PSIKODIMENSIA: Kajian Ilmiah Psikologi
P-ISSN:1411-6073|E-ISSN:2579-6321
Vol. 25, No. 1 (2026), 30-48; DOI:10.24167/psidim.v25i1.14451

Rasch Analysis of the Student Mental Health Scale (ISKM-R-2)

Lira Fessia Damaianti, Muhamad Arif Saefudin, Farhan Zakariyya, Helli Ihsan, Siti Chotidjah, Ghinaya Ummul Mukminin Hidayat, Aslama Tazkia, Atha Khalila Mustafid, Camila Azzahra Nurrohman, Nafisya Bulan Andari, Ni Kadek Ayu Regitha Pramesti

Psychology Department, Faculty of Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Abstract

Adolescent mental health is a critical issue, as this developmental stage is often marked by vulnerability to psychological problems, yet standardized measurement tools remain limited. This study aimed to evaluate the psychometric properties of the Student Mental Health Scale (ISKM-R-2) using the Rasch model. A methodological study design was adopted with a cross-sectional method, including 1,045 adolescents aged 15–19 years. The results suggested that the instrument had a multidimensional structure, with a two-factor model showing a better fit than the unidimensional model. Rasch analysis showed acceptable item fit and high reliability across dimensions. However, the instrument items were relatively difficult for the sample, as shown by the Wright Map analysis. The results of this study supported the importance of a multidimensional method in measuring student mental health and provided evidence for further refinement of the instrument.

Keywords: Multidimensionality, psychometrics, Rasch model, student mental health

Abstrak

Kesehatan mental remaja merupakan isu yang krusial, karena tahap perkembangan ini seringkali ditandai dengan kerentanan terhadap masalah psikologis, namun alat ukur standar yang tersedia masih terbatas. Penelitian ini bertujuan untuk mengevaluasi sifat psikometrik instrumen kesehatan mental remaja versi ke 2 (ISKM-R-2) menggunakan model Rasch. Penelitian ini menggunakan desain metodologis dengan pendekatan cross-sectional yang melibatkan 1.045 remaja berusia 15–19 tahun. Hasil penelitian menunjukkan bahwa instrumen memiliki struktur multidimensi, di mana model dua faktor memberikan kecocokan yang lebih baik dibandingkan model unidimensional. Analisis Rasch menunjukkan bahwa butir-butir instrumen memiliki kecocokan yang baik dan reliabilitas yang tinggi pada masing-masing dimensi. Namun, analisis Wright Map menunjukkan bahwa tingkat kesulitan butir relatif lebih tinggi dibandingkan karakteristik responden. Secara keseluruhan, temuan ini menegaskan pentingnya pendekatan multidimensi dalam pengukuran kesehatan mental pelajar serta menjadi dasar untuk pengembangan instrumen lebih lanjut.

Kata kunci*: Kesehatan mental pelajar; model rasch; multidimensionalitas; psikometri*

Corresponding author: Farhan Zakariyya | **E-mail:** farhanzakariyya@upi.edu
Received: November 12th, 2025 | **Accepted:** June 1st, 2026 | **Published:** June 22nd, 2026

1. Introduction

According to the World Health Organization (WHO, 2024), adolescence is a stage that occurs between ages 10 and 19, serving as a transition between childhood and adulthood. During this period, individuals experience many changes in social, psychological, and biological aspects. However, biological and physical development often precedes psychological or social development (Cronau & Brown, 1998). At this stage, individuals occupy an ambiguous social position between childhood and full adulthood, which can lead to internal conflicts between the desire for autonomy and the continuing need for support and guidance (Andriyani, 2020). Furthermore, the transition to adulthood exposes adolescents to diverse stressors, including social comparison, uncertainty about the future, and pressure to meet academic or societal expectations (Jones et al., 2021; Meggiolaro & Ongaro, 2014; Narmandakh et al., 2020). These cumulative challenges increase vulnerability to mental health problems, such as anxiety, depression, and emotional dysregulation. Consistent with this result, WHO (2024) reported that approximately 25% of adolescents experience mental health issues, showing adolescence as a critical period for the development of psychological difficulties. Therefore, understanding adolescent development cannot be separated from examining mental health, as the challenges inherent in this stage are closely related to the risk and manifestation of mental health problems.

The mental well-being of adolescents can be impacted during the transition to early adulthood. The psychological state is shaped not only by biological changes but also by social factors, such as family dynamics, peer relationships, and the school environment (Lamblin et al., 2017). As adolescents progress into adulthood, they need to adapt to novel situations and confront the challenges inherent in pursuing higher education. Achieving top rankings may foster a sense of accomplishment but also introduce additional stressors that can contribute to mental strain (MacIntyre et al., 2020). Particularly in the Indonesian context, the mental health landscape for adolescents presents unique cultural challenges. One in three adolescents experienced a mental health problem, while one in twenty met the full clinical criteria for a mental disorder (I-NAMHS, 2022). Approximately 95.4% of individuals in the age range 16-24 encountered anxiety, 88% grappling with depression symptoms, and 96.4% have reported inability to handle stress. According to previous studies, the stressors could be traced to academic pressure (Sari & Hazim, 2023), cultural stigma and literacy (Yani et al., 2025), and family dynamics (Sunarti et al., 2026).

The primary challenge by far in the current measurement of adolescent mental health is the lack of a validated, comprehensive, and context-sensitive instrument (Dubbeldeman et al., 2025). Several studies have pointed out persistent shortages of instruments that measure adolescent mental health and are explicitly based on adolescents' developmental traits and specific sociocultural backgrounds (Mansfield et al., 2020; Sequeira et al., 2022). Consequently, most studies and intervention practices used adapted instruments, which risk introducing measurement bias and may be less comprehensive in capturing adolescent mental health dimensions, including positive affective states, cognitive function, and psychosocial symptoms (Rose et al., 2017).

The prevalence of self-reported measures of mental health is also a separate challenge, as it is prone to subjective bias, inadequate understanding by the respondents, and the influence of social stigma on reporting psychological problems (Gagné et al., 2022). Although some studies have designed and tested the validity and reliability of a few instruments, the urgent need for more sensitive, multidimensional, and locally based measurement tools still exists (Hidayati et al., 2021). The information deficiency and measurement inadequacy constraints are certainly obstacles in acquiring an accurate understanding of adolescents' mental health. This situation leads ot the suboptimal development of evidence-based policies and interventions (Mansfield et al., 2020). Therefore, this study aimed to develop a measurement tool to collect comprehensive information on adolescents' mental health conditions in schools. Although efforts have been made to develop mental health assessment tools for students, recent studies (Chotidjah et al., 2024; Mukminin et al., 2024) still show significant psychometric limitations.

Currently available instruments report low explained variance (below 45%), indicating that student mental health constructs have not yet been comprehensively mapped. Furthermore, issues were found in the representation of certain dimensions, such as the 'isolation' factor, which had only two items with low reliability (0.627), as well as an over-reliance on index modifications to achieve model fit. These limitations suggested that the theoretical structure and item stability of the previous ISKM-R instrument were not yet robust enough for widespread practical use. Therefore, an in-depth re-evaluation using a more rigorous analytical method was necessary to address item redundancy and ensure greater construct stability. This study sought to develop a new instrument to assess mental health using the Sartorius method (Bhugra et al., 2013). Mental health is an integral aspect of human wellbeing, defined either as the absence of mental disorder that enables individuals to function effectively or as a state of balance within oneself and in relation to the physical and social environment. The core features of mental health include the ability to form affectionate relationships, manage emotions (e.g., sadness), and change constructively; a sense of worth; control and understanding of internal/ external functioning; and positive feelings toward self and others (Bhugra et al., 2013). The process for instrument development began with defining the construct, identifying key aspects and behavioral indicators for each aspect, and generating items based on those indicators and the blueprint. A total of 51 items were generated from this process and subsequently examined using Rasch analysis.

The selection of the Rasch model is based on the specific objectivity paradigm (Wright, 1977). Unlike descriptive IRT, which adapts the model to the unique characteristics of the data, the Rasch model is prescriptive and requires that the data meet measurement requirements. This process ensures that the respondent's total score is a sufficient statistic for estimating latent ability, thereby enabling the construction of a pure linear interval scale. The Rasch model is not only a 1-PL IRT formulation but an embodiment of Conjoint Measurement, used to transform ordinal data (raw scores) into interval units (logits), which constitute a mathematical prerequisite for further parametric statistical analysis. The application for analyzing scales related to adolescent mental health aims to improve the scales' validity and reliability and to identify biased items. By adopting this advanced analytical method, this study aimed to produce a comprehensive measurement tool for adolescent mental health. Furthermore, this study laid a strong foundation for refining adolescent mental health measurement tools for future assessments.

