

Psychometric Evaluation of an Expanded Meaning in Life Questionnaire (MLQ) in the Indonesian Context: A Rasch Model Approach

Faridah Awwaliyyah

Master of Arts in Education, Faculty of Education, International Islamic University of Indonesia, Depok, Indonesia
Email: faridah.awwaliyyah@students.uiii.ac.id

Proceedings of the Kalijaga International Conference on Education 2026

Abstract. Meaning in life is a widely studied construct, yet its measurement in the Indonesian context remains limited, with existing validation work relying only on classical test theory. This study evaluated the psychometric quality of the Meaning in Life Questionnaire (MLQ) in Indonesia using the Rasch model, and expanded the original ten-item instrument with a newly developed item set to broaden its construct coverage. The instrument was adapted through back-translation, reviewed by seven experts using the Many-Facet Rasch Model, and refined across two pilot tests ($n = 36$ and $n = 30$) before a final survey ($n = 108$). Item responses were analyzed with the Rasch Rating Scale Model, separately for the Presence and Search dimensions, examining unidimensionality, item and person fit, rating scale functioning, and separation and reliability, and comparing the original items against the newly developed set. Results showed that the newly developed Presence items achieved unidimensionality and stronger item reliability than the original items, but showed weaker person reliability, while the newly developed Search items improved person reliability but did not fully resolve dimensionality concerns present in the original version. The four-category rating scale functioned largely as intended across both versions. These findings suggest that neither the original nor the newly developed item set alone provides an optimal measure, and that a combined instrument drawing on the stronger-performing items from each version may better balance person and item discrimination. The study offers an initial Rasch-based evaluation of the MLQ for Indonesia and identifies specific items requiring further revision before the instrument can be recommended for wider use.

Keywords: Indonesian context; Meaning in Life Questionnaire (MLQ); meaning in life; psychometric evaluation; Rasch model; scale development

INTRODUCTION

Meaning in life has long been a subject of inquiry across multiple disciplines, including philosophy, theology, education, and psychology, yet no single definition has been universally agreed upon [1]. Scholars such as Schopenhauer, Dilthey, Nietzsche, and Heidegger have approached the construct from distinct philosophical traditions, while later psychologists reframed it as an empirically measurable dimension of human functioning [2]. Within existentialist thought, meaning in life has even been interpreted as an abstract or normative construction rather than an objective given [3]. Despite these conceptual variations, meaning in life is widely regarded as central to human wellbeing; individuals rely on it to interpret their experiences, direct their behavior, and evaluate the significance of their lives [4].

As a psychological construct, meaning in life cannot be observed directly; it is a latent variable inferred through indicators that reflect an individual's underlying cognitive and emotional states [5]. According to MacKenzie and Baumeister [6], meaning serves three interrelated functions: it helps individuals recognize patterns in their environment, facilitates social coordination, and supports self-regulation in decision-making and emotional control. Frankl [7], drawing on his experience as a Holocaust survivor and the founder of logotherapy, proposed that meaning can be attained through three avenues: creative values (producing or contributing something of worth), experiential values (encountering beauty, love, or valuable experience), and attitudinal values (the stance one adopts toward unavoidable suffering). Maslow [8] offered a complementary account, situating the pursuit of meaning at the top of the hierarchy of needs, arguing that self-actualization and by extension, meaning emerges only once deficit needs such as safety, belonging, and esteem have been satisfied.

Building on these foundations, Steger et al. [9] advanced a contemporary, empirically grounded model that distinguishes two independent dimensions of the construct: presence of meaning, referring to the degree to which individuals experience their lives as significant, and search for meaning, referring to the degree to which individuals actively strive to establish or deepen that significance. Steger et al. [9] further

noted that presence and search do not necessarily correlate in a fixed direction, and that their relationship is shaped by demographic and cultural context. To operationalize this two-dimensional model, they developed the Meaning in Life Questionnaire (MLQ), a 10-item self-report instrument that has demonstrated good internal reliability and a stable factor structure across diverse samples. The MLQ represented an improvement over earlier instruments such as the Purpose in Life Test [10] and the Life Regard Index [11] which failed to differentiate between having meaning and seeking it.

The MLQ has since been validated across a range of cultural and clinical contexts. Pezirkianidis et al. [12] confirmed its two-factor structure in a Greek sample through exploratory factor analysis, while Damásio and Koller [13] demonstrated measurement invariance across gender and age groups in a large Brazilian sample of over 3,000 respondents. Naghiyae et al. [14] extended validation to clinical populations, finding that among chronic illness patients the presence and search dimensions could, under certain conditions, coexist rather than diverge. Most recently, Le Vaou et al. [15] confirmed the instrument's validity and reliability among psychiatric crisis patients in France. Within the Indonesian context, however, empirical work on the MLQ remains limited. Rosyad et al. [16] provided one of the few available validation studies, testing the instrument's validity and reliability among a specific population in West Sumatra, yet this study relied solely on classical test theory (CTT) and did not examine deeper psychometric properties such as unidimensionality, item and person fit, differential item functioning, or rating scale category functioning.

