch model, so that the data obtained is more in line with the assumptions of the Rasch model. A good proportion will produce a good measure. A good proportion of values indicates that the data obtained is in accordance with the assumptions of the Rasch model, so that it can improve the accuracy of estimating item parameters and respondents' abilities. This will result in better measure values, thus improve the overall quality of measurement.

Further analysis was to examine the results of Wright Map aiming to find out which statement items were difficult and easy for participants to answer.

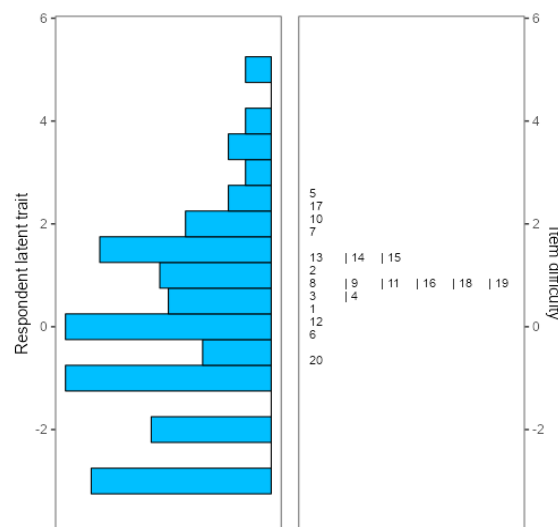


Figure 1. Wright Map



Based on figure 1, it is shown that there were no statement items that are lower than the -2 logit value. There are statement items that have a value of more than 2 logits, namely numbers 10, 17, and 5, which means that the statements in these numbers are difficult to answer. According to Linacre (2009, 2011), individual and item abilities are represented as points or lines on the Wright Map. The location of the individual on the x-axis indicates the level of ability, while the location of the item on the y-axis indicates the difficulty of the item. The distance between the individual's point and the item line illustrates the extent to which the individual can answer correctly on that item (Alagumalai, S. et al., 2005; Boone et al., 2014). If the subject's ability exceeds the item difficulty level, then the subject is considered capable of answering the item correctly. Meanwhile, if the subject's ability is lower than the item difficulty level, then the subject is considered unable to answer the item (Hilaliyah et al., 2019).

## DISCUSSION

Yessimov et al. (2023) mentioned that Rasch model is a probabilistic model used to study categorical data by taking into account the correlation between the ability of test takers and the level of item difficulty. The reliability coefficient obtained in this study is quite high, which is 0.823. However, there is no specific number that can be considered a satisfactory level of reliability coefficient. Previous research on SRQ-20 test quality conducted in primary health care in Eritrea resulted in a reliability coefficient of 0.78 (Netsereab et al., 2018). Meanwhile Scholte (2013), mentioned that the reliability coefficient of SRQ-20 used in Rwanda was 0.85 for the female group and 0.81 for men. Kröner et al., (2017) mentioned the results of SRQ-20 reliability analysis of adolescent participants in Germany gave a coefficient of 0.88. The reliability coefficient obtained in this study produces a number that is not much different from previous studies. The higher the reliability coefficient, the greater the confidence that can be given to the test results (Azwar, 2016). High stakes standardized tests are expected to have a reliability value of 0.9, while lower levels are expected to have a reliability value of 0.8 or 0.85 (Santamaría-Vázquez et al., 2023; Wells & Wollack, 2003). In conclusion, SRQ-20 is a screening instrument to determine the potential of mental disorders that has good reliability.

In addition, to find out unsuitable items can be seen from the mean square (MNSQ) value, where the accuracy of the item can be seen from the ideal item through the MNSQ value (Linacre, 2009). The guideline for assessing item suitability criteria according to the MNSQ outfit value is  $0.5 < \text{MNSQ} < 1.5$  (Ramdani et al., 2020). Linacre (2009) explained that MNSQ items are in the range of 0.5 to 1.5, In addition, it can also be known by summing the average value of the center square infit (mean MNSQ Infit) (Sumintono & Widhiarso, 2013). The results of the OUTFIT MNSQ analysis were 0.477 to 1.941. The proportion value refers to the percentage or proportion of the level of conformity between the responses given by participants in the test or questionnaire with the Rasch model being used (Suminto & Wening, 2022).

The Wright Map graph is a visualization used in the Rasch model to depict individual ability levels and item difficulty levels on a measurement scale. This graph illustrated the relative location of individuals and items on a single axis. Based on the results on figure 1, items number 5, 10, and 17 were difficult items. According to Prasetio et al. (2022), item number 5 is an anxiety factor, "*are your hands shaking?*". Item number 10 does not belong to any factor, which said "*do you cry more often than usual?*". Item number 17 is from the depression factor, stated "*do you have thoughts of ending your life?*". In the measurement context of SRQ-20, an individual's high score indicated a possible mental health problem. Low scores indicated that the participants have few or no mental health symptoms as measured by the questionnaire (van der Westhuizen et al., 2016; WHO, 1994). This means that the individual may not currently have a mental health problem or the symptoms are so mild that they cannot be detected.

*Item Characteristic Curve*

The Item Characteristic Curve (ICC) in the Rasch model is a logit-shaped curve that describes the correlation between an individual's measured ability and their probability of answering correctly on a particular test item (Nguyen et al., 2014). ICC is one of the important components in Rasch model used to quantify and validate measurement instruments. This curve can help identify items that are too easy or too difficult for the intended population, as well as items that cannot differentiate between participants with different ability levels. The curve has an almost S-shape as the probability of correct answers tends to be low at very low and very high skill levels, but increases sharply at intermediate skill levels. Each item will have a different shape from the others. The output of Rasch analysis results in the form of Item Characteristic Curve (ICC) can be seen in the following figure. Each graph has an x and y axis. The x-axis shows the ability or ability level. The y-axis shows the probability of answering correctly on each item. The black colored curve shows the empirical ICC and the blue colored curve shows the theoretical ICC. The slope shown by the blue graph indicates that the item has good sensitivity to measure individual ability.