2. Method

A methodological study design was adopted, with a focus on psychometric validation. Specifically, the study adopted the Rasch Measurement Model to transform ordinal raw data into interval-level measures, ensuring the instrument met the requirements of fundamental measurement, such as unidimensionality and item-person invariance (Bond & Fox, 2015). A cross-sectional method was used for data collection, enabling a rigorous assessment of the scale's structural validity and reliability in the target adolescent population. The population consisted of adolescents enrolled in junior high school (grade 3) and senior high school (grades 1–3), aged 15 to 19 years, and purposive sampling was used to select the sample. A total of 1,045 respondents participated, fulfilling the minimum requirement of 1,000 individuals as recommended by Reise & Yu (1990) for Rasch model analysis. Among the respondents, the largest age group was 17 years old (40.9%), followed by 16 years old (36.7%). Gender distribution showed that 59.5% and 40.5% were female and male, respectively, as shown in Table 1.

The Student Mental Health Scale was developed based on the mental health framework proposed by Bhugra et al. (2013). The initial version consisted of 51 items across three dimensions, namely (1) inner peace (40 items), (2) proficiency in social interactions and roles (8 items), and (3) acknowledgment of individual psychological needs (3 items). The scale used a five-point Likert format, and expert reviews were conducted with psychologists and lecturers experienced in psychometric development to ensure content accuracy. This was followed by a pre-test, administered both online (Google Forms) and offline (paper-based), to evaluate the instrument's clarity and applicability.

Table 1. Research Sample Demographic Data

Description	n	%
Age		
15	150	14.4
16	383	36.7
17	427	40.9
18	80	7.7
19	5	0.5
Gender		
Men	424	40.5
Women	621	59.5

As previously mentioned, the selected model was the only strictly defined method available for evaluating measurement quality. With strict assumptions, this method enabled examination not only of measurement quality but also served as a first step toward establishing measurement invariance for future studies. Data analysis was conducted in the R Studio with the MIRT Packages application in three stages, namely (1) conducting a test of assumptions, (2) evaluating the items with MNSQ infit-outfit criteria, where 0.6 – 1.4 was acceptable, (3) Item location and difficulty analysis and Wright Map for item mapping, and (4) Finalization for the items.

3. Results

Assumption Checking

A bottom-up method was adopted to test the assumptions, beginning with fundamental function verification for each item using monotonicity checks. The test was first conducted to ensure that each item had functional validity before more complex latent-structure testing. Ignoring monotonicity at the outset would be risky, as it could include items that were functionally "defective" in multidimensional model estimates, thereby obscuring the dimensionality test results. Given the large number of items, this study confirmed the factor/dimension structure. Through this process, the items were clustered into at least two major factors based on the results.

Table 2. Factor Structure Testing

Model	AIC	SABIC	HQ	BIC	logLik	X2	df	p
model_1f	105926.2	106294.8	106311.1	106942.7	-52759.10			
model_2f	102640.0	103098.8	103119.2	103905.6	-51065.98	3386.243	50	0.000

Akaike Information Criterion (AIC) decreased from 105926.2 (one factor) to 102640.2 (two factors). Meanwhile, the Bayesian Information Criterion (BIC) decreased from 106942.7 to 103119.2. Lower AIC and BIC values indicate a better-fitting model, with BIC being particularly stringent because it penalizes model complexity more heavily (Burnham & Anderson, 2004; Schwarz, 1978). The interpretation showed that the substantial decreases in AIC and BIC values, exceeding 3,000 points, provided very strong evidence that adding a second factor significantly improved the model's fit to the data structure. Moreover, a significant Chi-square difference test ($p < 0.05$) suggested that the more complex model provided a significantly better fit to the data than the simpler, nested model (Kline, 2015). A two-factor model with statistically significant latent factors showed better fit than a one-factor model. The increase in log-likelihood further confirmed that the observed data are more likely to occur under the assumption of two latent factors.

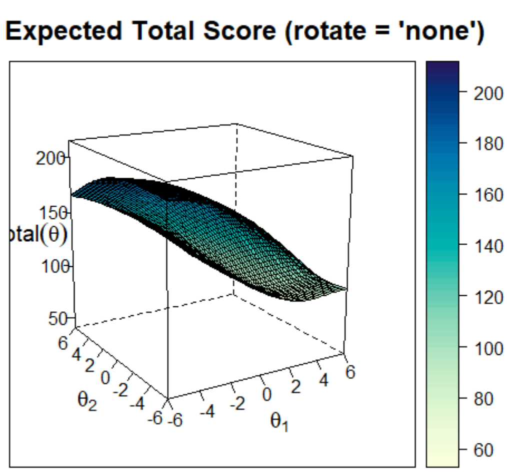


Figure 1 Monotonicity Graph for Expected Total Score

Considering that the instrument structure was found to be multidimensional through exploratory factor analysis (EFA), a monotonicity assumption test was conducted within the MIRT framework. This test was carried out to ensure that the probability of score improvement was a monotonic increasing function of both latent dimensions (Reckase, 2009). Visualization using Item Response Surfaces was necessary to verify that there were no regions in the latent space where an increase in ability decreased the probability of the expected score (Ackerman, 1996; Chalmers, 2012).

In general, the instruments show good monotonicity, and the curve (surface) on the graph continued to rise without any "dips" or flat spots. An individual who had a high score on both theta(θ)-1 and theta(θ)-2 was predicted to obtain a higher total score up to 200. Meanwhile, an individual in the yellow region was predicted to get a lower score of about 60. Although the graph showed good monotonicity, one or more items occasionally appeared problematic, but these weaknesses were compensated for by other items with strong performance. Further examination was necessary to determine whether any item exhibited negative discrimination. This study evaluated the criterion values for each item, selected items with a critical value (crit)> 0.80 and a number of violations (vi), and assessed the significance (zsig). Because these items violated the monotonicity assumption, the selection of higher response categories did not increase significantly with respondent ability. Items clustered in "Red Zone" were KM16 (#crit = 85; #vit = 17; #zsig = 3), KM30 (#crit = 111; #vit = 20; #zsig = 12), KM46 (#crit = 111; #vit = 14; #zsig = 4). The rest with #crit = 40-80 were clustered in "Yellow Zone", and which have a good #crit value = 0 were clustered in "Green Zone". To preserve the latent structure, adjustments were made only to items in the "Red Zone". Items with critical values greater than 0.80 were not retained. However, KM47 and KM48 were maintained despite values reaching the maximum threshold for the reasons explained earlier.

Local Independence

After the fundamental function of each item had been verified, the next analysis tested local independence (LI). Although unidimensionality is often considered the main assumption (Hambleton et al., 1991), in practice, violations of LI (high Q3 values) often indicate additional dimensions or unidentified nuisance factors (Yen, 1984). In MIRT, Q3 Statistics (Yen, 1984) were used for evaluating LI. In this case, the study examined whether residual correlations between items exceeded .20. Correlations above the mean value showed potential problems related to LI. The result of the LI test showed the correlation value, minimum = 0.037, 1st Quartile = 0.430, median = 0.536, mean = 0.529, 3rd quartile = 0.657, and maximum = 0.913. Based on this description, substantial violations of LI were identified, alongside a high degree of item redundancy. Residual correlations (Q3) should ideally remain close to 0, or at least below 0.20. Therefore, several items with high Q3 values were removed, as shown in Table 3.

Table 3. Q3 value

Items	Q3	Description	Items	Q3	Description
KM4 ~ KM5	0.907	Extreme redundancy (>.85)	KM10 ~ KM24	0.846	High redundancy (.80 - .85)
KM2 ~ KM4	0.906		KM11 ~ KM22	0.846	
KM11 ~ KM24	0.913		KM10 ~ KM22	0.932	
KM11 ~ KM12	0.898		KM13 ~ KM14	0.821	
KM14 ~ KM15	0.895		KM13 ~ KM15	0.817	

Items identified as the "core" sources of redundancy were removed to avoid discarding too many items while still achieving optimal results. These items frequently appeared in pairs with high residual correlation values, including KM4, KM11, KM14, KM24, and KM22. After dropping these items, the minimum, 1st Quartile, median, mean, 3rd quartile, and maximum were -0.116, 0.186, 0.361, 0.379, 0.592, and 0.855, respectively. Based on the current results, substantial violations remained, and removing additional items with high Q3 values would have created further problems. Therefore, further investigation was conducted to determine whether systemic redundancy was present. First, the model was evaluated for fit to two factors, yielding RMSEA = 0.07, 95% CI [0.068 - 0.072], CFI = 0.845, TLI = 0.828, and SRMR = 0.063. Based on these results, the two-factor model was a marginal fit or close to a good fit. The high Q3 value, as reflected in the CFI, was below 0.90. Excessive redundancy made it difficult to distinguish variance attributable to latent factors from variance resulting from item duplication.