This gap is methodologically significant. As Meguellati [17] notes, CTT remains a practical and accessible approach for instrument evaluation, but its estimates are inherently sample- and test-dependent, limiting the generalizability and diagnostic depth of its findings. The Rasch model addresses these limitations by transforming ordinal response data into interval-level measures, enabling item-respondent mapping, evaluation of response category functioning, and estimation of item and person reliability independent of the specific sample distribution [18]. Given that meaning in life is also associated with important educational and psychological outcomes—for instance, students with a stronger sense of meaning in life tend to exhibit greater intrinsic learning motivation [19] establishing a psychometrically robust, culturally appropriate measure of the construct in Indonesia carries both theoretical and practical value.

The present study therefore aims to evaluate the psychometric quality of the Meaning in Life Questionnaire (MLQ) within the Indonesian context using the Rasch model. In addition to testing the original 10-item MLQ, this study expands the instrument with newly developed items to broaden its construct coverage, and examines the resulting instrument's reliability, item-person fit, unidimensionality, differential item functioning, and rating scale functioning across pilot testing and a final survey phase. The findings are intended to inform the development of a more valid and reliable measure of meaning in life for use in future psychological and educational research in Indonesia.

METHODS

Study area

This study was conducted in Indonesia using a quantitative survey design. The research focused on Indonesian adults aged 18–40 years, this age was chosen because Hurlock [20] emphasized that this period is the foundation for every individual, where humans begin to think maturely, in addition, at this age, individuals have a more intense level of digital use, making it easier to collect data. Furthermore, data collection was carried out online in three stages: a first pilot test ($n = 36$), a second pilot test ($n = 30$), and a final survey ($n = 108$). Convenience sampling was employed, as the purpose of the study was to evaluate the psychometric properties of the instrument rather than to generalize the findings to the broader population.

Procedures

Sub-procedure 1: Instrument Adaptation and Item Development

The study began with the adaptation of the Meaning in Life Questionnaire (MLQ) into Indonesian using a back-to-back translation procedure. The original instrument, consisting of 10 items across two dimensions, was translated into Indonesian and subsequently back-translated into English by independent bilingual translators. The original and back-translated versions were compared to ensure semantic equivalence. Following the adaptation process, 10 new items were developed based on the theoretical framework of meaning in life, resulting in a preliminary instrument consisting of 20 items.

Sub-procedure 2: Expert Judgment

The preliminary 20-item instrument was evaluated by seven experts with expertise in psychology, education, and psychometric assessment. Expert judgments were analyzed using the Many-Facet Rasch

Model (MFRM) to examine item quality, expert severity, and agreement among raters. The findings were used to determine the content validity of the instrument and identify items requiring revision.

Sub-procedure 3: Item Revision I

Based on the expert judgment results, several items were revised to improve clarity, wording, and conceptual relevance. Revisions focused on ensuring that each item accurately represented the intended construct while maintaining linguistic appropriateness for Indonesian respondents.

Sub-procedure 4: Pilot Test I

The revised instrument was administered to 36 participants in the first pilot study. Rasch analysis was conducted to examine item fit, person fit, and evidence of unidimensionality. The results were used to identify problematic or overlapping items that required further refinement before the next stage of testing.

Sub-procedure 5: Item Revision II

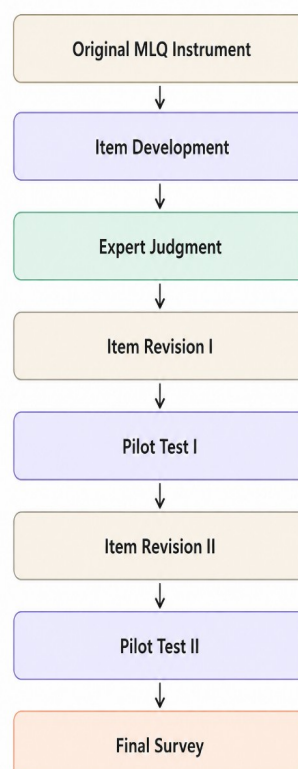
Following the first pilot test, items demonstrating poor psychometric performance or conceptual overlap were revised or removed. The revised instrument was refined to improve measurement quality while preserving the theoretical structure of the construct.

Sub-procedure 6: Pilot Test II

The refined instrument was administered to a second pilot sample of 30 participants. Rasch analysis was performed to confirm item functioning, evaluate the effectiveness of the rating scale, and ensure that the revised instrument met acceptable psychometric standards before large-scale administration.

Sub-procedure 7: Final Survey

The finalized Indonesian version of the instrument was administered to 108 respondents through an online survey. The final data were analyzed using the Rasch measurement model to evaluate the overall psychometric properties of the instrument, including item fit, person reliability, item reliability, separation indices, dimensionality, differential item functioning (DIF), and item comparison.



(Research Procedure Flowchart)

Data analysis

Data were analyzed in two stages using different Rasch-family models, corresponding to the two types of data collected in this study: expert ratings and respondent item responses.

For the expert judgment stage, ratings from the seven reviewers were analyzed using the Many-Facet Rasch Model (MFRM), which accommodates multiple facets items, raters, and rating criteria—within a single measurement framework [21]. This analysis produced a logit-based vertical map of item quality and

rater severity, allowing items and raters that deviated from expected patterns to be identified prior to the pilot testing stage.

