PSYMPATHIC : Jurnal Ilmiah Psikologi
Vol. 9, No. 1, 2022: 31-44

eISSN: 2502-2903, pISSN: 2356-3591
DOI: 10.15575/psy.v9i1.14270

Rasch Rating Scale Model: Bias Detection and Validation Test of Indonesian-Adolescent Life Satisfaction Scale

Yonathan Natanael[1*], Rahmaditya Salsabilla[1], Denisa Aulia[1], Dwi Khoirunnisa[1], Husna Nursaid Munawar[1], Nada Salsabila Hidayat[1], Romi Faizal Firdaus[1]
[1]Faculty of Psychology, Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
*e-mail: yonathan@uinsgd.ac.id

Abstract

Life satisfaction instrument among psychology related analyses has weaknesses in terms of gender bias testing, especially among adolescents. Therefore, this research aims to develop and validate a life satisfaction instrument in the Indonesian language which is new and undetected by gender bias. The modern construct-oriented scale construction method was selected as a guide in developing the instrument, and the participants were 474 adolescents from various cities in Indonesia. At the initial screening stage using the Rasch model analysis, 54 participants were detected as outliers, hence further analysis was only conducted on 420. The results show that I-ALSS fulfills the assumptions of unidimensional and local independence, with a good reliability score. Furthermore, I-ALSS correctly categorizes answer choices into 8 items, without detecting gender bias. Therefore, it is a fairly good instrument to be used in Indonesia.

Keywords: life satisfaction, modern-construct oriented scale construction, rating scale model

Abstrak

Instrumen kepuasan hidup yang sangat populer di kalangan peneliti Psikologi memiliki kelemahan dalam hal pengujian bias gender, khususnya pada kalangan remaja di berbagai negara. Oleh karena itu, penelitian ini bertujuan mengembangkan dan memvalidasi instrumen kepuasan hidup dalam bahasa Indonesia yang baru dan tidak terdeteksi mengalami bias gender. Metode *modern construct-oriented scale construction* dipilih sebagai panduan dalam mengembangkan instrumen. Partisipan sebanyak 474 remaja berasal dari berbagai kota di Indonesia. Pada tahap *screening* awal menggunakan analisis *Rasch model* terdapat 54 partisipan terdeteksi sebagai *outliers* penelitian, sehingga dalam penganalisisan lebih lanjut hanya dilakukan pada 420 partisipan. Analisis *Rating Scale Model (RSM)* menunjukkan I-ALSS memenuhi asumsi unidimensional dan lokal independensi, memiliki nilai reliabilitas yang baik, tepat dalam pengkategorisasian pilihan jawaban, serta terdapat 8 item yang baik dan tidak terdeteksi bias gender. I-ALSS merupakan instrumen yang cukup baik untuk digunakan di Indonesia.

Kata Kunci: kepuasan hidup, *modern-construct oriented scale construction*, *rating scale model*

Introduction

Life satisfaction is very important for individuals as a key part of subjective and relative well-being. There are different perceptions and understanding of this concept in relation to emotion, feelings and mood (Fernández-Portero et al., 2017). Rissanen (2015) explained that life satisfaction benefits an individual and the surrounding environment.

Diener and Diener (1995) stated that life satisfaction is one of the important components of producing a good experience for individuals. The concept is often used as a reference to evaluate the quality of life. Satisfied individuals tend to be adaptive even when facing difficult or stressful situations. Furthermore, they are also exploratory in carrying out various activities because life satisfaction consists of the assessment of life, confidence in getting a better life, and feeling satisfied with essential achievements in life (Diener et al., 1985; van Beuningen, 2012).

During the Covid-19 pandemic, many adolescents in Indonesia and around the world felt the lockdown effects or did not conduct their social life activities (van der Laan et al., 2021). Concerning the effect, adolescents

showed a very low value for life satisfaction in the first and second years of the pandemic (Borualogo & Casas, 2022). This is closely related to subjective well-being or adolescent happiness in Indonesia. In addition, there was a high tendency to experience anxiety and symptoms of depression (van der Laan et al., 2021). Therefore, when life satisfaction is low with increasing anxiety and depression, adolescents do not experience subjective well-being or are less content with their lives.

A well-known instrument for measuring individual life satisfaction is the Satisfaction with Life Scale (SWLS). Natanael and Novanto (2020) confirmed that SWLS is the most popular instrument used by many studies. In the field of Psychology (Google Scholar, 2022), this instrument has been cited in over thirty-five thousand studies worldwide for correlational, causality, experiments, test, instrument testing, mixed-mechanism, and mixed-method research. These studies involve various characteristics of participants because SWLS shows good psychometric properties (Natanael & Novanto, 2020). It is a simple instrument that briefly assesses individual life satisfaction in general.

The advantages of SWLS summarized from previous research consist of only five items. SWLS reliability value > .80, indicates a reliable or consistent instrument (Sufren & Natanael, 2014). Based on several analytical tests through several statistical programs, the items are valid and measure only one dimension in exploratory and confirmatory factor analysis tests. Furthermore, the best SWLS measurement model also reaches a tau-equivalent model, meaning the instrument can be calculated directly. This is because the loading factor value is similar for each item and the test model fits when the parameter constraints are applied (Natanael & Novanto, 2020).