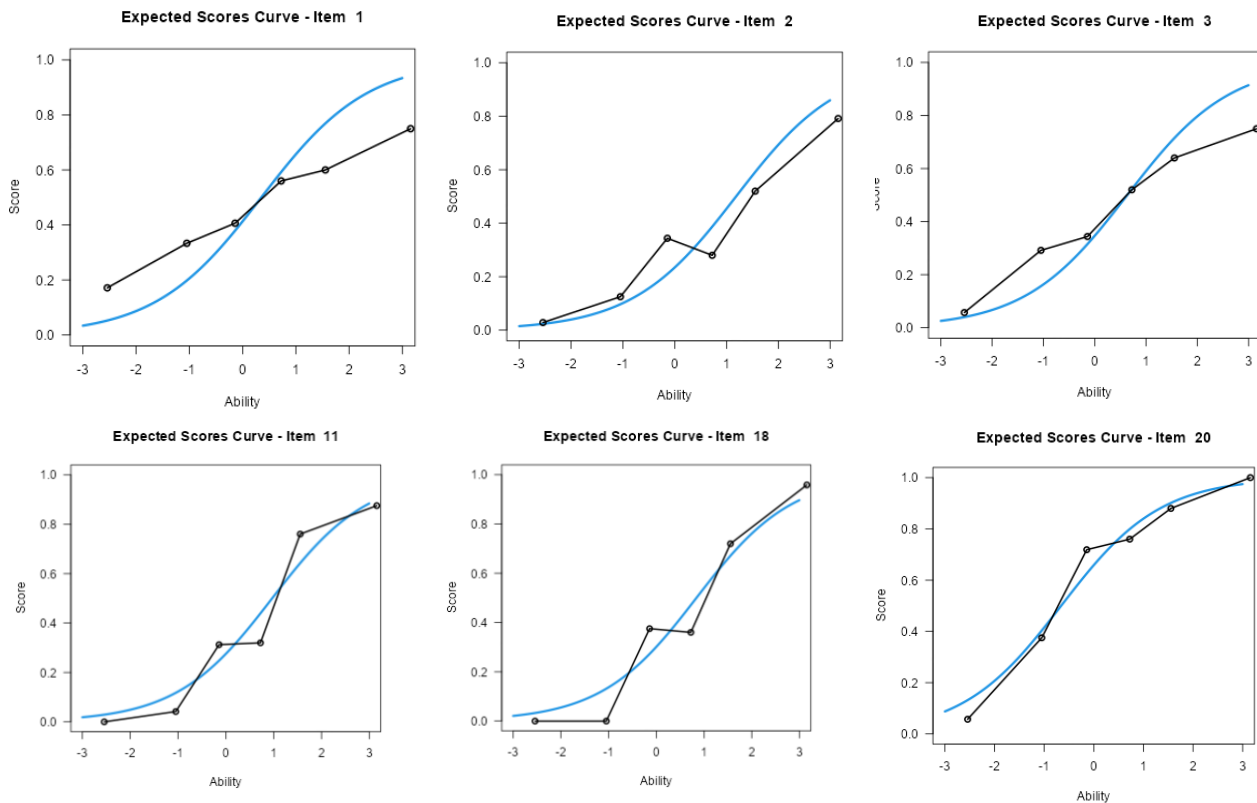


Figure 2. Item Characteristic Curve Factor 1: Energy



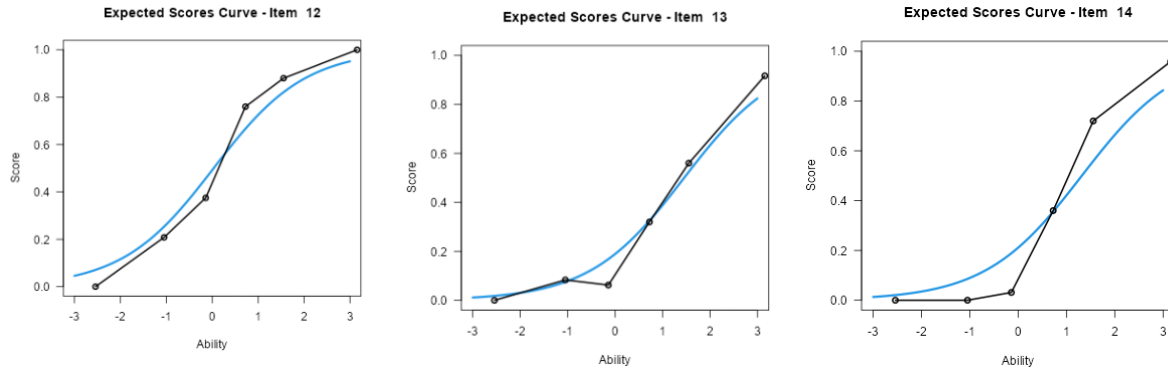


Figure 3. Item Characteristic Curve Factor 2: Cognitive

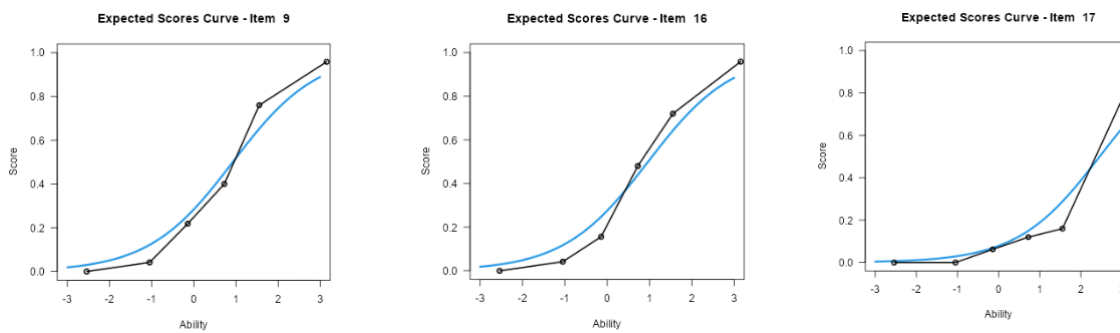


Figure 4. Item Characteristic Curve Factor 4: Depression

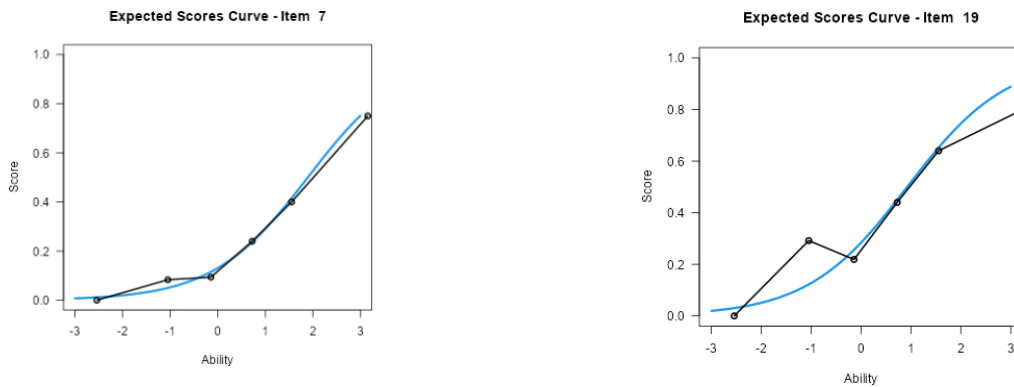


Figure 5. Item Characteristic Curve Factor 4: Physiology

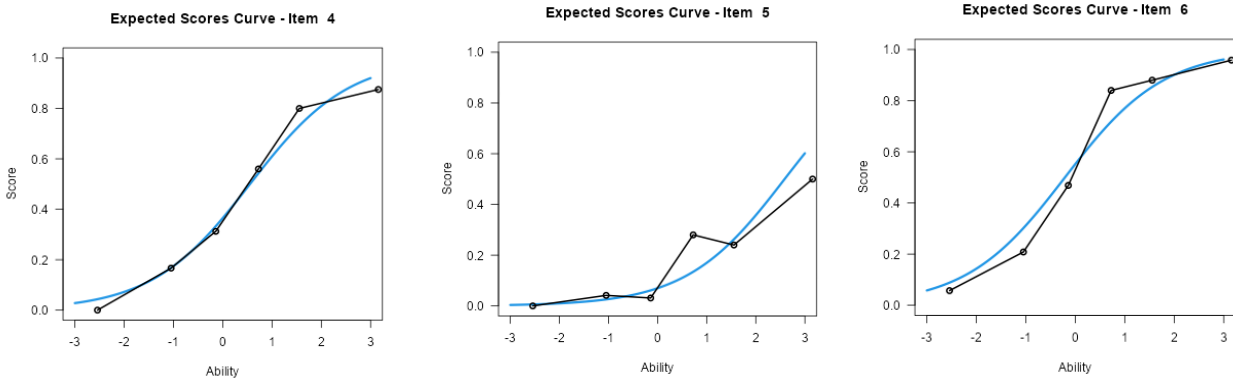


Figure 6. Item Characteristic Curve Factor 5: Anxiety

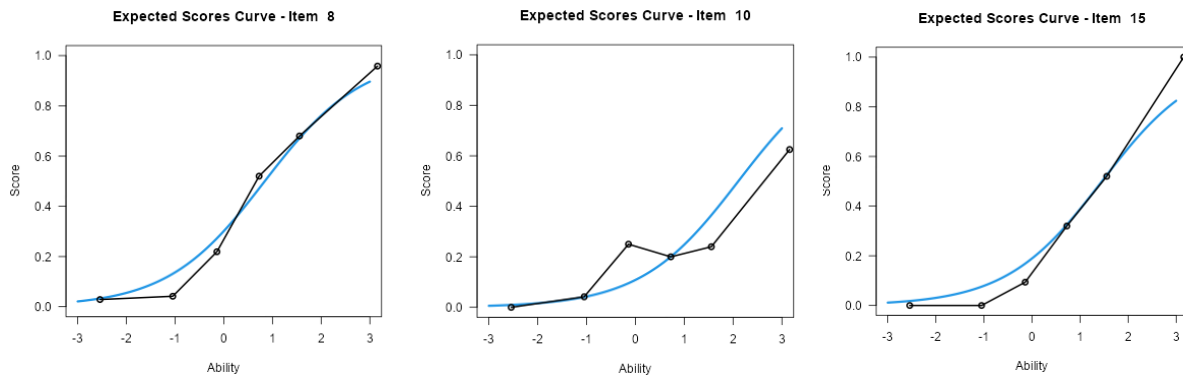


Figure 7. Item Characteristic Curve Without Factor

In figure 2 there were 6 Item Characteristic Curve (ICC) graphs. The ICC graph on item 1 showed a graphical pattern that was less close together. Item number 1 has Infit and Outfit MNSQ values of 1.502 and 1.941 where the limit is  $0.5 < \text{MNSQ} < 1.5$ , it can be concluded that item number 1 is less fit because the output value exceeds 1.5. Furthermore, statement item number 14 has MNSQ infit and outfit values of 0.705 and 0.477 where the item had an outfit value that was below the standard 0.5 so it is less fit. The ICC image for item number 14 can be seen in figure 3 where the curves were less close together. There was a relationship between ICC, infit, and outfit in Rasch analysis. In the context of the SRQ-20, the ICC would provide an overview of the difficulty level and discrimination of each item on the questionnaire. Then, infit and outfit were used to evaluate the extent to which participants' responses to each item fit the Rasch model and the sensitivity of the item to variations in participants' abilities. If an item has a good ICC (appropriate difficulty and discrimination), then it is expected that the infit and outfit values will also be good. Conversely, if the test item has a poor ICC (inappropriate difficulty and discrimination), then the infit and outfit values are likely to be poor as well.

*Differential Item Functioning (DIF)*

Based on the results of the analysis, it was known that DIF occurred in item number 8 ( $p=0.001$ ), number 10 ( $p<0.001$ ), and number 18 ( $p=0.028$ ). There were 3 items that indicate DIF. DIF or measurement bias occurred when individuals from different groups gave different responses to an item, resulting in variations in the estimated parameters within the same ability level (Nguyen et al., 2014).



### *Unidimensionality*

The dimensionality test using Jamovi did not show information about unidimensionality. I conducted a unidimensionality analysis using confirmatory factor analysis (CFA). The analysis results showed that the CFI value was 0.872 and the RMSEA was 0.078.

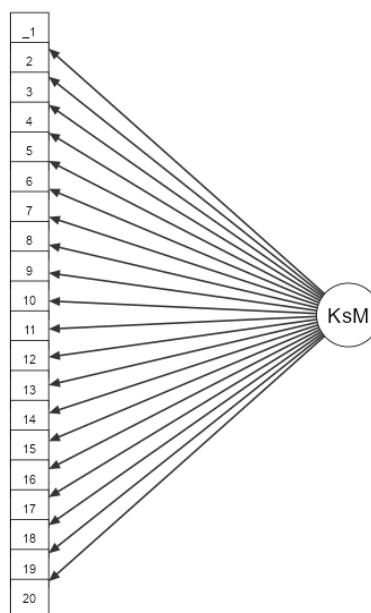


Figure 8. Path Diagram CFA

Based on figure 8, it can be seen that all items in SRQ-20 were in the same construct and labeled as mental health factors. The items in SRQ-20 were confirmed to fulfill the assumption of unidimensionality.

## **CONCLUSION**

Using Rasch model, this study focuses on analyzing the quality of SRQ-20 which is a screening instrument for mental health. Rasch model is a probabilistic model used to study categorical data by taking into account the correlation between test taker ability and item difficulty. The higher a person's ability compared to the difficulty level of the item, the more likely the person is to give the correct answer. The results of this study indicates that in general SRQ-20 has good reliability and the items that fit the model are as many as 18. This study still has population limitations so it would be better if the sample used is wider with a larger number. There is a DIF analysis conducted but there is a possibility of inaccurate results due to limited samples.