For the pilot testing (Phase I and Phase II) and the final survey, item responses were analyzed using the Rasch Rating Scale Model (RSM), an extension of the dichotomous Rasch model for polytomous items [22], using Winsteps/Ministep software. Analyses were conducted separately for the two MLQ dimensions—Presence of Meaning and Search for Meaning—because the instrument was designed to yield two relatively independent subscales [9].

Following the evaluation framework recommended in prior Rasch-based validation studies [23], this study examined several key aspects of measurement quality at each stage: (1) unidimensionality, assessed through the raw variance explained by the Rasch measure and the eigenvalues of unexplained variance in the first contrast of the principal component analysis of residuals; (2) rating scale functioning, evaluated through the ordering of Andrich thresholds and category fit statistics; (3) item and person fit, assessed using Infit and Outfit mean-square (MNSQ) and standardized (ZSTD) statistics, with acceptable ranges of 0.50–1.50 for MNSQ and –2.0 to +2.0 for ZSTD; (4) person–item targeting, examined through the Wright map to determine the correspondence between item difficulty and respondent ability; (5) reliability and separation indices, for both item and person estimates. In the final survey, an additional comparative analysis was conducted between the original MLQ items and the newly developed items to evaluate their relative psychometric contribution to the overall instrument.

RESULTS AND DISCUSSION

Result-1 (Expert Judgment Analysis)

In general, the results of expert judgment indicate that the MLQ has good quality based on the four previous test criteria, namely conformity to the construct, relevance, and representativeness of the indicators. By using MFRM, the results of expert judgment also show differences in quality between items, at least there are 5 items that show the lowest quality, namely items number 20, 16, 2, 9, and 15, however, these findings are not immediately considered to need revision but are used as a basis for evaluating items in the pilot test stage 1, so that all items are still maintained.

Result-2 (Pilot Test I)

The results of the first pilot test showed that the MLQ instrument generally functioned quite well, as indicated by the separation index, Cornbach's alpha value, and overall reliability value at the acceptable level. Furthermore, almost all items had an Infit-Outfit MNSQ value between 0.5 and 1.50, and an Infit-Outfit ZSTD value between -2.0 and 2.0. However, items A5 from and A8 had values above the threshold. Furthermore, in the multidimensionality test, the eigenvalue in the first contrast was 6.47, well above the threshold. The rating scale review revealed an irregular pattern shown by categories 2-3-4, indicating that respondents had difficulty distinguishing between these categories.

Based on these findings, the researcher reviewed all items and found that there was overlap between the original items and the developed items, for example the item "Hidup saya memiliki tujuan yang jelas" Has a similar meaning to the item "Saya memiliki tujuan yang memberi arah dan alasan bagi setiap langkah hidup saya" this is thought to reduce the clarity of the item's function in measuring the construct. So before conducting the second pilot test, a re-instrument was developed involving a waypoints framework, item criteria, and item guidelines to clarify the level of item difficulty, reduce overlapping meanings.

Item Development Using the Waypoints Framework

The overlap identified in Pilot Test I indicated that the item pool, comprising the original MLQ items and the first set of developed items, was conceptually redundant and concentrated within a narrow difficulty range, which explains the multidimensionality signal and the poor category discrimination observed earlier. To address this, a new set of items was developed for each dimension using a waypoints framework, in which items were deliberately constructed to represent five graded levels of construct intensity, very easy, easy, medium, hard, and very hard rather than restating the same idea in different wording. This approach was intended to broaden item-person targeting and allow the instrument to differentiate respondents across a wider range of meanings in life. The resulting items, along with the rationale underlying each difficulty level, are presented for the Presence of Meaning dimension in Table 1 and for the Search of Meaning dimension in Table 2.

Table 1. Waypoints-based item development for the Presence of Meaning dimension

No	Item	Difficulty level
1	Saya merasa hidup saya tidak sepenuhnya kosong.	Very easy (P)
2	Saya merasa apa yang saya jalani sekarang memiliki nilai bagi diri saya.	Easy (P)
3	Saya memahami apa yang membuat hidup saya terasa berarti.	Medium (P)

No	Item	Difficulty level
4	Makna hidup saya tidak mudah goyah oleh masalah yang saya hadapi.	Hard (P)
5	Makna hidup saya menjadi pedoman utama dalam mengambil keputusan penting.	Very hard (P)

Table 2. Waypoints-based item development for the Search of Meaning dimension

No	Item	Difficulty level
1	Saya pernah bertanya-tanya tentang tujuan hidup saya.	Very easy (S)
2	Saya sering merenungkan apa yang membuat hidup saya bermakna.	Easy (S)
3	Saya berusaha memahami tujuan hidup saya melalui pengalaman sehari-hari.	Medium (S)
4	Saya bersedia mengubah cara hidup saya demi menemukan makna yang lebih dalam.	Hard (S)
5	Mencari makna hidup merupakan bagian penting dari identitas diri saya.	Very hard (S)

Result-3 (Pilot Test 2)

The results of the second pilot test showed that the Presence of Meaning dimension had good reliability, separation, and Cronbach's alpha values, which were within the acceptable range. However, item A5, "Hidup saya tidak memiliki tujuan yang jelas," which was an original item from the MLQ, showed outfit and infit MNSQ and ZSTD values above the acceptable criteria. This result indicated that the item required further evaluation. Since item A5 is a negatively worded item and has a similar meaning to item A2, "Hidup saya memiliki tujuan yang jelas," its performance was carefully considered. However, rather than excluding the item at the pilot stage, item A5 was retained for further testing in the larger-scale survey to determine whether its performance would remain inconsistent or improve when administered to a larger sample.