Several items with high Q3 value, such as KM11, KM12, KM25, and KM49, were dropped. The following result of Q3, minimum = -0.131, 1st Quartile = 0.183, median = 0.351, mean = 0.351, 3rd quartile = 0.525, and maximum = 0.769. At first glance, no significant difference was observed relative to the pre–item-removal results. The two-factor model also showed worse fit, with RMSEA = 0.071 (95% CI [0.069–0.074]), CFI = 0.801, TLI = 0.823, and SRMR = 0.062. The observed data reflected a form of "reliability paradox," in which highly redundant items with Q3 values > 0.80 artificially inflated contributions to model fit, thereby producing misleadingly favorable indications of measurement quality. Item redundancy could artificially inflate reliability and fit indices but undermine construct validity by narrowing the scope of measurement (Boyle, 1991). However, to address this issue, the study did not remove the items; instead, it attempted to resolve the problem using a bifactor model. Reise (2012) argued that the bifactor model provided a comprehensive framework for evaluating whether an instrument was essentially unidimensional despite the presence of local inter-item independence. Furthermore, this model often addressed testlet effects or local dependencies that arise when items share specific content (Cai et al., 2011).

Before performing the bifactor model, the factor structure was re-examined using exploratory factor analysis (EFA). The bifactor model acknowledged the existence of a single general factor underlying all items, while allowing the items to remain grouped within specific factors. After conducting the analysis, the model fit results for EFA were RMSEA = 0.071, 95% CI [0.069 - 0.073], SRMR = 0.062, TLI = 0.855, and CFI = 0.871. The bifactor model yielded RMSEA = 0.055 (95% CI [0.053–0.057]), SRMR = 0.085, TLI = 0.881, and CFI = 0.895. Both the EFA and bifactor models yielded acceptable fit indices. This was particular to the bifactor model, which had better fit indices than the EFA. In other words, confirming the presence of a general factor while allowing two factors to be identified proved to be the appropriate method. This was consistent with the report by Rodriguez et al. (2016) that the bifactor model enabled analysts to retain the idea of a single dominant trait while accounting for multidimensionality arising from item-content clusters. The result was supported by the sharp drop in RMSEA (from 0.07 to 0.055) and the significant increase in CFI. The issue of high Q3 values in the previous model was "resolved" by bifactor because the common factor accounted for the covariance that had previously been considered residual/error. Reise (2012) explained that in psychological instruments with interrelated subscales, the bifactor model provided a better fit because it distinguished between target variance (general) and content variance (specific). Several indices, such as CFI and TLI, were incremental fit indices. When RMSEA indicated a good fit, a CFI near 0.90 was often acceptable for complex psychological constructs with minor multidimensionality. At the very least, an acceptable bifactor model made it possible to "tame" Q3, thereby addressing the assumption of LI without discarding many items..

Table 4. Rasch Analysis Result

Factor 1						Factor 2					
No	**Item**	**Outfit**	**z.outfit**	**infit**	**z.infit**	**No**	**Item**	**Outfit**	**z.outfit**	**infit**	**z.infit**
1	KM1	0.982	-0.34	0.971	-0.574	24	KM2	0.795	-5.327	0.79	-5.457
2	KM6	0.995	-0.086	0.974	-0.491	25	KM3	0.981	-0.461	0.971	-0.696
3	KM7	0.94	-1.048	0.947	-0.91	26	KM5	0.759	-5.728	0.811	-4.595
4	KM9	0.907	-2.158	0.946	-1.175	27	KM8	1.066	1.545	1.059	1.386
5	KM10	0.817	-4.444	0.83	-3.992	28	KM13	0.892	-2.768	0.893	-2.726
6	KM17	0.897	-2.057	0.921	-1.501	29	KM15	0.797	-5.435	0.796	-5.471
7	KM19	0.887	-2.574	0.889	-2.5	30	KM18	1.077	1.692	1.085	1.844
8	KM20	1.004	0.108	0.988	-0.252	31	KM21	1.262	5.362	1.234	4.787
9	KM23	0.887	-2.687	0.892	-2.564	32	KM29	1.039	0.879	1.046	1.053
10	KM26	0.956	-1.004	0.937	-1.467	33	KM31	0.992	-0.171	0.989	-0.252
11	KM27	0.911	-1.992	0.893	-2.417	34	KM32	0.987	-0.316	0.992	-0.194
12	KM28	0.886	-2.495	0.901	-2.076	35	KM33	0.776	-5.874	0.776	-5.868
13	KM38	1.213	4.679	1.167	3.738	36	KM34	0.759	-6.365	0.754	-6.518
14	KM39	1.119	2.394	1.093	1.873	37	KM35	0.867	-3.418	0.861	-3.572
15	KM41	0.948	-0.958	0.953	-0.867	38	KM36	0.722	-7.34	0.727	-7.186
16	KM42	1.066	1.587	1.046	1.109	39	KM37	1.232	5.311	1.207	4.917
17	KM43	1.072	1.789	1.072	1.789	40	KM40	1.12	2.881	1.105	2.536
18	KM44	0.923	-1.795	0.917	-1.946						
19	KM45	0.949	-1.101	0.946	-1.163						
20	KM47	0.969	-0.607	0.987	-0.246						
21	KM48	1.235	4.816	1.193	3.937						
22	KM50	0.969	-0.671	0.952	-1.083						
23	KM51	0.857	-2.773	0.877	-2.39						

Unidimensionality

Based on the results of the previous assumption tests, it was clear that the assumption of unidimensionality was not met due to the presence of two dimensions. Therefore, to meet this assumption, the study required examination of the Explained Common Variance (ECV). Rodriguez et al. (2016) explained that the ECV value > 0.60 and should be classified as Essentials Dimensionality. In this case, the general factors were highly dominant while specific factors contributed minimally. After the calculation, the ECV was only 0.27, indicating that the general factor was very weak. The data is fundamentally multidimensional, not unidimensional. The specific factors (factor 1 and factor 2) explained a greater proportion of the variance than the general factor. Low ECV values indicated that multidimensionality was too strong to ignore, and reporting a single composite score was discouraged on psychometric grounds (Reise, 2012). Forcing a unidimensional structure would constitute a serious violation of Rasch's basic assumptions and result in parameter estimation bias. When specific dimensions accounted for a larger share of variance than the general dimension, the use of composite (total) scores became misleading (Rodriguez et al., 2016; Ten Berge & Sočan, 2004). Therefore, the Rasch analysis was carried out separately for each dimension.

Rasch Analysis

Based on the results of the previous assumption tests, the unidimensionality assumption was not met, as two or more dimensions were identified. Therefore, Rasch Model analysis was carried out separately, with the results presented in Table 4 to meet this assumption.

Mean-square values (MNSQ) between 0.6 and 1.4 were considered acceptable for high-stakes or survey studies, showing that the items were productive for measurement." (Wright & Linacre, 1994; Bond & Fox, 2015). All 23 items in the S1 group had Infit MNSQ and Outfit MNSQ values ranging from 0.817 to 1.235. In general, none of the items fell outside the 0.6-1.4 range. MNSQ values close to 1.0 showed an ideal fit between the observed data and the Bond & Fox (2015) model. Values ranging from 0.8 to 1.2 showed that these items were highly effective in measuring the latent trait on Factor 1. However, items KM48 and KM38 had MNSQ values of 1.16–1.23, which remained below 1.4. This showed that the items exhibited slightly greater variation in unexpected responses while still contributing positively to the measurement structure (Wright & Linacre, 1994). Meanwhile, KM10 had an MNSQ value of 0.81-0.83, suggesting that the item was highly consistent. The response pattern also fit the model well, as values approaching 0.6 without falling below the threshold reflected strong model conformity.

For Factor 2, which consisted of 17 items, the MNSQ values for Infit and Outfit ranged from 0.722 to 1.262. No item exceeded the 1.4 threshold (underfit) or fell below 0.6 (overfit). According to Wright & Linacre (1994), items with an MNSQ below 1.4 were sufficient for clinical instruments or surveys. Generally, values of 1.0 confirmed that each item provided consistent information for measuring factor 2 (Bond & Fox, 2015). KM36 (0.722), KM34 (0.759), and KM5 (0.759) had MNSQ values close to 0.6. These items exhibited a response pattern that was "too neat" or slightly redundant. Although the items fit the model very well, the questions were highly similar, leading respondents to provide highly consistent answers. Linacre (2002) stated that overfit items (MNSQ < 1.0) did not compromise validity but provided less of a "challenge" or variation in information compared to items with values close to 1.0.

The z.outfit and z.infit columns showed Z-scores, with some items having values < -1.96 or> +1.96, indicating statistical significance. However, in large samples, Z-scores often become significant (exceeding the 1.96 threshold) even when the MNSQ was good. Smith et al. (2008) recommended that analysts prioritize the MNSQ value when there was a discrepancy between the MNSQ (which indicates clinical/practical significance) and the Z-score (which indicates statistical significance). This was not only because MNSQ values were more reliable for polytomous data, but also because Z-scores were highly sensitive to sample size and prone to Type I errors.