## Analisis Properti Psikometris pada Instrumen *Self-Reported Questionnaire* (SRQ)-20 dengan Rasch Model

Masalah kesehatan mental menjadi sebuah tantangan global yang signifikan karena tingginya prevalensi dan tingkat penderitaan yang dialami oleh individu, keluarga, komunitas, dan negara. Terdapat peningkatan yang terlihat secara global dalam kasus gangguan mental (Handayani dkk., 2020; Prasetyo dkk., 2022). Menurut penelitian Dianovinina (2018), prevalensi depresi pada remaja sebesar 8% serta remaja memiliki potensi mengalami gangguan mental sebesar 28%. Gangguan mental pun memburuk dengan adanya situasi pandemi Covid-19. Penyintas Covid-19 mengungkapkan bahwa meskipun dinyatakan pulih dari Covid-19, proses pemulihan secara mental tidak terjadi secara instan (Santoso dkk., 2023). Gangguan mental merujuk pada pola perilaku atau kejiwaan yang menyebabkan penderitaan cukup besar dan mengganggu fungsi individu secara pribadi (Xiao dkk., 2023). Menurut Nurhaeni dkk., (2022), siapa pun bisa mengalami gangguan mental. Gangguan mental tidak memandang usia, jenis kelamin, latar belakang sosial, atau faktor lainnya. Hal ini menunjukkan pentingnya untuk meningkatkan pemahaman dan dukungan terhadap masalah kesehatan mental dalam masyarakat.

Kesehatan mental adalah elemen pokok dalam konsepsi kesehatan secara umum (Ayuningtyas dkk., 2018). Menurut Undang-Undang Nomor 18 Tahun 2014 tentang Kesehatan Jiwa, kesehatan mental mencakup kondisi di mana individu dapat mengalami perkembangan yang seimbang secara fisik, mental, spiritual, dan sosial (Kemenkes RI, 2014). Undang-undang ini menggarisbawahi pentingnya menjaga keseimbangan dan kesejahteraan dalam berbagai aspek kehidupan, termasuk aspek fisik, mental, spiritual, dan sosial. Kesehatan mental semakin dikenal sebagai isu serius yang tidak hanya mencakup pasien, tetapi juga populasi secara umum (Lee dkk., 2020). Individu yang sehat secara mental memiliki kesadaran akan potensi diri, mampu mengatasi tekanan, dapat berfungsi secara produktif, dan mampu berkontribusi bagi komunitasnya. Berdasarkan undang-undang, kesehatan mental ditekankan sebagai komponen penting dalam menjaga keseimbangan dan kualitas hidup seseorang. Melalui pemahaman ini, diharapkan pemerintah dan masyarakat dapat memberikan perhatian yang lebih besar terhadap kesehatan mental, serta memastikan adanya dukungan dan akses yang memadai bagi individu dalam menjaga keseimbangan dan kesejahteraan mentalnya.

Berdasarkan penjabaran di atas, tindakan skrining status kesehatan mental diperlukan. Skrining status kesehatan mental bertujuan untuk mengidentifikasi dini potensi gangguan mental pada individu, pencegahan, pemberian rujukan yang tepat, dan juga pengelolaan stres. Dengan memperhatikan pentingnya skrining status kesehatan mental, diharapkan masyarakat dapat lebih proaktif dalam menjaga kesehatan mental mereka dan memberikan dukungan kepada orang-orang di sekitar yang mungkin membutuhkan. Sebuah instrumen yang digunakan untuk keperluan skrining diharapkan memiliki kualitas yang bagus, baik secara keseluruhan maupun per butir. Kualitas tersebut dapat diperoleh dengan melakukan analisis butir untuk mengetahui kualitas butir pada instrumen, salah satunya dengan melakukan analisis tingkat kesukaran atau *difficulty index* (Suryana, 2015). Menurut Embretson & Reise (2000), tingkat kesukaran ideal untuk instrumen skrining atau diagnostik bergantung pada berbagai faktor, seperti karakteristik populasi yang diuji, tujuan penggunaan tes, dan jenis konstruk yang ingin diukur. Butir pernyataan ideal memiliki tingkat kesulitan seimbang, tidak terlalu mudah dan tidak terlalu sulit (Amalia & Widayati, 2012). Butir pernyataan yang ideal adalah jenis butir yang dirancang dengan cermat hingga memiliki tingkat kesulitan seimbang untuk dijawab partisipan. Dengan tingkat kesulitan yang seimbang, tes dapat memberikan informasi lebih akurat tentang tingkat pemahaman atau kemampuan partisipan. Jika butir terlalu mudah, maka tidak ada perbedaan yang signifikan dalam performa peserta yang berbeda sehingga sulit untuk membedakan antara tingkat kemampuan mereka. Di sisi lain, jika butir terlalu sulit, banyak peserta mungkin gagal

menjawab dan penyelenggara tes akan kesulitan dalam membedakan kemampuan di antara peserta tes. Dengan tingkat kesulitan yang seimbang, tes dapat lebih tepat mengukur rentang kemampuan peserta, membantu mengidentifikasi peserta yang unggul, yang memerlukan perhatian, atau yang memerlukan bantuan tambahan dalam pemahaman materi yang diuji. Tingkat kesulitan yang seimbang juga memberikan pengalaman yang lebih adil bagi semua peserta, memastikan bahwa mereka diuji secara proporsional sesuai dengan kemampuan mereka. Oleh karena itu, dalam merancang tes perlu mempertimbangkan dan memastikan tingkat kesulitan yang tepat demi mencapai tujuan evaluasi yang lebih baik dan informasi yang lebih berharga tentang peserta.

*Self-Reported Questionnaire* (SRQ-20) adalah sebuah instrumen skrining atau penapisan terhadap gangguan mental yang terdiri dari 20 pernyataan (Netsereab dkk., 2018). Penggunaan skrining dengan SRQ-20 sangat luas karena mudah digunakan dan membutuhkan waktu yang singkat (Sunjaya dkk., 2022). SRQ-20 dikembangkan untuk skrining gangguan non-psikotik sebagai bagian dari studi kolaboratif WHO tentang strategi perluasan perawatan kesehatan mental, terutama di negara-negara berkembang (Husain dkk., 2016). SRQ-20 umumnya diisi oleh individu sendiri, di mana mereka merespons pertanyaan-pertanyaan dengan mengacu pada pengalaman dan keadaan diri mereka sendiri. Tujuan perancangan SRQ-20 adalah untuk mengidentifikasi potensi gangguan mental umum maupun spesifik, seperti depresi, kecemasan, dan masalah emosional lainnya (Chua dkk., 2022; Dharmayanti dkk., 2018). SRQ-20 telah digunakan dalam berbagai penelitian dan konteks klinis sebagai alat awal untuk mengidentifikasi individu yang mungkin memerlukan evaluasi lebih lanjut terkait kesehatan mental mereka.

SRQ-20 memiliki lima faktor: energi, kognitif, depresi, fisiologis, dan kecemasan (Prasetio dkk., 2022). Faktor energi mencerminkan beberapa ciri orang yang mengalami penurunan energi fisik dan mental. Faktor kognitif menggambarkan beberapa tanda-tanda mengenai masalah kognitif, seperti penurunan kualitas hasil kerja, kehilangan minat, dan kesulitan dalam mengambil keputusan. Faktor depresi menggambarkan beberapa gejala emosional dan kognitif yang terkait dengan depresi, seperti perasaan tidak bahagia dan merasa tidak berharga, serta pemikiran untuk bunuh diri. Faktor fisiologis dicirikan oleh pengaruh gangguan psikologis pada sistem pencernaan. Faktor kecemasan menggambarkan beberapa gejala fisik dan psikologis, seperti perasaan tegang atau takut, dan tangan gemeteran. Mengingat tingginya prevalensi gangguan mental, penting untuk memiliki instrumen pengukuran yang akurat dan efektif guna mengidentifikasi gejala gangguan mental. Penggunaan instrumen ini bermanfaat dalam meningkatkan pencegahan dan/atau pengobatan yang dapat segera dilakukan pada individu yang memiliki permasalahan mental, sehingga mereka dapat mengembangkan potensi mereka dengan lebih baik.

Terdapat beberapa penelitian terdahulu untuk mengidentifikasi kualitas SRQ-20. Penelitian Prasetio dkk. (2022) bertujuan untuk menginvestigasi struktur faktor SRQ-20 serta mengumpulkan properti psikometrisnya. Model SRQ-20 satu faktor menyatakan bahwa variabel yang diukur adalah gangguan mental umum, sementara model lima faktor menunjukkan adanya faktor spesifik yang dapat membantu psikolog pada tindakan yang lebih terarah. Kedua model dapat digunakan sebagai referensi untuk interpretasi. Pada model lima faktor, masih dibutuhkan analisis lebih lanjut dan penggunaan sampel populasi lebih luas. Stratton dkk. (2013) telah mengevaluasi properti psikometris dari SRQ-20 dalam survei epidemiologi pada orang dewasa di Vietnam. Hasil penelitian Stratton dkk. (2013) menyebutkan bahwa *plot scree* dan hasil EFA, dua struktur laten dianggap layak dan kemudian dikenakan pemodelan lebih lanjut. Model bi-faktor (BFM) dan solusi model tiga faktor terkorelasi (*Negative Affect*, *Somatic Complaints*, dan *Hopelessness*) memberikan hasil yang memadai. BFM menentukan faktor Dominan Distress Umum tunggal (semua *item* SRQ-20) dengan faktor kelompok ortogonal untuk subset *item*: *Negative Affect* (9 butir), *Somatic Complaints* (8 butir), dan *Hopelessness* (3 butir). Model ini sesuai

dengan data atau bahkan lebih baik daripada model tiga faktor. Netsereab dkk. (2018) menyebutkan bahwa SRQ-20 memiliki reliabilitas internal yang baik ( $\alpha=0,78$ ). Analisis faktor menghasilkan dua faktor yang menjelaskan 31,2% dari varians total. Instrumen ini memiliki performa baik dalam mendeteksi gangguan mental umum, dengan area di bawah kurva (AUC) sebesar 0,879 (SE=0,23, 95% CI 0,83–0,92) pada sampel keseluruhan, dengan skor pemotongan optimal pada 5/6 dengan sensitivitas 78,6% dan spesifisitas 81,5%. Skor pemotongan berbeda antara perempuan (5/6) dan laki-laki (4/5). Untuk partisipan laki-laki, statistik AUC adalah 0,877 (SE=0,04, 95% CI 0,79–0,96) dan 0,871 (SE = 0,02, 95% CI 0,81–0,92) untuk partisipan perempuan.