Furthermore, the Search for Meaning dimension also demonstrated good quality, as indicated by its reliability, separation, and Cronbach's alpha values, which were within the acceptable range. Item number 6, "Saya pernah bertanya-tanya tentang tujuan hidup saya," which was developed for the instrument, showed outfit and infit MNSQ and ZSTD values above the acceptable criteria. This finding indicated that the item required further evaluation based on the Rasch model. Therefore, rather than excluding the item solely based on the pilot test results, item 6 was retained for further examination in the final large-scale survey. The larger sample was expected to provide additional evidence regarding the consistency and appropriateness of the item within the measurement model.

At this stage, all dimensions showed eigenvalues below 2, indicating that the instrument demonstrated unidimensionality. In addition, all four rating scale categories functioned appropriately. Therefore, the rating scale structure was retained for the final stage of the survey. Overall, the results of the second pilot test provided sufficient preliminary evidence to proceed to the large-scale survey, while the two items that showed potential misfit were retained for further evaluation using the larger sample.

Results of Final Survey

Dimensionality

According to the Rasch model, a measurement is considered optimal when the construct being measured is unidimensional and the items satisfy the assumption of local independence [24]. In this study, unidimensionality was examined through Principal Component Analysis (PCA) of residuals, where a scale is considered to fundamentally meet the unidimensionality assumption if the variance explained by the Rasch measure exceeds 40%, and the eigenvalue of the first contrast in the unexplained variance is below 2.

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)				
		-- Empirical --		Modeled
Total raw variance in observations	=	8.3	100.0%	100.0%
Raw variance explained by measures	=	3.3	39.9%	41.1%
Raw variance explained by persons	=	1.7	21.0%	21.6%
Raw Variance explained by items	=	1.6	18.9%	19.4%
Raw unexplained variance (total)	=	5.0	60.1%	58.9%
Unexplned variance in 1st contrast	=	2.0	24.1%	40.1%
Unexplned variance in 2nd contrast	=	1.3	15.4%	25.6%
Unexplned variance in 3rd contrast	=	.9	11.1%	18.5%
Unexplned variance in 4th contrast	=	.8	9.5%	15.8%
Unexplned variance in 5th contrast	=	.0	.0%	.0%

(presence original)

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --	Modeled
Total raw variance in observations	=	9.9 100.0%	100.0%
Raw variance explained by measures	=	4.9 49.4%	48.2%
Raw variance explained by persons	=	1.1 11.1%	10.8%
Raw Variance explained by items	=	3.8 38.4%	37.4%
Raw unexplained variance (total)	=	5.0 50.6% 100.0%	51.8%
Unexplnd variance in 1st contrast	=	1.9 18.9%	37.4%
Unexplnd variance in 2nd contrast	=	1.5 15.0%	29.7%
Unexplnd variance in 3rd contrast	=	1.0 10.5%	20.8%
Unexplnd variance in 4th contrast	=	.6 6.1%	12.1%
Unexplnd variance in 5th contrast	=	.0 .0%	.1%

(search original)

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --	Modeled
Total raw variance in observations	=	9.9 100.0%	100.0%
Raw variance explained by measures	=	4.9 49.4%	48.2%
Raw variance explained by persons	=	1.1 11.1%	10.8%
Raw Variance explained by items	=	3.8 38.4%	37.4%
Raw unexplained variance (total)	=	5.0 50.6% 100.0%	51.8%
Unexplnd variance in 1st contrast	=	1.9 18.9%	37.4%
Unexplnd variance in 2nd contrast	=	1.5 15.0%	29.7%
Unexplnd variance in 3rd contrast	=	1.0 10.5%	20.8%
Unexplnd variance in 4th contrast	=	.6 6.1%	12.1%
Unexplnd variance in 5th contrast	=	.0 .0%	.1%

(Presence – Newly Developed)

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --	Modeled
Total raw variance in observations	=	18.4 100.0%	100.0%
Raw variance explained by measures	=	13.4 72.8%	71.9%
Raw variance explained by persons	=	9.1 49.6%	49.0%
Raw Variance explained by items	=	4.3 23.2%	22.9%
Raw unexplained variance (total)	=	5.0 27.2% 100.0%	28.1%
Unexplnd variance in 1st contrast	=	2.3 12.7%	46.6%
Unexplnd variance in 2nd contrast	=	1.5 8.0%	29.5%
Unexplnd variance in 3rd contrast	=	.8 4.3%	15.9%
Unexplnd variance in 4th contrast	=	.4 2.2%	7.9%
Unexplnd variance in 5th contrast	=	.0 .0%	.1%

(Search – Newly Developed)

The PCA results for the original version showed that the Presence subscale explained 39.9% of the variance with a first contrast eigenvalue of 2.0, while the Search subscale explained 56.2% of the variance with a first contrast eigenvalue of 2.5. Both subscales therefore fell short of fully meeting the unidimensionality criteria, with Presence borderline on variance explained and both subscales showing a first contrast at or above the 2.0 threshold.