Table 5. Items Difficulty

Factor 1						Factor 2					
Items	***b***	**Items**	***b***	**Items**	***b***	**Items**	***b***	**Items**	***b***	**Items**	***b***
KM1	2.705	KM23	2.685	KM43	1.505	KM2	2.414	KM29	2.432	KM37	3.021
KM6	1.768	KM26	1.771	KM44	2.351	KM3	2.829	KM31	1.497	KM40	2.168
KM7	2.416	KM27	2.327	KM45	2.434	KM5	2.044	KM32	1.025		
KM9	1.98	KM28	2.255	KM47	2.492	KM8	2.687	KM33	2.194		
KM10	3.125	KM38	1.79	KM48	2.414	KM13	1.268	KM34	2.053		
KM17	1.935	KM39	2.091	KM50	1.928	KM15	2.072	KM35	2.157		
KM19	2.93	KM41	2.097	KM51	2.662	KM18	2.317	KM36	2.527		
KM20	2.263	KM42	2.129			KM21	2.306				

In Table 5, a list of items was compiled along with the respective difficulty levels. Most difficult items included KM10 (3.125), KM37 (3.021), KM19 (2.93), and KM3 (2.829). Moderately difficult items were clustered around 2.2–2.6, including KM36 (2.527), KM47 (2.492), and KM51 (2.662). The easiest items included KM32 (1.025), KM31 (1.497), KM43 (1.505), and KM13 (1.268). This hierarchy showed that KM10 and KM37 were the hardest, while KM32 and KM13 were the easiest. In polytomous models (Rating Scale Model or Partial Credit Model), item difficulty reflected the threshold location of categories, showing that higher b values corresponded to harder categories to endorse consistently across respondents.

The Wright Map visualization (see Figure 2) for both factors showed an off-target discrepancy between the item distribution and respondents' abilities. Most respondents had latent trait levels that fell below the item locations (logit < 1.5). In the context of non-cognitive measurement, this result indicated that the items on Factors 1 and 2 had a very high endorsement threshold. Therefore, respondents tended to rate at the low end of the score range (Bond & Fox, 2015). This phenomenon also revealed a measurement gap at low-to-moderate trait levels, which limited the instrument's sensitivity to distinguishing among respondent characteristics within that range (Sumintono & Widhiarso, 2014).

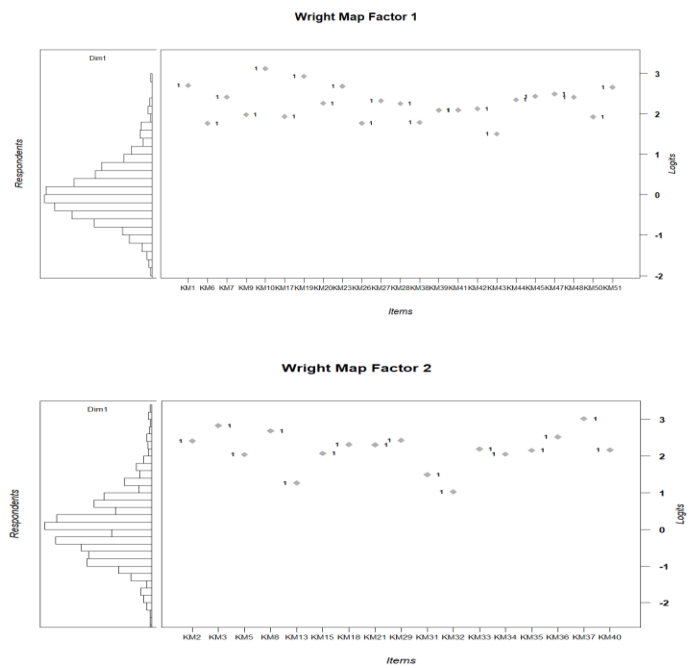


Figure 2 Wright Map Factor 1 and Factor 2

None of the items accurately measured respondents with low ability (negative logit scores) because no items were at that level. Therefore, instruments tend to exhibit a ceiling effect, in which the items are too difficult for respondents. This test was better suited for distinguishing individuals with high trait levels. Factors 1 and 2 were unable to accurately distinguish respondents with low ability, showing that the instrument was too difficult for the student sample.

Person Reliability and Separation Index

The reliability of the instrument was evaluated using the Rasch marginal reliability coefficient. The results of the analysis for Factors 1 and 2 showed a reliability value of 0.936 and 0.887, respectively. According to the criteria of Bond & Fox (2007), reliability values above 0.80 indicate strong internal consistency and the model's ability to effectively distinguish respondents' levels of ability. For the separation index, Factor 1 was 2.81 and Factor 2 was 3.85. This situation was understandable given the presence of off-target effects, as indicated by the Wright Map, which could reduce the reliability of separation. In general, neither Factor 1 nor 2 provided a more detailed breakdown of the respondents.

4. Discussion

This study provides comprehensive psychometric evaluation of a student mental health instrument by integrating multiple advanced analytical frameworks, including factor analysis, bifactor modeling, and

Rasch analysis. The results consistently indicate that the construct being measured is inherently multi-dimensional, as evidenced by the superior fit of the two-factor model relative to the unidimensional alternative. Consistent with the growing body of knowledge, this result suggests that mental health, particularly in student populations, is not a singular construct but rather a constellation of interrelated domains encompassing emotional, cognitive, and behavioral dimensions (Aziz & Mangestuti, 2025; Wei et al., 2016).

The multidimensional structure extended previous results by showing that even when a general factor was statistically identifiable, substantive dominance was not necessarily present (Widaman, 1993). The low ECV value (0.27) observed in this study confirms that the general factor explains only a limited proportion of the common variance, thereby challenging the appropriateness of depending on a total composite score. This result is consistent with contemporary psychometric literature indicating that imposing unidimensionality in the presence of strong multidimensionality can yield biased parameter estimates and misleading substantive conclusions (Bonifay et al., 2015). In this regard, the present study contributes to the literature by empirically reinforcing the importance of dimension-specific interpretation in mental health assessment (Montagni et al., 2026).

The two main dimensions were identified, with Factor 1 reflecting positive self and social functioning, including self-worth, emotional regulation, maintenance of interpersonal relationships, effective performance of social roles, and awareness of personal limitations. Factor two tends to be a negative situation or an inability to perform a behavior that reflects a good mentality. Factor 2 includes several mental health features, with both dimensions reflecting behavioral indicators of mental health. The dimensions appeared to capture whether respondents exhibited these characteristics, suggesting a generally unified structure despite conceptual differences among features. Item correlations within both factors varied across correlation indices (Figures 1 and 2 in Appendix 1), and several possible explanations for this pattern exist. Additionally, the previous Q3 analysis showed several violations, which were not unexpected. Overlapping item content leads to inflated estimates of internal consistency in both the mental health measurement instrument and the educational assessment (Rabin et al., 2021; Wei et al., 2016). However, this is consistent with Iasiello et al. (2024), who reported a substantial overlap among constructs after reviewing 155 measures and 410 dimensions of positive mental health. The implication was that, even though these domains were theoretically distinct, empirical convergence occurred, while correlation patterns varied by population and instrument (Iasiello et al. 2026). Furthermore, empirical studies often show that the domain of mental health loads onto a single higher-order factor, though correlations vary across cultural samples (Keyes, 2005; Keyes et al., 2008). The cultural context shaped how mental health features interrelate. For example, collectivist cultures tend to prioritize social harmony, while individualist cultures place greater focus on autonomy (Fave et al., 2016). This study acknowledges that, as an initial effort to assess the highly complex field of mental health, potential overlap and redundancy make the task herculean.

This study extends earlier work by showing that high reliability and acceptable RMSR and Comparative Fit Index (CFI) values may partly result from redundant item content rather than meaningful representation of the construct (Lai & Green, 2016; Sathayanarayana & Mohanasundaram, 2024). A good overall fit does not necessarily reflect sound measurement, making this an important methodological issue. The application of the bifactor model proved instrumental in resolving this issue. Consistent with a previous study, the bifactor method effectively partitioned shared variance into general and specific components, thereby improving model fit (Bornovalova et al., 2020). Importantly, the bifactor model provided a more nuanced understanding of the data structure by showing that what initially appeared as residual dependence could largely be attributed to a general latent factor (Markon, 2019). However, the weak dominance of this general factor suggests that the instrument should not be treated as essentially unidimensional (Slocum-Gori & Zumbo, 2010). These results refine previous assumptions in the literature by showing that bifactor models, while statistically advantageous, do not automatically justify unidimensional scoring.