Dari beberapa penelitian yang pernah dilakukan untuk menguji kualitas properti psikometris SRQ-20, penelitian menggunakan model Rasch belum banyak dilakukan. Tujuan dari analisis Rasch bukanlah untuk mengidentifikasi dimensi-dimensi yang muncul dalam data sebagai akibat dari adanya varians bersama antar indikator, tetapi untuk menguji dan mengukur karakteristik psikometris dari instrumen pengukuran yang bergantung pada asumsi bahwa probabilitas seseorang untuk berhasil menjawab suatu butir terkait dengan kemampuan individu dan tingkat kesulitan butir tersebut (Stolt dkk., 2022; Widhiarso, 2018). Dalam konteks penelitian yang menggunakan analisis model Rasch pada instrumen SRQ-20, tujuannya adalah menguji validitas dan kualitas psikometris SRQ-20 menggunakan pendekatan Rasch yang menghasilkan tingkat kesulitan butir dalam unit logit (Urbayatun & Widhiarso, 2012). Sebagai bagian dari *Item Response Theory*, analisis model Rasch memiliki beberapa kelebihan dibandingkan dengan analisis model klasik. Pertama, ketergantungan pada karakteristik kelompok. Menurut Magno (2009) dan Steyer (2001), ketergantungan pada karakteristik kelompok dalam teori tes klasik berarti diasumsikan bahwa kelompok yang diuji adalah cerminan dari seluruh populasi yang ingin diukur. Generalisasi hasil dapat ditingkatkan karena hasil yang diperoleh lebih spesifik untuk masing-masing individu, bukan hanya terbatas pada kelompok uji tertentu (Steyer, 2001). Kedua, adanya kesulitan dalam memperkirakan parameter butir. Teori tes klasik menghadapi permasalahan dalam mengestimasi dengan tepat tingkat kesulitan suatu *item*. Perkiraan ini bergantung pada kinerja kelompok yang tidak selalu mencerminkan tingkat kesulitan aktual butir tersebut secara akurat (Magno, 2009).

Berdasarkan uraian yang telah disampaikan, tujuan peneliti adalah untuk mengetahui kualitas SRQ-20 yang dikembangkan oleh WHO dengan melakukan analisis menggunakan pendekatan Rasch. Meskipun dianggap sebagai pendekatan yang ketat, pemodelan Rasch masih dapat dipengaruhi oleh interaksi individu yang mengerjakan dengan skala, serta heterogenitas karakteristik individu (Preinerstorfer & Formann, 2011; Widhiarso, 2013). Penelitian analisis kualitas SRQ-20 dengan pendekatan Rasch pada kelompok usia remaja di Indonesia masih belum banyak dilakukan sehingga peneliti memandang hal ini perlu untuk diuji.

## METODE

### *Desain Penelitian*

Penelitian ini menggunakan pendekatan kuantitatif, yakni analisis model Rasch. Pengumpulan data dilakukan dengan menggunakan bantuan Google Form. Pengumpulan data berlangsung dari tanggal 30 Agustus 2022 hingga 5 November 2022.

### *Partisipan*

Total partisipan adalah 165 Warga Negara Indonesia, dengan rincian 113 perempuan dan 52 laki-laki, dengan rentang usia 12 – 44 tahun. Prosedur pengambilan sampel menggunakan metode *convenience sampling*, yaitu metode sampling non probabilitas dimana partisipan merupakan individu yang



berkehendak mengerjakan atas keputusan sendiri setelah mendapatkan akses pengisian skala ini (Firmansyah & Dede, 2022; Stratton, 2021).

Dalam laman Google Form, peneliti memberikan penjelasan mengenai *informed consent* sebelum pengisian identitas. Peneliti menuliskan bahwa partisipan penelitian diperkenankan menuliskan nama lengkap atau inisial, serta seluruh data terkumpul akan dijaga kerahasiaannya dan hanya digunakan untuk kepentingan penelitian. Partisipan diperkenankan menghubungi peneliti melalui kontak yang tertera apabila memiliki pertanyaan. Partisipan diizinkan untuk tidak melanjutkan pengisian apabila partisipan merasa kurang nyaman selama pengisian skala.

#### *Pengukuran*

*Self-Reported Questionnaire* (SRQ-20) adalah instrumen yang dikembangkan oleh WHO dan terdiri atas 20 pernyataan dengan bentuk respons jawaban ya/tidak (WHO, 1994). Setiap jawaban benar mendapatkan skor 1 dan skor 0 untuk jawaban salah. Secara teoretis, nilai maksimum dari SRQ-20 adalah 20 poin dan nilai minimum adalah 0. Semakin tinggi nilai, menunjukkan adanya peningkatan jumlah gejala yang dialami (Fatahya & Abidin, 2017).

#### *Analisis Data*

Analisis data dilakukan dengan menggunakan analisis Rasch untuk mendapatkan nilai reliabilitas, *item fit*, *Wright Map*, serta dimensionalitas skala. Analisis data dilakukan dengan bantuan software Jamovi versi 2.3.18.

## HASIL PENELITIAN

Hasil analisis Rasch dari Jamovi menunjukkan *item statistic* untuk *proportion*, *measure*, *SE measure*, *infit*, dan *outfit*. Rincian *output* dapat dilihat pada tabel berikut.

Tabel 1. *Item Statistic*

	<b>Proportion</b>	<b>Measure</b>	<b>S.E. Measure</b>	<b>Infit</b>	<b>Outfit</b>
1	0,448	0,352	0,200	1,502	1,941
2	0,327	1,185	0,210	1,212	1,421
3	0,406	0,635	0,202	1,281	1,469
4	0,418	0,554	0,202	0,998	0,991
5	0,164	2,587	0,256	1,200	1,287
6	0,533	-0,208	0,200	0,924	0,940
7	0,236	1,895	0,228	1,042	1,050
8	0,376	0,842	0,205	0,907	0,808
9	0,364	0,926	0,206	0,843	0,665
10	0,212	2,109	0,235	1,213	1,307
11	0,358	0,968	0,206	0,921	0,781
12	0,497	0,032	0,200	0,854	0,708
13	0,291	1,455	0,215	0,848	0,760
14	0,309	1,318	0,212	0,705	0,477
15	0,291	1,455	0,215	0,747	0,523
16	0,358	0,968	0,206	0,835	0,697
17	0,176	2,460	0,249	0,859	0,706

	Proportion	Measure	S.E. Measure	Infit	Outfit
18	0,376	0,842	0,205	0,899	0,697
19	0,364	0,926	0,206	1,184	1,258
20	0,600	-0,656	0,204	0,976	0,942

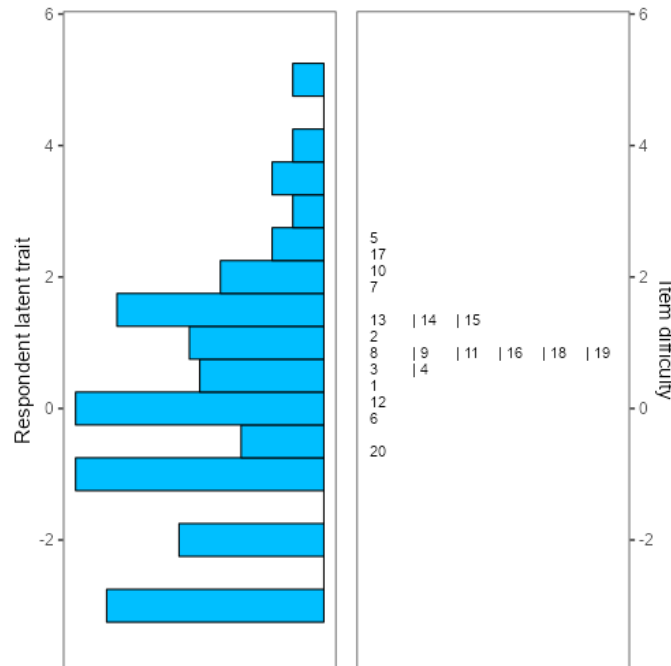
Note. Infit: Information-weighted mean square statistic; Outfit: Outlier-sensitive means square statistic.

Berdasarkan hasil analisis, SRQ-20 memiliki nilai reliabilitas yang baik (0,823) dan nilai *person separation reliability* sebesar 0,837. Menurut Widhiarso (2013), nilai reliabilitas tersebut dapat diterima karena berada pada rentang  $0,81 \leq p < 0,90$  (bagus).