After item addition, the Presence subscale improved to 49.4% variance explained with a first contrast eigenvalue of 1.9, meeting both criteria for unidimensionality. The Search subscale also improved substantially, with variance explained increasing to 72.8%; however, its first contrast eigenvalue remained above the threshold at 2.3, indicating that a secondary dimension was still present despite the stronger construct measurement. These findings suggest that the newly developed items strengthened the unidimensionality of the Presence subscale, while the Search subscale, although more strongly measured overall, may still capture more than one underlying dimension and warrants further examination of item content in the first contrast.

Item Properties

Based on the discussion in Akhtar & Sumintono [25], the analysis of item characteristics in the Rasch model can be understood to cover several important aspects, such as item fit, local dependence, and the correlation between the item and the construct being measured (point-measure correlation).

Item fit analysis (item fit) aims to identify items that are worth retaining or that need to be eliminated. An item is considered to meet the criteria if its fit statistics fall within the established range. Referring to [26], the criteria used are: (a) an Outfit Mean Square (MNSQ) value between 0.5 and 1.5; (b) an Outfit Z-Standard (ZSTD) value between -2.0 and +2.0; and (c) a point-measure correlation (Pt Mean Corr) value between 0.4 and 0.85.

Original Item Set

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
5	350	108	-.01	.16	2.48	7.5	2.48	7.1	A	.36	.66	40.7	52.8	A5
4	318	108	.71	.15	.77	-1.8	.79	-1.6	B	.74	.72	56.0	50.8	A4
1	361	108	-.29	.16	.68	-2.4	.71	-2.0	C	.74	.64	64.8	55.6	A1
2	351	108	-.04	.16	.63	-2.9	.62	-2.8	b	.77	.66	65.9	52.7	A2
3	364	108	-.37	.16	.49	-4.2	.49	-3.8	a	.77	.63	78.0	56.0	A3
MEAN	348.8	108.0	.00	.16	1.01	-.8	1.02	-.6				61.1	53.6	
S.D.	16.3	.0	.38	.01	.74	4.2	.74	3.9				12.4	1.9	

(Presence – Original)

The item fit results for the original version of the Presence subscale show that item A5 did not meet the fit criteria, with an Outfit MNSQ of 2.48 and an Outfit ZSTD of 7.1, far exceeding the upper limit, together with a point-measure correlation of 0.36, below the minimum threshold of 0.4. The other four items (A1, A2, A3, A4) showed Outfit MNSQ and point-measure correlation values that met the criteria, although the Outfit ZSTD values for items A1 (-2.0), A2 (-2.8), and A3 (-3.8) fell outside the recommended range, indicating an overly predictable (overfitting) response pattern for these three items.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Item	
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
1	375	108	-.46	.20	1.27	1.4	1.37	1.8	A	.68	.75	56.9	62.7	A1
4	349	108	.50	.18	1.09	.5	.93	-.4	B	.82	.82	65.3	59.8	A4
5	345	108	.63	.18	.96	-.2	.84	-.9	C	.84	.82	69.4	56.9	A5
3	358	107	.04	.19	.77	-1.3	.89	-.6	b	.81	.79	69.4	60.9	A3
2	381	108	-.71	.21	.87	-.7	.79	-1.0	a	.75	.73	66.7	63.3	A2
MEAN	361.6	107.8	.00	.19	.99	.0	.96	-.2				65.6	60.7	
S.D.	14.2	.4	.52	.01	.17	1.0	.21	1.0				4.6	2.3	

(Search – Original)

In the original version of the Search subscale, all five items (A1 to A5) met the three item fit criteria, with Outfit MNSQ values ranging from 0.79 to 1.37, Outfit ZSTD values ranging from -1.0 to 1.8, and point-measure correlation values ranging from 0.73 to 0.82. This indicates that all items in the original Search subscale functioned consistently with the Rasch model.

Newly Developed Item Set

Item STATISTICS: MISFIT ORDER

ENTRY	TOTAL	TOTAL		MODEL	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item
4	119	35	1.37	.17	1.69	2.5	1.68	2.4	A-.11	.54	22.9	38.1	A4
5	166	32	-.36	.17	1.23	1.0	1.18	.8	B.75	.57	34.4	35.6	A5
3	189	36	-.44	.16	.84	-.7	.82	-.8	C.55	.56	38.9	35.5	A3
2	187	36	-.39	.16	.60	-2.0	.62	-1.9	b.77	.56	38.9	35.5	A2
1	179	36	-.18	.16	.57	-2.2	.55	-2.2	a.76	.57	36.1	35.0	A1
MEAN	168.0	35.0	.00	.16	.99	-.3	.97	-.3			34.2	35.9	
S.D.	25.8	1.5	.69	.00	.42	1.8	.42	1.7			5.9	1.1	

(Presence – Newly Developed)

For the newly developed item set, the Presence subscale shows that item A4 experienced significant misfit, with an Outfit MNSQ of 1.68 and an Outfit ZSTD of 2.4, exceeding the upper limit, together with a negative point-measure correlation of -0.11, far below the minimum criterion. Item A1 also showed an Outfit ZSTD of -2.2, outside the recommended range, although its Outfit MNSQ and point-measure correlation still met the criteria. Items A2, A3, and A5 met all three fit criteria overall.