The Rasch analysis further confirmed that the instrument showed adequate internal functioning at the item level, with all items falling within acceptable fit ranges. This result shows that the items are productive for measurement within the respective dimensions. However, the Wright Map analysis showed a critical limitation in targeting, with item difficulty levels substantially exceeding respondents' ability levels (Boone et al., 2014). This pattern suggests that the instrument is more sensitive to individuals with higher levels of the latent trait, while lacking precision in differentiating individuals at lower to moderate levels.

Based on a mental health perspective, the results of this study have significant practical implications. Instruments that are poorly targeted may fail to identify students experiencing early or mild psychological difficulties, thereby limiting the utility for preventive screening (Costello, 2016). This concern is particularly salient considering that early detection is a cornerstone of mental health intervention frameworks. The current results suggest that the instrument may be better suited to identifying severe cases than to serving as a broad screening tool across the full spectrum of student mental health.

Despite high reliability coefficients, the relatively modest separation indices show limited discrimination across respondent strata (Wei et al., 2010). This further suggests the need to improve item distribution, particularly by incorporating items sensitive to lower levels of the latent trait. Consistent with previous studies, an optimal instrument should balance measurement precision across the entire continuum of the construct, rather than being skewed toward extreme levels.

Another important implication of this study relates to measurement fairness and invariance. Considering the multidimensional structure and evidence of item redundancy, future studies should incorporate Differential Item Functioning (DIF) analysis to examine whether items function equivalently across subgroups, such as gender, academic level, or socio-cultural background. This is particularly relevant in mental health assessment, where variations in emotional expression and response styles may influence how individuals interpret and respond to items. Integrating DIF within a multidimensional IRT framework would provide a more rigorous test of measurement invariance and ensure that observed differences reflect true differences in the underlying construct rather than systematic bias. Furthermore, due to the study's limited resources, the sample coverage remains concentrated in West Java. It cannot be denied that the landscape of Indonesian adolescents is highly diverse, with varying cultural settings.

Adolescents' mental health is closely related to cultural and social contexts. Therefore, it is important to remember that mental health is closely tied to social context. Expanding the sample's scope in terms of size and cultural diversity is key to understanding and measuring mental health issues among adolescents in Indonesia.

5. Conclusion

In conclusion, this study provides an in-depth psychometric evaluation of the instrument (ISKM-R-2) using Rasch modeling. The main results show that students' mental health is a multidimensional construct, with a two-factor model providing a significantly better fit than a unidimensional model. A low Estimated Common Value (ECV) of 0.27 confirms that the general factor is not sufficiently dominant. Therefore, interpretation of the results must be based on specific dimensions rather than a single composite score to avoid estimation bias. The application of the two-factor model successfully addressed issues of LI violations and item redundancy, as shown by high Q3 residual correlation values. Although all items in the final version showed good fit statistics, Wright Map analysis suggests significant off-target issues. Item difficulty levels were found to far exceed respondents' average ability. This result shows that the instrument has excellent sensitivity for identifying severe mental health conditions (clinical symptoms), but is less sensitive for initial screening in populations with mild to moderate symptoms. Further studies are recommended to recalibrate the instrument by adding items with lower difficulty thresholds, investigating DIF and measurement invariance, and drawing a broader sample or one from a different cultural context.

Ethical Statement: This study upholds research ethics in accordance with the Declaration of Helsinki. The study was conducted with voluntary participation, informed consent, confidentiality, and the protection of respondents' rights and welfare. The ethics also comply with the research standards established by Universitas Pendidikan Indonesia.

Acknowledgment: The authors are grateful to the Bandung city education office and the high schools that participated in this study for helping reach the subjects. This study is also funded by Universitas Pendidikan Indonesia.

Author Contributions: FZ and HI as project leaders, and constructing the research. FZ and MAS wrote the article. MAS and GH conducted data analysis. LS, SC, FZ, HI, GH, AT, AKM, CAN, NB, NK created measurement tool items. AT, AKM, CAN, NB, and NK carried out administration and data collection.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Declaration of Artificial Intelligence (AI) Use: During the preparation of this manuscript, the author used generative artificial intelligence (ChatGPT and OpenL) for language editing and to enhance the clarity and readability of the text. The author critically reviewed, revised, and approved the final manuscript and takes full responsibility for the accuracy, originality, and integrity of the content of the manuscript.

References

Ackerman, T. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, *20*(4), 311–329. https://doi.org/10.1177/014662169602000402

Andriyani, J. (2020). Peran Lingkungan Keluarga dalam Mengatasi Kenakalan Remaja. *At-Taujih : Bimbingan Dan Konseling Islam*, *3*(1), 86–98. https://doi.org/10.22373/taujih.v3i1.7235

Aziz, R., & Mangestuti, R. (2025). Validating psychometric properties of dual-continua models for university student mental health assessment. *International Journal of Public Health Science (IJPHS)*, *14*(1), 425–433. https://doi.org/10.11591/ijphs.v14i1.24601

Beck, M. F., Albano, A. D., & Smith, W. M. (2019). Person-Fit as an Index of Inattentive Responding: A Comparison of Methods Using Polytomous Survey Data. *Applied Psychological Measurement*, *43*(5), 374–387. https://doi.org/10.1177/0146621618798666

Bhugra, D., Till, A., & Sartorius, N. (2013). What is mental health? *International Journal of Social Psychiatry*, *59*(1), 3–4. https://doi.org/10.1177/0020764012463315

Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When Are Multidimensional Data Unidimensional Enough for Structural Equation Modeling? An Evaluation of the DETECT Multidimensionality Index. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(4), 504–516. https://doi.org/10.1080/10705511.2014.938596

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Wright Maps: First Steps. *Rasch Analysis in the Human Sciences*, 111–136. https://doi.org/10.1007/978-94-007-6857-4_6

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate Use of Bifactor Analysis in Psychopathology Research: Appreciating Benefits and Limitations. *Biological Psychiatry*, *88*(1), 18–27. https://doi.org/10.1016/j.biopsych.2020.01.013

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized Full-Information Item Bifactor Analysis. *Psychological Methods*, *16*(3), 221–248. https://doi.org/10.1037/A0023350

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Chotidjah, S., Ihsan, H., Zakariyya, F., Damaianti, L. F., & Mukminin, G. U. (2024). Penyusunan Instrumen Status Kesehatan Mental Pelajar. *Jurnal Ilmu Kesehatan Masyarakat*, *13*(03), 254–262. https://doi.org/10.33221/jikm.v13i03.2278

Costello, E. J. (2016). Early Detection and Prevention of Mental Health Problems: Developmental Epidemiology and Systems of Support. *Journal of Clinical Child & Adolescent Psychology*, *45*(6), 710–717. https://doi.org/10.1080/15374416.2016.1236728

Cronau, H., & Brown, R. T. (1998). Growth And Development: Physical, Mental, And Social Aspects. *Primary Care: Clinics in Office Practice*, *25*(1), 23–45. https://doi.org/10.1016/S0095-4543(05)70324-9

Dubbeldeman, Eveline. M., Boelens, M., Gurp, E. J. B., Dierx, J., Spreeuwenberg, M. D., & Jong, J. C. K. (2025). A Longitudinal Psychometric Evaluation of a Context-Sensitive Positive Health Questionnaire for Measuring Broad Health in

Dutch Adults. *MedRxiv*, 2025.08.23.25334290. https://doi.org/10.1101/2025.08.23.25334290

Fave, A. D., Brdar, I., Wissing, M. P., Araujo, U., Solano, A. C., Freire, T., Hernández-Pozo, M. D. R., Jose, P., Martos, T., Nafstad, H. E., Nakamura, J., Singh, K., & Soosai-Nathan, L. (2016). Lay definitions of happiness across nations: The primacy of inner harmony and relational connectedness. *Frontiers in Psychology*, *7*(JAN), 159794. https://doi.org/10.3389/fpsyg.2016.00030

Gagné, T., Henderson, C., & McMunn, A. (2022). Is the self-reporting of mental health problems sensitive to public stigma towards mental illness? A comparison of time trends across English regions (2009–19). *Social Psychiatry and Psychiatric Epidemiology 2022 58:4*, *58*(4), 671–680. https://doi.org/10.1007/S00127-022-02388-7

Hanlbleton, R. K., Swaminathan, H., Rogers, D. J., & Hambleton, R. K. (1991). *Fundamentals of Item Response Theory Library of Congress Cataloging-in-Publication Data*.

Hidayati, H. N., Hayat, B., & Rahayu, W. (2021). Assessment of the Validity and Reliability of Mental Health Instruments of High School Student in Indonesia. *European Journal of Educational Research*, *10*(2), 729–742. https://doi.org/10.12973/eu-jer.10.2.729

Iasiello, M., Ali, K., Agteren, J. van, Kolovos, E., Kyrios, M., Kashdan, T. B., & Fassnacht, D. B. (2024). What's the difference between measures of wellbeing, quality of life, resilience, and coping? An umbrella review and concept map of 155 measures of positive mental health. *International Journal of Weel Being*, *14*(2), 1–25. https://doi.org/10.5502/ijw.v14i2.3621

Indonesia-National Adolescent Mental Health Survey (I-NAMHS) Report. (n.d.).