Partisipan mendapatkan skor dengan rentang 2 sampai 20, dengan mean 7,1 dan SD sebesar 5,83. Nilai *infit* berkisar dari 0,705 hingga 1,502. Seluruh butir fit karena tidak ditemukan nilai INFIT MNSQ > 2. Hasil analisis OUTFIT MNSQ sebesar 0,477 hingga 1,941 dan ditemukan 2 butir pernyataan yang tidak valid karena memiliki nilai OUTFIT MNSQ > 1,5 yaitu butir nomor 1 dengan nilai 1,941 dan butir nomor 14 dengan nilai 0,477. Kesesuaian butir mengacu pada sejauh mana butir-butir individual dalam tes atau instrumen sejalan dengan model Rasch yang digunakan untuk mengukur kemampuan individu. *Item fit* berhubungan dengan residual, yang merupakan selisih antara respons aktual dengan respons yang diharapkan. Respons yang diharapkan dihitung berdasarkan probabilitas respons tertentu berdasarkan kemampuan individu dan tingkat kesulitan butir. Menurut Müller (2020), *item fit statistics* menyajikan ringkasan dari residual yang sudah distandardisasi untuk semua butir dalam tes atau instrumen. Statistik ini memberikan ukuran tentang seberapa baik respons yang diamati cocok dengan respons yang diharapkan berdasarkan model Rasch dan dapat digunakan untuk mengukur kemampuan subjeknya. Jika butir-butir dalam tes memiliki *item fit* yang baik, hal itu menandakan bahwa tes tersebut secara keseluruhan sesuai dengan model Rasch dan dapat diandalkan untuk mengukur kemampuan individu dengan akurat (Linda Crocker, 1986; Müller, 2020). Sebaliknya, jika pola respons yang diamati tidak sesuai dengan yang diharapkan menunjukkan adanya masalah pada butir tersebut, tidak cocok dengan model Rasch dan perlu direvisi atau dihapus (Kurniawan & Andriyani, 2018; Linacre, 2011).

Nilai *proportion* berkisar dari 0,164 sampai 0,6. Proporsi nilai ini berfungsi untuk menilai sejauh mana data yang diperoleh sesuai dengan asumsi model Rasch yang sedang digunakan. Menurut Veas dkk. (2016), nilai *measure* dan proporsi dalam analisis model Rasch saling berhubungan dan keduanya memiliki peranan penting dalam memastikan akurasi dan validitas hasil pengukuran. *Measure* yang baik akan menghasilkan proporsi yang baik pula. Nilai *measure* yang baik menunjukkan bahwa partisipan memiliki kemampuan yang baik dalam menjawab pertanyaan sehingga cenderung memberikan jawaban yang sesuai dengan model Rasch. Hal ini akan meningkatkan proporsi nilai yang sesuai dengan model Rasch sehingga data yang diperoleh lebih sesuai dengan asumsi model Rasch. Proporsi yang baik akan menghasilkan *measure* yang baik. Proporsi nilai yang baik menunjukkan bahwa data yang diperoleh sesuai dengan asumsi model Rasch sehingga dapat meningkatkan akurasi estimasi parameter butir dan kemampuan responden. Hal ini akan menghasilkan nilai *measure* yang lebih baik dan pada gilirannya memperbaiki kualitas pengukuran secara keseluruhan.

Analisis selanjutnya adalah melihat hasil pengujian dengan Wright Map yang tujuannya untuk mengetahui butir pernyataan yang sulit dan mudah dijawab oleh partisipan.



Gambar 1. Wright Map

Berdasarkan grafik tersebut diketahui bahwa tidak ada butir pernyataan yang lebih rendah dari nilai -2 logit. Terdapat butir pernyataan yang memiliki nilai lebih dari 2 logit yaitu nomor 10, 17, dan 5. Artinya, pernyataan pada nomor ini sulit untuk dijawab dengan benar. Menurut Linacre (2009, 2011), kemampuan individu dan butir direpresentasikan sebagai titik atau garis pada Wright Map. Lokasi individu pada sumbu x menunjukkan tingkat kemampuan, sedangkan lokasi butir pada sumbu y menunjukkan tingkat kesulitan butir. Jarak antara titik individu dan garis butir menggambarkan sejauh mana individu dapat menjawab dengan benar pada butir tersebut (Alagumalai, S. dkk., 2005; Boone dkk., 2014). Jika kemampuan subjek melebihi tingkat kesulitan butir, maka subjek dianggap mampu menjawab butir tersebut dengan benar. Namun, jika kemampuan subjek lebih rendah dari tingkat kesulitan butir, maka subjek dianggap tidak mampu menjawab butir tersebut (Hilaliyah dkk., 2019).

## DISKUSI

Yessimov dkk., (2023) menyebutkan bahwa model Rasch adalah suatu model probabilistik yang digunakan untuk mempelajari data kategorial dengan memperhitungkan hubungan antara kemampuan peserta uji dan tingkat kesulitan butir. Koefisien reliabilitas yang diperoleh dalam penelitian ini cukup tinggi, yaitu sebesar 0,823. Meskipun begitu, tidak terdapat angka tertentu yang dapat dianggap sebagai tingkat koefisien reliabilitas yang memuaskan. Penelitian sebelumnya mengenai kualitas tes SRQ-20 yang dilakukan di *primary health care* di Eritrea menghasilkan koefisien reliabilitas sebesar 0,78 (Netsereab dkk., 2018). Sementara itu Scholte (2013) menyebutkan bahwa koefisien reliabilitas instrumen SRQ-20 yang digunakan di Rwanda sebesar 0,85 untuk kelompok perempuan dan 0,81 untuk laki-laki. Kröner dkk. (2017) menyebutkan hasil analisis reliabilitas SRQ-20 dari partisipan remaja di Jerman memberikan hasil koefisien sebesar 0,88. Koefisien reliabilitas yang diperoleh dalam penelitian ini menghasilkan angka yang tidak jauh berbeda dengan penelitian-penelitian sebelumnya. Semakin tinggi koefisien reliabilitas, semakin besar kepercayaan yang dapat diberikan terhadap hasil tes tersebut (Azwar, 2016). Tes standar yang memiliki konsekuensi besar (*high stakes*) diharapkan

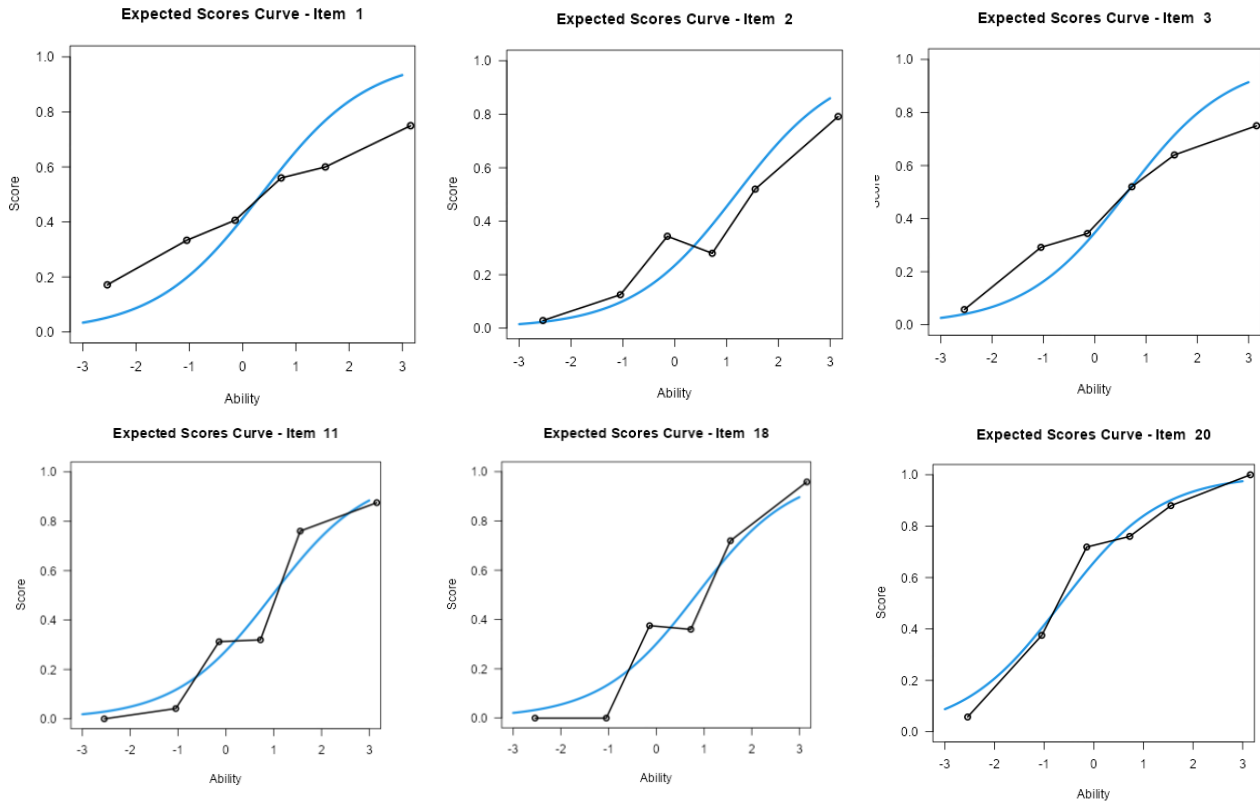
memiliki nilai reliabilitas sebesar 0,9, sementara untuk level lebih rendah diharapkan memiliki nilai reliabilitas 0,8 atau 0,85 (Santamaría-Vázquez dkk., 2023; Wells & Wollack, 2003). Kesimpulannya, SRQ-20 adalah instrumen skrining untuk mengetahui kemungkinan adanya gangguan mental dan memiliki reliabilitas yang baik dan mampu menghasilkan data yang tepercaya.

Selain itu, untuk mengetahui butir yang tidak sesuai dapat dilihat dari nilai *mean square* (MNSQ), dimana ketepatan butir dapat dilihat dari butir ideal melalui nilai MNSQ (Linacre, 2009). Pedoman penilaian kriteria kesesuaian butir menurut nilai outfit MNSQ adalah  $0,5 < \text{MNSQ} < 1,5$  (Ramdani dkk., 2020). Linacre (2009) menjelaskan bahwa butir MNSQ berada dalam kisaran 0,5 hingga 1,5. Selain itu, dapat juga diketahui dengan menjumlahkan nilai rata-rata infit kuadrat tengah (*mean MNSQ Infit*) (Sumintono & Widhiarso, 2013). Hasil analisis OUTFIT MNSQ sebesar 0,477 hingga 1,941. Nilai *proportion* merujuk pada persentase atau proporsi tingkat kesesuaian antara respons atau jawaban yang diberikan oleh peserta dalam uji atau kuesioner dengan model Rasch yang sedang digunakan (Suminto & Wening, 2022).