ENTRY	TOTAL	TOTAL		MODEL	INFIT	OUTFIT	PT-MEASURE	EXACT	MATCH					
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item	
1	177	34	.54	.21	1.46	1.5	1.66	2.1	A .75	.84	34.5	43.9	A1	
3	204	35	-.65	.23	1.16	.6	.85	-.5	B .81	.78	56.7	50.5	A3	
4	186	34	.14	.22	.83	-.6	.84	-.5	C .83	.81	44.8	42.6	A4	
2	197	35	-.30	.22	.76	-.9	.72	-1.0	b .83	.80	50.0	46.4	A2	
5	183	34	.27	.21	.61	-1.6	.64	-1.4	a .87	.82	37.9	42.3	A5	
MEAN	189.4	34.4	.00	.22	.96	-.2	.94	-.3			44.8	45.1		
S.D.	9.8	.5	.42	.01	.31	1.1	.37	1.2			8.0	3.0		

(Search – Newly Developed)

In the newly developed version of the Search subscale, item A1 showed misfit, with an Outfit MNSQ of 1.66 and an Outfit ZSTD of 2.1, exceeding the upper limit, although its point-measure correlation (0.84) remained within the recommended range. The other four items (A2, A3, A4, A5) met all three item fit criteria, with Outfit MNSQ values ranging from 0.64 to 0.85, Outfit ZSTD values ranging from -1.4 to -0.5, and point-measure correlation values ranging from 0.78 to 0.82.

Rating Scale Function

The rating scale diagnostics were evaluated using the criteria suggested by Linacre [27], namely a minimum of ten observations per category, a monotonic increase in average measures, and an outfit mean square (MNSQ) below 2.0. All four categories showed sufficient observations, ranging from 85 (category 1) to 1,166 (category 4). The average measures and Andrich thresholds increased monotonically across all categories (-2.17, -0.69, 0.58, 2.33), with no threshold disordering, indicating that respondents used the four categories in a manner consistent with their intended order. The outfit MNSQ values were within the acceptable range for categories 2 through 4 (.81, .78, and 1.01, respectively), while category 1 showed a slightly elevated outfit MNSQ of 2.12, along with the lowest proportion of observations (4%). Overall, the four-category rating scale functioned largely as intended, with ordered category measures and thresholds, though category 1 showed marginally more unpredictability in responses than the other categories.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	ANDRICH	CATEGORY	
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	THRESHOLD
1	1	85	4	.45	.07	1.31	2.12	NONE
2	2	237	11	.37*	.53	.82	.81	-.73
3	3	731	33	1.04	1.07	.85	.78	-.34
4	4	1166	53	1.88	1.85	1.03	1.01	1.07
MISSING		1	0	1.92				

Person and Item Separations and Reliability

Reliability analysis is used to describe how consistent the measurement results would be if the instrument were used more than once [28], while item separation shows whether the response data are good enough to determine the order of item difficulty consistently [29]; both refer to the ability of the test to define a distinction hierarchy of items along the measured variable [30].

Original MLQ

	Person	Item
N	108	5
SD	1.19	0.00
Separation	0.92	1.89
Reliability	0.46	0.78
Cronbach's Alpha	0.72	

(Presence Original)

	Person	Item
N	108	5
SD	1.16	0.00
Separation	1.49	2.40
Reliability	0.69	0.85

	Person	Item
Cronbach's Alpha	0.87	

(Search Original)

Newly Developed MLQ

	Person	Item
N	108	5
SD	0.78	0.00
Separation	0.77	3.71
Reliability	0.37	0.93
Cronbach's Alpha	0.65	

(Presence – Newly Developed)

	Person	Item
N	108	5
SD	1.27	0.00
Separation	2.22	1.56
Reliability	0.83	0.71
Cronbach's Alpha	0.98	

(Search – Newly Developed)

Reliability analysis was used to describe how consistent the measurement results would be if the test were used more than once [28], while item separation shows whether the response data is good enough to consistently order the difficulty of the items [29]; both refer to the ability of the test to define a distinct hierarchy of items along the measured variable [30]. Person separation and reliability describe how well the instrument distinguishes respondents with different ability levels, while item separation and reliability describe how well the difficulty hierarchy among items is defined, with values above 2.0 for separation and above 0.8 for reliability generally considered good.

For the original version, the Presence subscale showed a person separation of 0.92 (reliability = 0.46) and an item separation of 1.89 (reliability = 0.78, Cronbach's alpha = 0.72), indicating that while the items were reasonably well separated by difficulty, the instrument was not yet strong enough to distinguish respondents with different levels of presence of meaning. The Search subscale performed somewhat better, with a person separation of 1.49 (reliability = 0.69) and an item separation of 2.40 (reliability = 0.85, Cronbach's alpha = 0.87), though the person side still fell below the ideal threshold.