Jones, E. A. K., Mitra, A. K., & Bhuiyan, A. R. (2021). Impact of COVID-19 on Mental Health in Adolescents: A Systematic Review. *International Journal of Environmental Research and Public Health 2021, Vol. 18, Page 2470*, *18*(5), 2470. https://doi.org/10.3390/ijerph18052470

Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, *73*(3), 539–548. https://doi.org/10.1037/0022-006x.73.3.539

Keyes, C. L. M., Wissing, M., Potgieter, J. P., Temane, M., Kruger, A., & van Rooy, S. (2008). Evaluation of the mental health continuum-short form (MHC-SF) in Setswana-speaking South Africans. *Clinical Psychology and Psychotherapy*, *15*(3), 181–192. https://doi.org/10.1002/cpp.572

Kline, R. B. (2015). TXTBK Principles and practices of structural equation modelling Ed. 4. In *Methodology in the social sciences*.

Lai, K., & Green, S. B. (2016). The Problem with Having Two Watches: Assessment of Fit When RMSEA and CFI Disagree. *Multivariate Behavioral Research*, *51*(2–3), 220–239. https://doi.org/10.1080/00273171.2015.1134306

Lamblin, M., Murawski, C., Whittle, S., & Fornito, A. (2017). Social connectedness, mental health and the adolescent brain. *Neuroscience & Biobehavioral Reviews*, *80*, 57–68. https://doi.org/10.1016/j.neubiorev.2017.05.010

MacIntyre, P. D., Gregersen, T., & Mercer, S. (2020). Language teachers' coping strategies during the Covid-19 conversion to online teaching: Correlations with stress, wellbeing and negative emotions. *System*, *94*, 102352. https://doi.org/10.1016/j.system.2020.102352

Mansfield, R., Patalay, P., & Humphrey, N. (2020). A systematic literature review of existing conceptualisation and measurement of mental health literacy in adolescent research: current challenges and inconsistencies. *BMC Public Health 2020 20:1*, *20*(1), 607-. https://doi.org/10.1186/S12889-020-08734-1

Markon, K. E. (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annual Review of Clinical Psychology*, *15*(Volume 15, 2019), 51–69. https://doi.org/10.1146/annurev-clinpsy-050718-095522

Meggiolaro, S., & Ongaro, F. (2014). Family contexts and adolescents' emotional status. *Journal of Youth Studies*, *17*(10), 1306–1329. https://doi.org/10.1080/13676261.2014.918246

Montagni, I., Schwartz, A., Grové, C., & Marinucci, A. (2026). Exploration of factors associated with mental health literacy in French students: a dimension-specific approach. *Global Health Promotion*. https://doi.org/10.1177/17579759261419118

Mukminin, G. U., Zakariyya, F., Ihsan, H., Damaianti, L. F., Chotidjah, S., Fitria, A. C., Izzatulhaq, A. A., Naomi, D. P., Hayuningtyas, K., & Putri, W. W. (2024). Development of student mental health scale. *Inovasi Kurikulum*, *21*(1), 443–454. https://doi.org/10.17509/jik.v21i1.64298

Narmandakh, A., Roest, A. M., de Jonge, P., & Oldehinkel, A. J. (2020). Psychosocial and biological risk factors of anxiety disorders in adolescents: a TRAILS report. *European Child & Adolescent Psychiatry 2020 30:12*, *30*(12), 1969–1982. https://doi.org/10.1007/S00787-020-01669-3

Rabin, L. A., Miles, R. T., Kamata, A., Krishnan, A., Elbulok-Charcape, M., Stewart, G., & Compton, M. T. (2021). Development, item analysis, and initial reliability and validity of three forms of a multiple-choice mental health literacy assessment for college students (MHLA-c). *Psychiatry Research*, *300*, 113897. https://doi.org/10.1016/j.psychres.2021.113897

Reckase, M. D. (2009). Multidimensional Item Response Theory. *Multidimensional Item Response Theory*. https://doi.org/10.1007/978-0-387-89976-3

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, *27*(2), 133–144. https://doi.org/10.1111/j.1745-3984.1990.tb00738.x

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rose, T., Joe, S., Williams, A., Harris, R., Betz, G., & Stewart-Brown, S. (2017). Measuring Mental Wellbeing Among Adolescents: A Systematic Review of Instruments. *Journal of Child and Family Studies 2017 26:9*, *26*(9), 2349–2362. https://doi.org/10.1007/S10826-017-0754-0

Sathyanarayana, S., & Mohanasundaram, T. (2024). Fit Indices in Structural Equation Modeling and Confirmatory Factor Analysis: Reporting Guidelines. *Asian Journal of Economics, Business and Accounting*, *24*(7), 561–577. https://doi.org/10.9734/AJEBA/2024/V24I71430

Sari, M. P., & Hazim, H. (2023). The Relationship Between Academic Anxiety and Academic Procrastination in Students of the Faculty of Psychology and Education, University of Muhammadiyah Sidoarjo: Hubungan Antara Kecemasan Akademik dengan Prokrastinasi Akademik pada Mahasiswa Fakultas P…. *Journal of Islamic and Muhammadiyah Studies*, *5*, https://doi.org/10.21070/JIMS.V5I0.1581

Şengül Avşar, A. (2019). Comparison of Person-Fit Statistics for Polytomous Items in Different Test Conditions. *Journal of Measurement and Evaluation in Education and Psychology*, *10*(4), 377–393. https://doi.org/10.21031/EPOD.525647

Şengül Avşar, A. (2023). Aberrant individuals' effects on fit indices both of confirmatory factor analysis and polytomous IRT models. *Current Psychology*, *42*(3), 2157–2166. https://doi.org/10.1007/S12144-021-01563-4

Sequeira, M., Singh, S., Fernandes, L., Gaikwad, L., Gupta, D., Chibanda, D., & Nadkarni, A. (2022). Adolescent Health Series: The status of adolescent mental health research, practice and policy in sub-Saharan Africa: A narrative review. *Tropical Medicine and International Health*, *27*(9), 758–766. https://doi.org/10.1111/TMI.13802

Slocum-Gori, S. L., & Zumbo, B. D. (2010). Assessing the Unidimensionality of Psychological Scales: Using Multiple Criteria from Factor Analysis. *Social Indicators Research 2010 102:3*, *102*(3), 443–461. https://doi.org/10.1007/S11205-010-9682-8

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology 2008 8:1*, *8*(1), 33-. https://doi.org/10.1186/1471-2288-8-33

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial (2nd ed.). Cimahi: Trim Komunikata Publishing House.* (November).

Sunarti, F. S., Almunsiri, O. Y., & Fadilah, L. (2026). The Relationship Between Parenting Style And Peer Social Support With Adolescent Mental Health At Smp Negeri 2 Pakuhaji. *Proceeding of Banten Internasional Conference on Health Advancement* *and Research*, *1*(1), 100–107. https://binar.poltekkesbanten.ac.id/index.php/proceeding/article/view/25

Ten Berge, J. M. F., & Sočan, G. (2004). The Greatest Lower Bound to the Reliability of a Test and the Hypothesis of Unidimensionality. *Psychometrika*, *69*(4), 613–625. https://doi.org/10.1007/BF02289858

World Health Organisation. (2024). Indonesia Adolescent Health Profile 2024. https://www.unicef.org/indonesia/health/reports/adolescent-health-profile

Wei, M., Alvarez, A. N., Ku, T. Y., Russell, D. W., & Bonett, D. G. (2010). Development and validation of a Coping with Discrimination Scale: factor structure, reliability, and validity. *Journal of Counseling Psychology*, *57*(3), 328–344. https://doi.org/10.1037/A0019969

Wei, Y., McGrath, P. J., Hayden, J., & Kutcher, S. (2016). Measurement properties of tools measuring mental health knowledge: a systematic review. *BMC Psychiatry 2016 16:1*, *16*(1), 297-. https://doi.org/10.1186/S12888-016-1012-5

Wright, B. D. (1977). Solving Measurement Problems With The Rasch Model. *Journal of Educational Measurement*, *14*(2), 97–116. https://doi.org/10.1111/j.1745-3984.1977.tb00031.x

Yani, D. I., Chua, J. Y. X., Wong, J. C. M., Pikkarainen, M., Goh, Y. S. S., & Shorey, S. (2025). Perceptions of Mental Health Challenges and Needs of Indonesian Adolescents: A Descriptive Qualitative Study. *International Journal of Mental Health Nursing*, *34*(1), e13505. https://doi.org/10.1111/INM.13505

Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800

How to cite:
Zakariyya, F., Saefudin, M. A., Ihsan, H., Damaianti, L. F., Chotidjah, S., Hidayat, G. U. M., Tazkia, A., Mustafid, A. K., Nurrohman, C. A., Andari, N. B., Pramesti, N. K. A. R. (2026). Rasch Analysis of the Student Mental Health Scale (ISKM-R-2). *PSIKODIMENSIA: Kajian Ilmiah Psikologi*, *25*(1), 30-48. https://doi.org/10.24167/psidim.v25i1.14451

Appendices

Appendix 1. Monotonicity Check Result

Items	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
KM1	0.23	105	1	0.01	0.05	0.05	0.0005	0.95	0	17
KM2	0.3	100	0	0	0	0	0	0	0	0
KM3	0.22	108	4	0.04	0.04	0.14	0.0013	1.49	0	26
KM4	0.32	102	0	0	0	0	0	0	0	0
KM5	0.32	87	0	0	0	0	0	0	0	0
KM6	0.2	97	0	0	0	0	0	0	0	0
KM7	0.21	96	2	0.02	0.06	0.11	0.0012	1.53	0	26
KM8	0.16	108	5	0.05	0.05	0.2	0.0018	1	0	29
KM9	0.17	67	3	0.04	0.04	0.12	0.0018	0.54	0	24
KM10	0.32	79	0	0	0	0	0	0	0	0
KM11	0.32	98	0	0	0	0	0	0	0	0
KM12	0.32	83	0	0	0	0	0	0	0	0
KM13	0.24	102	1	0.01	0.04	0.04	0.0004	0.6	0	13
KM14	0.32	108	1	0.01	0.04	0.04	0.0004	1.34	0	13
KM15	0.3	102	3	0.03	0.07	0.13	0.0013	1.5	0	24

Items	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
KM16	0.12	101	17	0.17	0.08	0.78	0.0078	2.05	3	85
KM17	0.17	91	4	0.04	0.06	0.18	0.002	0.96	0	29
KM18	0.15	101	5	0.05	0.09	0.29	0.0029	1.57	0	39
KM19	0.23	89	2	0.02	0.05	0.08	0.0009	1.2	0	21
KM20	0.15	108	2	0.02	0.03	0.06	0.0006	1.19	0	23
KM21	0.12	108	6	0.06	0.06	0.28	0.0026	1.55	0	38
KM22	0.32	76	0	0	0	0	0	0	0	0
KM23	0.26	87	1	0.01	0.04	0.04	0.0005	1.02	0	15
KM24	0.32	96	0	0	0	0	0	0	0	0
KM25	0.31	97	1	0.01	0.08	0.08	0.0008	1.29	0	19
KM26	0.2	108	5	0.05	0.06	0.25	0.0023	1.54	0	33
KM27	0.22	108	2	0.02	0.04	0.08	0.0008	1.6	0	23
KM28	0.26	91	0	0	0	0	0	0	0	0
KM29	0.21	108	0	0	0	0	0	0	0	0
KM30	0.04	108	20	0.19	0.21	2.49	0.023	3.37	12	158
KM31	0.2	108	1	0.01	0.05	0.05	0.0005	0.81	0	18
KM32	0.19	103	5	0.05	0.06	0.24	0.0023	1.79	1	45
KM33	0.26	108	3	0.03	0.05	0.11	0.001	1.58	0	23
KM34	0.29	108	2	0.02	0.04	0.07	0.0007	1.08	0	16
KM35	0.25	102	3	0.03	0.05	0.14	0.0014	1.88	1	37
KM36	0.3	108	3	0.03	0.05	0.13	0.0012	1.87	1	35
KM37	0.13	102	9	0.09	0.1	0.61	0.006	1.81	1	64
KM38	0.11	108	6	0.06	0.05	0.23	0.0021	1.47	0	37
KM39	0.18	106	1	0.01	0.03	0.03	0.0003	1.35	0	19
KM40	0.18	108	4	0.04	0.05	0.15	0.0013	1.01	0	26
KM41	0.16	108	1	0.01	0.06	0.06	0.0006	1.3	0	24
KM42	0.14	108	5	0.05	0.1	0.38	0.0035	1.71	1	53
KM43	0.16	108	12	0.11	0.08	0.6	0.0056	1.84	1	62
KM44	0.22	108	1	0.01	0.08	0.08	0.0007	1.82	1	36
KM45	0.18	108	5	0.05	0.05	0.21	0.002	0.95	0	28
KM46	0.09	79	14	0.18	0.16	1.14	0.0144	2.67	4	111
KM47	0.15	100	10	0.1	0.13	0.65	0.0065	2.31	1	70
KM48	0.11	106	7	0.07	0.07	0.37	0.0034	1.44	0	42
KM49	0.29	82	0	0	0	0	0	0	0	0
KM50	0.25	108	3	0.03	0.06	0.13	0.0012	1.59	0	25
KM51	0.26	92	1	0.01	0.04	0.04	0.0004	1.14	0	16
KM51	0.26	92	1	0.01	0.04	0.04	0.0004	1.14	0	16

Appendix 2. Monotonicity Check After Drop Items (KM16, KM30, KM46)

Items	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
KM1	0.24	94	8	0.09	0.09	0.36	0.0039	1.9	1	53
KM2	0.32	100	0	0	0	0	0	0	0	0
KM3	0.23	108	5	0.05	0.05	0.19	0.0017	1.45	0	28
KM4	0.34	100	1	0.01	0.03	0.03	0.0003	0.59	0	8
KM5	0.33	82	0	0	0	0	0	0	0	0