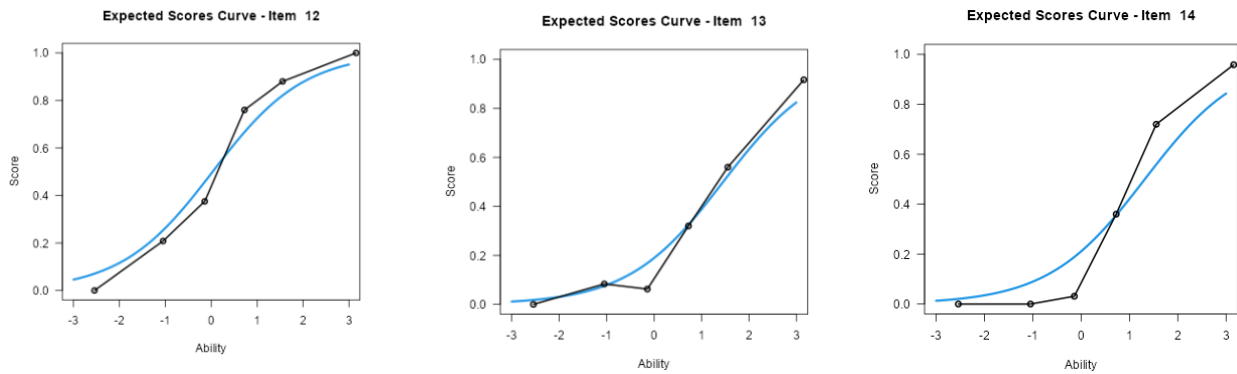
Grafik Wright Map adalah visualisasi yang digunakan dalam Rasch Model untuk menggambarkan tingkat kemampuan individu dan tingkat kesulitan butir pada suatu skala pengukuran. Grafik ini menggambarkan lokasi relatif individu dan butir pada suatu sumbu tunggal. Berdasarkan hasil yang muncul dan dapat dilihat pada gambar 1, diketahui bahwa butir pernyataan nomor 5, 10, dan 17 termasuk butir yang sulit untuk dijawab. Menurut Prasetio dkk. (2022), butir pernyataan nomor 5 berasal dari faktor kecemasan dengan bunyi, "*apakah tangan Anda gemetar?*". Butir pernyataan nomor 10 tidak termasuk dalam faktor mana pun, berbunyi, "*apakah Anda lebih sering menangis dari biasanya?*". Butir pernyataan nomor 17 berasal dari faktor depresi dengan bunyi pernyataan, "*apakah Anda memiliki pemikiran untuk mengakhiri hidup?*". Dalam konteks pengukuran SRQ-20, perolehan nilai tinggi individu menunjukkan kemungkinan adanya masalah atau gangguan kesehatan mental. Nilai rendah pada SRQ-20 mengindikasikan bahwa partisipan memiliki sedikit atau bahkan tidak ada gejala kesehatan mental yang terukur oleh kuesioner tersebut (van der Westhuizen dkk., 2016; WHO, 1994). Hal ini mengartikan bahwa individu tersebut mungkin saat ini tidak memiliki masalah kesehatan mental, atau gejala yang dialami sangat ringan sehingga tidak dapat terdeteksi.

#### *Item Characteristic Curve*

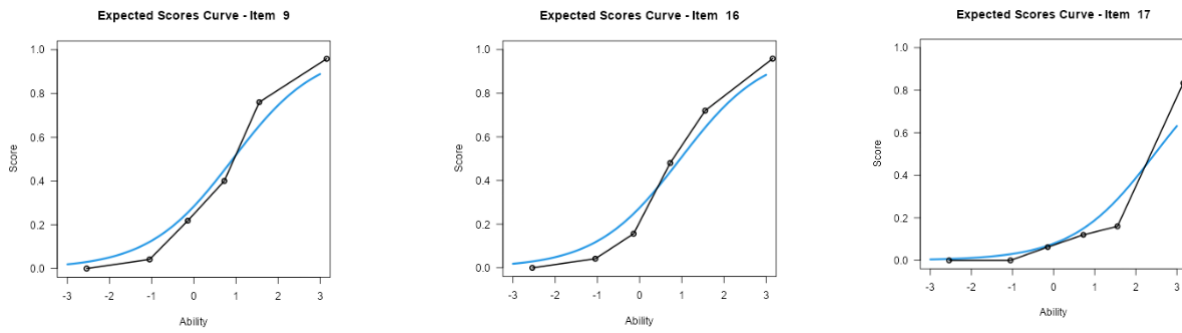
*Item Characteristic Curve (ICC)* dalam Rasch model adalah kurva berbentuk logit yang menggambarkan hubungan antara kemampuan individu yang diukur dan probabilitas mereka untuk menjawab dengan benar pada suatu butir tes tertentu (Nguyen dkk., 2014). ICC merupakan salah satu komponen penting dalam Rasch model yang digunakan untuk mengukur dan memvalidasi instrumen pengukuran. Kurva ini dapat membantu mengidentifikasi butir yang terlalu mudah atau terlalu sulit bagi populasi yang dituju, serta butir yang tidak dapat membedakan antara peserta dengan tingkat kemampuan yang berbeda. Kurva ini memiliki bentuk menyerupai huruf S karena probabilitas jawaban benar cenderung rendah pada tingkat keterampilan yang sangat rendah dan sangat tinggi, tetapi meningkat tajam pada tingkat keterampilan menengah. Setiap butir akan memiliki bentuk yang berbeda satu sama lain. Output hasil analisis Rasch berupa *Item Characteristic Curve (ICC)* dapat dilihat pada gambar berikut. Setiap grafik memiliki sumbu x dan y. Sumbu x menunjukkan tingkat kemampuan. Sumbu y menunjukkan probabilitas menjawab benar pada setiap butir. Kurva berwarna hitam menunjukkan ICC empiris dan kurva berwarna biru menunjukkan ICC teoretis. Kemiringan yang ditunjukkan oleh grafik berwarna biru menunjukkan bahwa butir pernyataan memiliki sensitivitas yang baik untuk mengukur kemampuan individu.



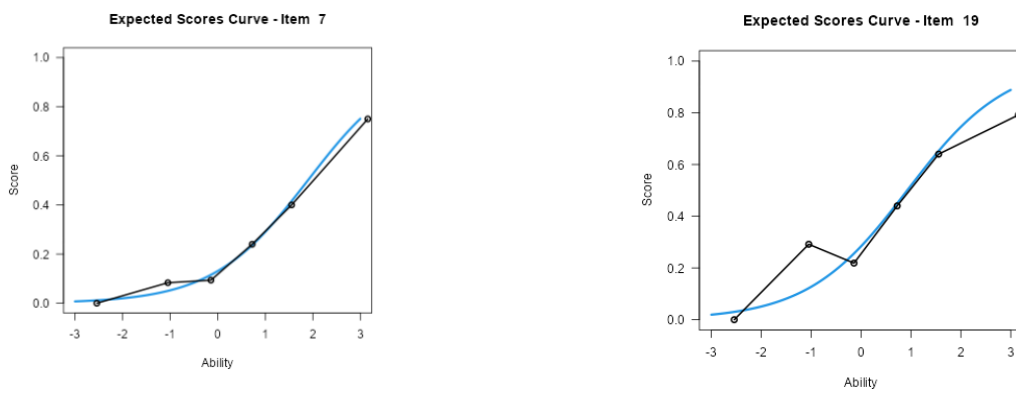
Gambar 2. Item Characteristic Curve Faktor 1: Energi



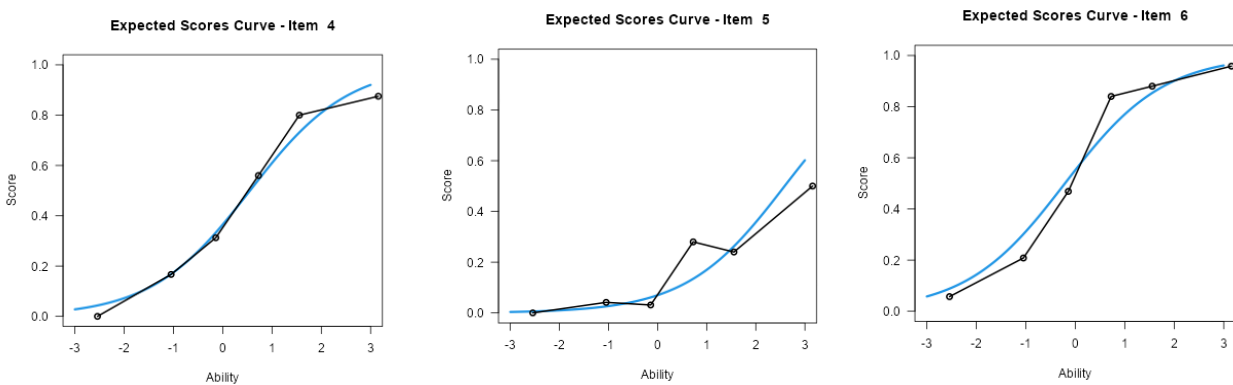
Gambar 3. Item Characteristic Curve Faktor 2: Kognitif