For the newly developed item set, the two subscales showed different patterns. The Presence subscale's item separation increased sharply to 3.71 (reliability = 0.93), but its person separation dropped to 0.77 (reliability = 0.37, Cronbach's alpha = 0.65), meaning the newly developed items sharpened the difficulty hierarchy among items but weakened the instrument's ability to differentiate respondents, likely related to the reduced spread of person measures (SD = 0.78, compared to 1.19 in the original version). The Search subscale showed the opposite pattern: person separation improved to 2.22 (reliability = 0.83), now meeting the recommended threshold, while item separation decreased to 1.56 (reliability = 0.71, Cronbach's alpha = 0.98).

CONCLUSION

This study evaluated the psychometric properties of the Meaning in Life Questionnaire (MLQ) in the Indonesian context using the Rasch model and examined whether newly developed items could improve the measurement of the two dimensions of meaning in life: Presence of Meaning and Search for Meaning. The findings demonstrate that the psychometric performance of the original and newly developed item sets differed across dimensions.

The newly developed Presence of Meaning items showed improved unidimensionality and substantially stronger item separation and reliability compared with the original item set. However, their low person separation and reliability indicated limited ability to distinguish respondents with different levels of meaning in life. In contrast, the newly developed Search for Meaning items substantially improved person separation and reliability, although the dimensionality analysis indicated that the subscale still did not fully satisfy the assumption of unidimensionality. Several problematic items were also identified across both versions, including the negatively worded original Presence item and specific newly developed items that showed misfit.

The rating scale analyses indicated that the response categories generally functioned according to their intended ordering. However, the interpretation of rating scale functioning should distinguish between the original MLQ, which uses a seven-category response format, and the newly developed item set, which uses a four-category response format. Therefore, the two versions should not be interpreted as having identical category structures.

Overall, the findings indicate that neither the original MLQ nor the newly developed item set independently provides an optimal measurement instrument for the Indonesian context. Rather, the results suggest that combining the strongest-performing items from both versions may provide a more balanced instrument with improved construct coverage, item quality, and ability to differentiate respondents.

This study provides an initial Rasch-based psychometric evaluation of the MLQ in Indonesia and contributes evidence beyond previous validation studies relying primarily on Classical Test Theory. Further research should revise or remove the identified misfitting items, examine the source of the secondary dimension within the Search for Meaning subscale, and test a refined combined item set using a larger and more diverse Indonesian sample. Future studies should also carefully evaluate the most appropriate response scale for the refined instrument before it can be recommended for wider use in psychological and educational research in Indonesia.

The interpretation of meaning-in-life measurement and Rasch diagnostics was also situated within the broader psychometric literature [31], [32], [33].