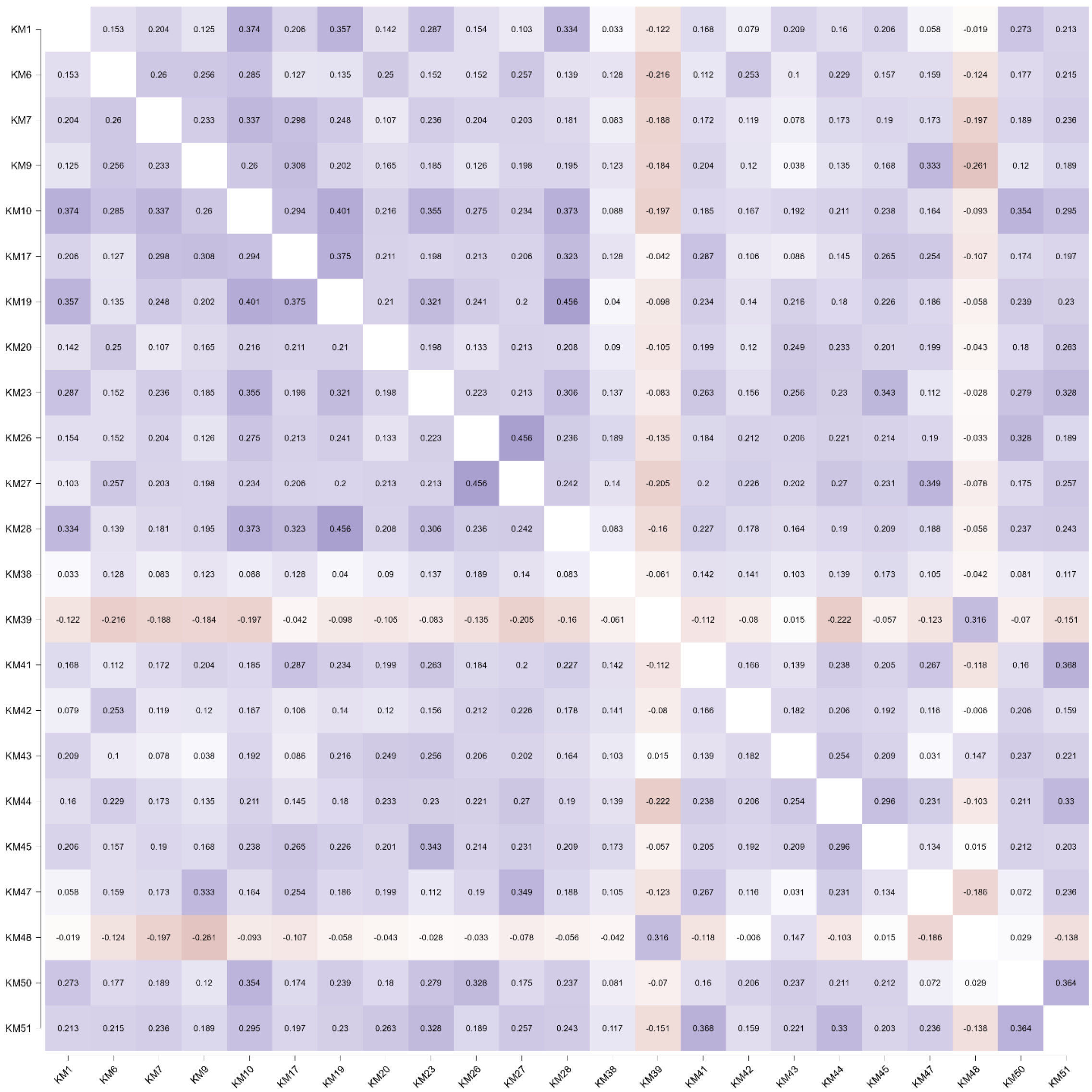
Items	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
KM6	0.2	78	0	0	0	0	0	0	0	0
KM7	0.21	101	7	0.07	0.06	0.31	0.0031	1.33	0	35
KM8	0.17	108	3	0.03	0.05	0.12	0.0011	1.14	0	26
KM9	0.16	78	4	0.05	0.07	0.24	0.0031	1.06	0	34
KM10	0.33	79	0	0	0	0	0	0	0	0
KM11	0.34	96	0	0	0	0	0	0	0	0
KM12	0.33	80	0	0	0	0	0	0	0	0
KM13	0.26	101	1	0.01	0.04	0.04	0.0004	1.41	0	16
KM14	0.33	101	1	0.01	0.03	0.03	0.0003	1.12	0	11
KM15	0.31	101	1	0.01	0.05	0.05	0.0005	0.73	0	12
KM17	0.16	93	2	0.02	0.06	0.09	0.001	0.91	0	24
KM18	0.16	101	8	0.08	0.07	0.42	0.0042	1.68	1	55
KM19	0.23	74	0	0	0	0	0	0	0	0
KM20	0.15	108	1	0.01	0.04	0.04	0.0003	1.09	0	20
KM21	0.13	108	5	0.05	0.09	0.25	0.0023	1.63	0	39
KM22	0.33	78	0	0	0	0	0	0	0	0
KM23	0.26	93	1	0.01	0.04	0.04	0.0005	1.16	0	16
KM24	0.33	92	0	0	0	0	0	0	0	0
KM25	0.31	99	2	0.02	0.07	0.1	0.0011	1.39	0	20
KM26	0.2	108	5	0.05	0.09	0.29	0.0026	2.06	3	58
KM27	0.22	108	4	0.04	0.08	0.19	0.0018	1.63	0	31
KM28	0.26	92	0	0	0	0	0	0	0	0
KM29	0.22	108	6	0.06	0.06	0.26	0.0024	1.27	0	31
KM31	0.21	108	0	0	0	0	0	0	0	0
KM32	0.21	103	3	0.03	0.05	0.14	0.0014	2.06	2	45
KM33	0.28	108	2	0.02	0.07	0.1	0.001	1.57	0	23
KM34	0.31	101	1	0.01	0.03	0.03	0.0003	1.36	0	14
KM35	0.26	103	4	0.04	0.04	0.16	0.0016	1.69	1	37
KM36	0.32	108	2	0.02	0.04	0.07	0.0007	1.04	0	14
KM37	0.14	103	6	0.06	0.08	0.33	0.0032	1.93	1	53
KM38	0.11	108	12	0.11	0.06	0.52	0.0048	1.31	0	48
KM39	0.19	99	4	0.04	0.04	0.14	0.0015	1.48	0	28
KM40	0.2	108	3	0.03	0.05	0.12	0.0011	1.46	0	26
KM41	0.16	103	4	0.04	0.04	0.15	0.0015	1.56	0	30
KM42	0.15	102	4	0.04	0.06	0.18	0.0017	1.53	0	33
KM43	0.17	108	12	0.11	0.08	0.63	0.0058	2.3	2	69
KM44	0.23	108	0	0	0	0	0	0	0	0
KM45	0.18	108	5	0.05	0.08	0.28	0.0026	1.92	2	54
KM47	0.14	94	7	0.07	0.12	0.64	0.0068	2.1	4	79
KM48	0.11	108	10	0.09	0.11	0.59	0.0055	1.96	2	71
KM49	0.3	84	0	0	0	0	0	0	0	0
KM50	0.26	108	2	0.02	0.04	0.07	0.0007	0.9	0	17
KM51	0.26	92	1	0.01	0.04	0.04	0.0005	1.22	0	16

Appendix 3. Final Items

Factor 1		
Items	***b***	**Question**
KM1	2.705455	1. Saya berpikir bahwa saya dapat menerima kelemahan dan kekurangan saya
KM6	1.768456	6. Saya senang berteman dengan teman saya karena dia baik
KM7	2.416184	7. Ketika orang lain memarahi saya, saya yakin bahwa terdapat alasan baik yang mendasari hal tersebut
KM9	1.979591	9. Saya memberikan apresiasi kepada teman saya ketika dia mengalami kesuksesan
KM10	3.124732	10. Saya merasa bersyukur dengan apa yang saya miliki atau jalankan saat ini
KM17	1.934796	17. Saya memandang bahwa setiap masalah yang dialami merupakan bagian dari kehidupan
KM19	2.929976	19. Saya tetap menghargai diri yang telah berusaha sungguh-sungguh walaupun hasilnya kurang 3 harapan
KM20	2.263232	20. Saya merasa bisa mengatakan hal-hal lucu untuk membuat orang di sekitar saya tertawa
KM23	2.685118	23. Saya merasa senang dan bersemangat dalam menghadapi tantangan
KM26	1.770809	26. Saya merasa terhubung secara emosional dengan keluarga saya
KM27	2.326584	27. Saya merasa terhubung secara emosional dengan teman-teman saya
KM28	2.254679	28. Saya mencoba mencintai diri saya dalam segala situasi
KM38	1.789538	38. Saya lebih tertarik untuk menjalin relasi romantis dengan lawan jenis
KM39	2.091344	39. Saya mengabaikan teman-teman saya dengan menolak berkumpul
KM41	2.097045	41. Saya mengambil inisiatif untuk menyelesaikan konflik yang terjadi dalam hubungan saya dengan orang lain
KM42	2.129276	42. Saya sepenuhnya percaya pada orang terdekat saya
KM43	1.505364	43. Saya dapat mengekspresikan emosi saya dengan bebas tanpa takut akan dihakimi oleh orang lain
KM44	2.350632	44. Saya suka terlibat dalam berbagai kegiatan bersama orang lain
KM45	2.43441	45. Saya merasa dapat meningkatkan kemampuan saya ketika melakukan sesuatu di luar zona nyaman
KM47	2.492307	47. Saya memberikan dukungan emosional kepada teman-teman saya yang mengalami kesulitan
KM48	2.414494	48. Saya merasa biasa saja ketika orang di sekitar saya mengalami kesulitan, karena itu tidak terjadi pada saya
KM50	1.927954	50. Saya dapat menyelesaikan permasalahan yang ada di rumah
KM51	2.661834	51. Saya dapat menyelesaikan perselisihan yang terjadi dengan/pada teman di sekolah
Factor 2		
Items	***b***	**Question**
KM2	2.413529	2. Saya berpikir bahwa saya tidak berguna untuk orang lain atau lingkungan saya
KM3	2.829389	3. Saya sulit berpikir rasional ketika menghadapi masalah
KM5	2.043528	5. Saya berpikir bahwa saya tidak bisa melakukan apa-apa dalam kehidupan dan merasa menyesal karena telah lahir
KM8	2.687414	8. Saya berpikir bahwa saran atau kritik dari orang lain adalah bentuk ketidaksukaan mereka pada saya
KM13	1.26754	13. Ketika membandingkan diri saya dengan orang lain, saya merasa cemas atau tidak bahagia
KM15	2.072347	15. Saya merasa tidak memiliki keunggulan
KM18	2.317083	18. Saya merasa masalah yang terjadi pada saya merupakan masalah yang paling berat dan tidak ada orang lain yang mengalaminya
KM21	2.305922	21. Saya cenderung tidak mudah tertawa dengan candaan yang dibuat orang lain
KM29	2.432243	29. Saya merasa tidak nyaman ketika terlibat dalam berbagai kegiatan bersama orang lain
KM31	1.497189	31. Saya merasa mudah cemas dan khawatir ketika menghadapi situasi menantang
KM32	1.025046	32. Saya merasa marah kepada diri sendiri ketika tahu gagal mencapai sesuatu
KM33	2.193598	33. Saya merasa tidak pantas mendapatkan penghargaan dan perhatian dari orang lain
KM34	2.053034	34. Saya merasa tidak bisa lepas dari masalah yang saya alami
KM35	2.157188	35. Saya merasa rendah diri terhadap keadaan diri saya

Factor 1		
Items	***b***	**Question**
KM36	2.526779	36. Saya merasa tidak berdaya dan tidak bisa mengubah situasi yang saya hadapi
KM37	3.020806	37. Saya merasa sulit untuk menentukan apakah orientasi seksual saya berubah atau tetap stabil seiring waktu
KM40	2.167886	40. Saya suka memisahkan diri dari keramaian

Appendix 4. Heatmap Matrix Correlation Between Item in Factor 1



Appendix 5. Heatmap Matrix Correlation Between Item in Factor 2