Gambar 4. Item Characteristic Curve Faktor 4: Depresi

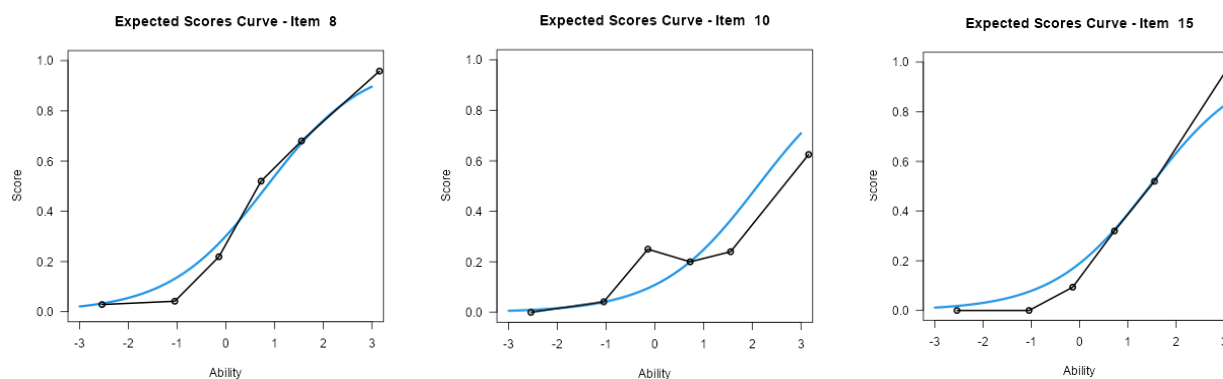


Gambar 5. Item Characteristic Curve Faktor 4: Fisiologis



Gambar 6. Item Characteristic Curve Faktor 5: Kecemasan





Gambar 7. *Item Characteristic Curve* Tanpa Faktor

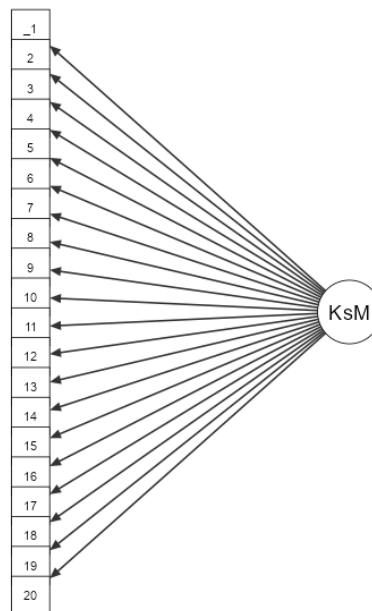
Pada gambar 2 terdapat enam grafik *Item Characteristic Curve* (ICC) yang dapat dilihat. Grafik ICC pada butir 1 menunjukkan pola grafik yang kurang berhimpit. Butir nomor 1 memiliki nilai Infit dan Outfit MNSQ sebesar 1,502 dan 1,941 dimana batasnya adalah  $0,5 < \text{MNSQ} < 1,5$ , dapat disimpulkan bahwa butir nomor 1 kurang fit karena nilai output melebihi 1,5. Selanjutnya, butir pernyataan nomor 14 memiliki nilai infit dan outfit MNSQ sebesar 0,705 dan 0,477 dimana butir tersebut memiliki nilai outfit yang berada di bawah standar 0,5 sehingga kurang fit. Gambar ICC untuk butir nomor 14 dapat dilihat pada gambar 3 dimana kurva kurang berhimpit. Terdapat keterkaitan antara ICC, infit, dan outfit dalam analisis Rasch. Dalam konteks SRQ-20, ICC akan memberikan gambaran tentang tingkat kesulitan dan diskriminasi masing-masing butir pada kuesioner tersebut. Setelah itu, infit dan outfit digunakan untuk mengevaluasi sejauh mana respons peserta pada masing-masing butir sesuai dengan model Rasch dan sejauh mana sensitivitas butir terhadap variasi kemampuan peserta. Jika butir memiliki ICC yang baik (tingkat kesulitan dan diskriminasi yang sesuai), maka diharapkan nilai infit dan outfit juga akan baik. Sebaliknya, jika *item* tes memiliki ICC yang buruk (tingkat kesulitan dan diskriminasi yang tidak sesuai), maka nilai infit dan outfit cenderung buruk pula.

#### *Differential Item Functioning (DIF)*

Berdasarkan hasil analisis, diketahui terjadi DIF pada butir nomor 8 ( $p=0,001$ ), nomor 10 ( $p<0,001$ ), dan nomor 18 ( $p=0,028$ ). Terdapat 3 butir yang terindikasi DIF. DIF atau bias pengukuran terjadi saat individu dari kelompok yang berbeda memberikan tanggapan yang berbeda terhadap suatu butir, sehingga mengakibatkan variasi dalam parameter yang diestimasi dalam tingkat kemampuan yang sama (Nguyen dkk., 2014).

#### *Unidimensionalitas*

Uji dimensionalitas yang dilakukan oleh peneliti menggunakan Jamovi tidak menunjukkan informasi mengenai unidimensionalitas. Peneliti melakukan analisis unidimensionalitas dengan menggunakan *confirmatory factor analysis* (CFA). Hasil analisis menunjukkan bahwa nilai CFI sebesar 0,872 dan RMSEA sebesar 0,078.



Gambar 8. *Path Diagram CFA*

Berdasarkan gambar 8 dapat dilihat bahwa seluruh butir dalam SRQ-20 berada dalam satu konstruk yang sama dan diberi label faktor kesehatan mental (KsM) oleh peneliti. Butir-butir dalam SRQ-20 terkonfirmasi memenuhi asumsi unidimensionalitas.

## SIMPULAN

Dengan menggunakan model Rasch, studi ini berfokus untuk menganalisis kualitas SRQ-20 yang merupakan instrumen skrining pengukuran kesehatan mental. Model Rasch adalah suatu model probabilistik yang digunakan untuk mempelajari data kategorikal dengan memperhitungkan hubungan antara kemampuan peserta uji dan tingkat kesulitan butir. Semakin tinggi kemampuan seseorang dibandingkan dengan tingkat kesulitan butir, semakin besar kemungkinan orang tersebut menjawab benar. Hasil penelitian ini menunjukkan bahwa secara umum SRQ-20 memiliki reliabilitas yang baik dan butir-butir yang fit dengan model sebanyak 18. Penelitian ini masih memiliki keterbatasan populasi sehingga akan lebih baik lagi apabila sampel yang digunakan lebih luas dengan jumlah yang lebih banyak. Terdapat analisis DIF yang dilakukan namun terdapat kemungkinan ketidakakuratan hasil karena sampel yang terbatas.

## ACKNOWLEDGEMENTS / UCAPAN TERIMA KASIH

Praise be to God Almighty who has provided opportunities and help in every step taken in the process of this research and writing this article. I would also like to thank my family and colleagues who provided support. / *Puji syukur atas kehadiran Tuhan Yang Maha Esa yang telah memberikan kesempatan dan pertolongan dalam setiap langkah yang ditempuh peneliti dalam proses penelitian dan penulisan artikel jurnal. Peneliti juga mengucapkan terima kasih kepada keluarga dan rekan-rekan yang memberikan dukungan dalam bentuk apa pun sehingga penelitian berjalan lancar.*

**DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST / DEKLARASI  
POTENSI TERJADINYA KONFLIK KEPENTINGAN**

Nurul Fajriyah Prahastuti tidak bekerja, menjadi konsultan, memiliki saham, atau menerima dana dari perusahaan atau organisasi manapun yang mungkin akan mengambil untung dari diterbitkannya naskah ini. / *Nurul Fajriyah Prahastuti tidak bekerja, menjadi konsultan, memiliki saham, atau menerima dana dari perusahaan atau organisasi manapun yang mungkin akan mengambil untung dari diterbitkannya naskah ini.*

**REFERENCES / PUSTAKA ACUAN**

- Alagumalai, S., Curtis, D. D., & Hungi, N. (2005). *Applied Rasch Measurement: A Book of Exemplars*. Springer.
- Amalia, A. ., & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, 10(1), 1–26.
- Ayuningtyas, D., Misnaniarti, M., & Rayhani, M. (2018). Analisis Situasi Kesehatan Mental Pada Masyarakat Di Indonesia Dan Strategi Penanggulangannya. *Jurnal Ilmu Kesehatan Masyarakat*, 9(1), 1–10. <https://doi.org/10.26553/jikm.2018.9.1.1-10>
- Azwar, S. (2016). *Reliabilitas dan Validitas* (4th ed.). Pustaka Pelajar.
- Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. In *Rasch Analysis in the Human Sciences*. <https://doi.org/10.1007/978-94-007-6857-4>
- Chua, K.-C., Hahn, J. S., Farrell, S., Jolly, A., Khangura, R., & Henderson, C. (2022). Mental health literacy: A focus on daily life context for population health measurement. *SSM - Mental Health*, 2(October 2021), 100118. <https://doi.org/10.1016/j.ssmmh.2022.100118>
- Dharmayanti, I., Tjandrarini, D. H., Hidayangsih, P. S., & Nainggolan, O. (2018). Pengaruh Kondisi Kesehatan Lingkungan Dan Sosial Ekonomi Terhadap Kesehatan Mental Di Indonesia. *Jurnal Ekologi Kesehatan*, 17(2), 64–74. <https://doi.org/10.22435/jek.17.2.149.64-74>
- Dianovinina, K. (2018). Depresi pada Remaja : Gejala dan Permasalahannya Depression in Adolescent : Symptoms and the Problems. *Jurnal Psikogenesis*, 6(1), 69–78.
- Embretson, S. E., & Reise, S. . (2000). *Item Response Theory for Psychologists*. Psychology Press. <https://doi.org/https://doi.org/10.4324/9781410605269>
- Fatahya, & Abidin, F. A. (2017). Higeia Journal of Public Health. *Higeia Journal of Public Health Research and Development*, 1(3), 625–634.
- Firmansyah, D., & Dede. (2022). Teknik Pengambilan Sampel Umum dalam Metodologi Penelitian: Literature Review. *Jurnal Ilmiah Pendidikan Holistik (JIPH)*, 1(2), 85–114. <https://doi.org/10.55927/jiph.v1i2.937>
- Handayani, T., Ayubi, D., & Anshari, D. (2020). Literasi Kesehatan Mental Orang Dewasa dan Penggunaan Pelayanan Kesehatan Mental. *Perilaku Dan Promosi Kesehatan: Indonesian Journal of Health Promotion and Behavior*, 2(1), 9. <https://doi.org/10.47034/ppk.v2i1.3905>