REFERENCES

- [1] Fridayanti, F. (2013). PEMAKNAAN HIDUP (MEANING IN LIFE) DALAM KAJIAN PSIKOLOGI. *Psikologika: Jurnal Pemikiran Dan Penelitian Psikologi*, 18(2). <https://doi.org/10.20885/psikologika.vol18.iss2.art8>
- [2] Auhagen, A. E. (2000). On the psychology of meaning of life. *Swiss Journal of Psychology*, 59(1), 34–48. <https://doi.org/10.1024//1421-0185.59.1.34>
- [3] Camus, A. (1991). *The Myth of Sisyphus and Other Essays* (J. O'Brien, Trans.). Vintage International.
- [4] Steger, M. F., Oishi, S., & Kashdan, T. B. (2009). Meaning in life across the life span: Levels and correlates of meaning in life from emerging adulthood to older adulthood. *The Journal of Positive Psychology*, 4(1), 43–52. <https://doi.org/10.1080/17439760802303127>
- [5] Spirtes, P. (2001). Latent Structure and Causal Variables. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 8395–8400). Elsevier. <https://doi.org/10.1016/B0-08-043076-7/00429-0>
- [6] MacKenzie, M. J., & Baumeister, R. F. (2014). Meaning in Life: Nature, Needs, and Myths. In A. Batthyany & P. Russo-Netzer (Eds.), *Meaning in Positive and Existential Psychology* (pp. 25–37). Springer. https://doi.org/10.1007/978-1-4939-0308-5_2
- [7] Frankl, V. E. (2006). *Man's Search for Meaning*. Beacon Press.
- [8] Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396. <https://doi.org/10.1037/h0054346>
- [9] Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, 53(1), 80–93. <https://doi.org/10.1037/0022-0167.53.1.80>
- [10] Crumbaugh, J. C., & Maholick, L. T. (1964). An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis. *Journal of Clinical Psychology*, 20(2), 200–207. [https://doi.org/10.1002/1097-4679\(196404\)20:2<200::AID-JCLP2270200203>3.0.CO;2-U](https://doi.org/10.1002/1097-4679(196404)20:2<200::AID-JCLP2270200203>3.0.CO;2-U)
- [11] Battista, J., & Almond, R. (1973). The development of meaning in life. *Psychiatry*, 36(4), 409–427. <https://doi.org/10.1080/00332747.1973.11023774>
- [12] Pezirkianidis, C., Galanakis, M., Karakasidou, I., & Stalikas, A. (2016). Validation of the Meaning in Life Questionnaire (MLQ) in a Greek Sample. *Psychology*, 07(13), 1518–1530. <https://doi.org/10.4236/psych.2016.713148>
- [13] Damásio, B. F., & Koller, S. H. (2015). Meaning in Life Questionnaire: Adaptation process and psychometric properties of the Brazilian version. *Revista Latinoamericana de Psicología*, 47(3), 185–195. <https://doi.org/10.1016/j.rlp.2015.06.004>
- [14] Naghiyae, M., Bahmani, B., & Asgari, A. (2020). The Psychometric Properties of the Meaning in Life Questionnaire (MLQ) in Patients with Life-Threatening Illnesses. *The Scientific World Journal*, 2020, 8361602. <https://doi.org/10.1155/2020/8361602>
- [15] Le Vaou, P., et al. (2025). Validation of the French version of the Meaning of Life Questionnaire (MLQ) in patients admitted for crisis hospitalization in a psychiatric emergency service (crisis center). *L'Encephale*. Advance online publication. <https://doi.org/10.1016/j.encep.2025.09.001>
- [16] Rosyad, Y. S., Malini, H., & Sarfika, R. (2019). Validity and reliability the Meaning in Life Questionnaire (MLQ) ; Men who have sex with men at West Sumatera Indonesia. *Riset Informasi Kesehatan*, 8(1), 16. <https://doi.org/10.30644/rik.v8i1.218>
- [17] Meguellati, S. (2024). A Critical Analysis of the Use of Classical Test Theory (CTT) in Psychological Testing: A Comparison with Item Response Theory (IRT). *Pakistan Journal of Life and Social Sciences (PJLSS)*, 22(2). <https://doi.org/10.57239/PJLSS-2024-22.2.00715>
- [18] Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd edition). Routledge, Taylor & Francis Group.
- [19] Liu, S., Gong, X., Ji, Q., Li, Q., Kong, M., Du, X., Xiang, L., & Zhu, Z. (2025). Meaning in life, positive cognition, and learning motivation: A mediational analysis among Chinese college students. *PLOS One*, 20(9), e0330447. <https://doi.org/10.1371/journal.pone.0330447>
- [20] Hurlock, E. B. (1981). *Developmental psychology: A life-span approach* (5th ed.). Tata McGraw-Hill.

- [21] Linacre, J. M., Heinemann, A. W., Wright, B. D., Granger, C. V., & Hamilton, B. B. (1994). The structure and stability of the functional independence measure. *Archives of Physical Medicine and Rehabilitation*, 75(2), 127–132. [https://doi.org/10.1016/0003-9993\(94\)90384-0](https://doi.org/10.1016/0003-9993(94)90384-0)
- [22] Andrich, D. (1978). Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers. *Applied Psychological Measurement*, 2(4), 581–594. <https://doi.org/10.1177/014662167800200413>
- [23] Mui Lim, S., Rodger, S., & Brown, T. (2009). Using Rasch analysis to establish the construct validity of rehabilitation assessment tools. *International Journal of Therapy and Rehabilitation*, 16(5), 251–260. <https://doi.org/10.12968/ijtr.2009.16.5.42102>
- [24] Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, 46(3), 380–393. <https://doi.org/10.1016/j.ijnurstu.2008.10.007>
- [25] Akhtar, H., & Sumintono, B. (2023). Rasch analysis of the International Personality Item Pool Big Five Markers Questionnaire: Is longer better? *Primenjena Psikologija*, 16(1), 3–28. <https://doi.org/10.19090/pp.v16i1.2401>
- [26] Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch pada Assessment Pendidikan. *Trim Komunikata*.
- [27] Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- [28] Sugiono, Noerdjanah, & Arifianti. (2020). Uji Validitas dan Reliabilitas Alat Ukur SG Posture Evaluation. *Jurnal Keterapian Fisik*, 5(1), 55–61. <https://doi.org/10.37341/jkf.v5i1.167>
- [29] Pada, A. U. T., Kartowagiran, B., & Subali, B. (2016). Separation index and fit items of creative thinking skills assessment. *REID (Research and Evaluation in Education)*, 2(1), 1–12. <https://doi.org/10.21831/reid.v2i1.8260>
- [30] Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed). Lawrence Erlbaum Associates Publishers.
- [31] Nugent, W. R. (2017). Understanding DIF and DTF: Description, Methods, and Implications for Social Work Research. *Journal of the Society for Social Work and Research*, 8(2), 305–334. <https://doi.org/10.1086/691525>
- [32] Schutte, L., Wissing, M. P., Ellis, S. M., Jose, P. E., & Vella-Brodrick, D. A. (2016). Rasch analysis of the Meaning in Life Questionnaire among adults from South Africa, Australia, and New Zealand. *Health and Quality of Life Outcomes*, 14(1), 12. <https://doi.org/10.1186/s12955-016-0414-x>
- [33] Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358–1362. <https://doi.org/10.1002/art.23108>