- Hilaliyah, H., Agustin, Y., Setiawati, S., Hapsari, S. N., Rangka, I. B., & Ratodi, M. (2019). Wright-Map to investigate the actual abilities on math test of elementary students. *Journal of Physics: Conference Series*, 1318(1). <https://doi.org/10.1088/1742-6596/1318/1/012067>
- Husain, N., Chaudhry, N., Rhouma, A., Sumra, A., Tomenson, B., & Waheed, W. (2016). Validation of the self-reporting questionnaire (SRQ 20) in British Pakistani and White European population in the United Kingdom. *Journal of Affective Disorders*, 189, 392–396. <https://doi.org/10.1016/j.jad.2015.08.068>
- Kemenkes RI. (2014). *Stop stigma dan diskriminasi terhadap orang dengan gangguan jiwa (ODGJ)*.
- Kröner, J., Goussios, C., Schaitz, C., Streb, J., & Sosic-Vasic, Z. (2017). The construct validity of the German academic Self-Regulation Questionnaire (SRQ-A) within primary and secondary school children. *Frontiers in Psychology*, 8(JUN). <https://doi.org/10.3389/fpsyg.2017.01032>
- Kurniawan, & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model. *Jurnal Statistika*, 6(1), 34–39.
- Lee, H. Y., Hwang, J., Ball, J. G., Lee, J., Yu, Y., & Albright, D. L. (2020). Mental health literacy affects mental health attitude: Is there a gender difference? *American Journal of Health Behavior*, 44(3), 283–291. <https://doi.org/10.5993/AJHB.44.3.1>
- Linacre, J. . (2009). Local independence and residual covariance: A study of olympic figure skating ratings. *Journal of Applied Measurement*, 10(2), 157–169.
- Linacre, J. M. (2011). *Winsteps Help for Rasch Analysis*. <http://homes.jcu.edu.au/~edtgb/%5Cnpapers3://publication/uuid/D56B724A-62FF-4D00-84E1-ECC888298B70>
- Linda Crocker, J. A. (1986). *Introduction to Classical and Modern Test Theory*. Harcourt Brace Jovanovich College.
- Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1–11.
- Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications*, 7(1). <https://doi.org/10.1186/s40488-020-00108-7>
- Netsereab, T. B., Kifle, M. M., Tesfagiorgis, R. B., Habteab, S. G., Weldeabzgi, Y. K., & Tesfamariam, O. Z. (2018). Validation of the WHO self-reporting questionnaire-20 (SRQ-20) item in primary health care settings in Eritrea 11 Medical and Health Sciences 1117 Public Health and Health Services. *International Journal of Mental Health Systems*, 12(1), 1–9. <https://doi.org/10.1186/s13033-018-0242-y>
- Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *Patient*, 7(1), 23–35. <https://doi.org/10.1007/s40271-013-0041-0>
- Nurhaeni, A., Erna Marisa, D., & Oktiany, T. (2022). Peningkatan Pengetahuan Tentang Gangguan Kesehatan Mental Pada Remaja. *Jurnal Pengabdian Masyarakat Kesehatan*, 1(1), 29–34. <http://download.garuda.kemdikbud.go.id/article.php?article=3234124&val=28314&title=PENIN>

## GKATAN PENGETAHUAN TENTANG GANGGUAN KESEHATAN MENTAL PADA REMAJA

- Prasetyo, C. E., Triwahyuni, A., & Prathama, A. G. (2022). Psychometric Properties of Self-Report Questionnaire-20 (SRQ-20) Indonesian Version. *Jurnal Psikologi*, 49(1), 69. <https://doi.org/10.22146/jpsi.69782>
- Preinerstorfer, D., & Formann, A. K. (2011). Parameter recovery and model selection in mixed Rasch models. *British Journal of Mathematical and Statistical Psychology*, No-No., 65. <https://doi.org/10.1111/j.2044-8317.2011.02020.x>
- Ramdani, R., Hanurawan, F., Ramli, M., Lasan, B. B., & Afdal, A. (2020). Development and Validation of Indonesian Academic Resilience Scale Using Rasch Models. *International Journal of Instruction*, 14(1), 105–120. <https://doi.org/10.29333/IJI.2021.1417A>
- Santamaría-Vázquez, M., Guijo-Blanco, V., Puente-Martínez, A., & Ubillos-Landa, S. (2023). Psychometric properties of the Spanish version of the Activities Scale for Kids (ASK): reliability, validity and the Rasch model. *BMJ Open*, 13(4). <https://doi.org/10.1136/bmjopen-2022-069248>
- Santoso, P. S., Utami, A. B., & Rista, K. (2023). *Kecenderungan gangguan pasca trauma pada penyintas covid-19 berusia dewasa akhir: Menguji peran dukungan sosial Pendahuluan*. 2(4), 808–814.
- Scholte, P. (2013). Psychometric properties and longitudinal validation of the Self-Reporting Questionnaire (SRQ-20) in a Rwandan community setting: a validation study. *Mental Health in War-Affected Populations*, 91–111.
- Steyer, R. (2001). Classical (Psychometric) Test Theory. *International Encyclopedia of the Social & Behavioral Sciences*, 1, 1955–1962. <https://doi.org/10.1016/B0-08-043076-7/00721-X>
- Stolt, M., Kottorp, A., & Suhonen, R. (2022). International Journal of Nursing Studies The use and quality of reporting of Rasch analysis in nursing research : A methodological scoping review. *International Journal of Nursing Studies*, 132, 104244. <https://doi.org/10.1016/j.ijnurstu.2022.104244>
- Stratton, K. J., Aggen, S. H., Richardson, L. K., Acierno, R., Kilpatrick, D. G., Gaboury, M. T., Tran, T. L., Trung, L. T., Tam, N. T., Tuan, T., Buoi, L. T., Ha, T. T., Thach, T. D., & Amstadter, A. B. (2013). Evaluation of the psychometric properties of the Self-Reporting Questionnaire (SRQ-20) in a sample of Vietnamese adults. *Comprehensive Psychiatry*, 54(4), 398–405. <https://doi.org/10.1016/j.comppsy.2012.10.011>
- Stratton, S. J. (2021). Population research: convenience sampling strategies. *Prehospital and Disaster Medicine*, 36(4), 373–374. <https://doi.org/10.1017/S1049023X21000649>
- Suminto, M. F., & Wening, S. (2022). The Rasch Model for Analysing the Indonesian Language Examination Instrument. *Proceedings of the 5th International Conference on Current Issues in Education (ICCIE 2021)*, 640(Iccie), 383–388. <https://doi.org/10.2991/assehr.k.220129.070>
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model Rasch untuk Penelitian Sosial*. Trim Komunikata Publishing House.
- Sunjaya, D. K., Sumintono, B., Gunawan, E., Herawati, M. D., & Hidayat, T. (2022). Online Mental Health Survey for Addressing Psychosocial Condition During the COVID-19 Pandemic in Indonesia : Instrument Evaluation Online Mental Health Survey for Addressing Psychosocial Condition During the COVID-19 Pandemic in Indonesia : Instrument Ev. *Psychology Research and Behavior*

*Management*, 15, 161–170. <https://doi.org/10.2147/PRBM.S347386>

- Suryana, Y. (2015). Re-analisis tingkat kebakan item tes intelegensi: Advanced progresive matrices. *Jurnal Pendidikan Pascasarjana Magister PAI*, 2(1), 1–21. <https://journal.unsika.ac.id/index.php/pendidikan/article/view/1033>
- Urbayatun, S., & Widhiarso, W. (2012). Variabel mediator dan moderator dalam penelitian psikologi kesehatan masyarakat. *Jurnal Psikologi*, 39(2), 180–188.
- van der Westhuizen, C., Wyatt, G., Williams, J. K., Stein, D. J., & Sorsdahl, K. (2016). Validation of the Self Reporting Questionnaire 20-Item (SRQ-20) for Use in a Low- and Middle-Income Country Emergency Centre Setting. *International Journal of Mental Health and Addiction*, 14(1), 37–48. <https://doi.org/10.1007/s11469-015-9566-x>
- Veas, A., Gilar, R., Miñano, P., & Castejón, J. L. (2016). Estimation of the proportion of underachieving students in compulsory secondary education in Spain: An application of the Rasch model. *Frontiers in Psychology*, 7(MAR). <https://doi.org/10.3389/fpsyg.2016.00303>
- Wells, C., & Wollack, J. (2003). *An Instructor's Guide to Understanding Test Reliability, Paper, Testing, & Evaluation Services*. University of Winconsin.
- WHO. (1994). *A user's guide to The Self-Report Questionnaire (SRQ)*. World Health Organization.
- Widhiarso, W. (2013). Aplikasi Model Rasch Campuran Dalam Mengevaluasi Pengukuran Harga Diri. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 17(1), 172–187. <https://doi.org/10.21831/pep.v17i1.1367>
- Widhiarso, W. (2018). Penggunaan Testlet dalam Pengembangan Tes Psikologi. *INSAN Jurnal Psikologi Dan Kesehatan Mental*, 3(1), 44. <https://doi.org/10.20473/jpkm.v3i12018.44-61>
- Xiao, Y., Yu, Y., Zhou, Z., Liu, W., Luo, Z., Tang, C., Xiao, P., Wen, S. W., Zeng, X., Zhong, W., & Shen, M. (2023). Hospital spending and length of hospital stay for mental disorders in Hunan, China. *Heliyon*, 9(4). <https://doi.org/10.1016/j.heliyon.2023.e14968>
- Yessimov, B., Hussein, R. A., Mohammed, A., & Hassan, A. Y. (2023). *Detecting Measurement Disturbances : Graphical Illustrations of Item Detecting Measurement Disturbances: Graphical Illustrations of Item Characteristic Curves*. May. <https://doi.org/10.22034/IJLT.2023.391731.1247>