The use of SWLS, with a focus on testing item bias or measurement invariance, found conflicting results regarding gender bias. There was no gender bias on SWLS items using the measurement invariance technique on Brazilian students (Zanon et al., 2014). In Angola, research on item bias by testing configurational, metric, and scalar invariance in SWLS showed the absence of item bias (Tomás et al., 2015), unlike research in Norway (Arikan & Zorbaz, 2020; Moksnes et al., 2014). Furthermore, SWLS also experiences cultural bias, as reported by research comparing the life satisfaction of Brazilian and United States students (Schnettler et al., 2017). Some contradicting research above uses the same testing technique and criteria for participants, namely students included in the adolescent development stage. The search for conflicting results was not explored due to the absence of a meta-analysis discussing SWLS from the initial development in 1985 to the present in 2022.

In addition to having advantages from the psychometric side, SWLS also has shortcomings in item bias testing. The test on the attitude scale used by social research is still very rarely analyzed or proven with a quantitative approach. Arifin (2017) stated that in determining the feasibility of an instrument, evidence is needed in terms of validity, reliability, distinguishing power, and distractors. However, the expert in measurement showed that the principle of fairness is needed in the instrument for measuring psychological attributes. This principle of justice is known as measurement equivalence, which is also very necessary to be the main requirement for instrument testing (Raju et al., 2002).

The principle of fairness in Education and Psychology can be analogized as follows: when an instrument is tested or compared to different groups based on the ability and probability to answer the items, no difference in probability should be reported. The principle of justice in measurement does not occur when there are different opportunities to provide answers in one group. Meanwhile, differences in responses suggest injustice and bias in the measuring instrument.

Testing for item bias is conducted on instruments that measure individual cognitive or achievement. The first example can be seen

from research concerning the principle of fairness on the New Student Admissions Selection (NSAS) test. In the study of doctoral students at Gajah Mada University, there was a gender bias in the New Student Admission Selection items that measured analogy, analytic, arithmetic, and geometric components (Ridho, 2014). Moreover, when given questions that measure these components, the male group benefits more than the female group. Another example in the form of a cognitive test indicating gender bias is the Multidimensional Aptitude Battery (MAB). Tresnawaty (2013) detected item bias on the MAB instrument and found nine items. In both examples of these instruments, an item bias test has been carried out based on gender.

Given the conflicting research on gender bias in SWLS and the limited test of attitude scales, an instrument should be developed to measure life satisfaction free from bias and be used in Indonesia. The life satisfaction instrument was developed based on the limited permission and opportunity to revise or confirm the SWLS bias because the person who developed SWLS had died in 2021. Therefore, developing an instrument based on the life satisfaction theory proposed by Diener is possible.

In another view, this research assesses the importance of having new information to analyze instruments with the principle of fairness or detecting item bias. SWLS prioritizes the principle of justice by using measurement invariance analysis with Multiple Group Confirmatory Factor Analysis (MGCFA). Previous research conducted a confirmatory factor analysis technique by comparing the loading factor of the model test on the tested group based on gender or cultural groups. Therefore, this research uses Differential Item Functioning (DIF) testing to produce a different result. The DIF test on Rasch modelling directly shows the difference in scores between the groups tested by displaying the values to test for bias and the instrument validation in more detail.

Methods

Research Design

The design used is a modern construct-oriented scale construction (John & Bent-Martinez, 2013) in developing the questionnaire. It has advantages in testing the instrument and integrating construct validity analysis. For example, external validity can be tested by convergent and discriminant validity. Furthermore, it can be conducted with other types of validity testing including face, content, criterion-oriented, construct, substantive, structural, generalizability, consequential, and external. This design has six stages, namely: (a) making research hypotheses, (b) creating measurement and alternative models, (c) creating items from constructing definitions to be examined and using content validity as a guide, (d) data analysis, (e) confirming the appropriate final test model, and (f) generalizing the test to produce a "good enough" model (John & Bent-Martinez, 2013). The explanation of each stage is explained in the research procedure.

Participants and Ethics Test

The initial participants were 474 students in the adolescent development stage between the ages of 17-19. This is under the adolescent age range suggested, and the age of 12-19 is a development stage (Karnik & Kanekar, 2012). All participants are willing to fill out the google form link, which consists of a column of willingness or informed consent as a participant, demographic data, and items used for research. Participants were taken using a non-probability technique called quota sampling, which has the same characteristics. Quota sampling is used because the number of participants is expected to be fulfilled in online research (Etikan & Bala, 2017). The test requirements for developing a measuring instrument for the number of participants are at least ten times the items (10 x n-item).

Demographic data in Table 1 shows that participants consist of two levels of education, namely Senior/Vocational High School

students and in the Diploma 3, Diploma 4, or Bachelor levels. They were also from various cities in Indonesia, dominated by the cities of Bandung, Sukabumi, and Jakarta. The origin of the area/city was obtained from data dissemination using online questionnaires distributed through social media, especially WhatsApp and Instagram groups. This research does not provide incentives for all but is only given to 5 selected participants in the form of Gopay balances worth IDR 25,000 each.

This research has also passed the ethics test with evidence of Ethical Clearance No: 020.2022 Ethics/KPIN conducted by an institution in Indonesia that specifically examines the ethics of psychological analyses in Indonesia. It follows the results of a review provided by an ethical testing institute related to testing life satisfaction instruments to validate the instrument. Therefore, it uses two instruments, further explained in another section.

Table 1
Participant Demographic Data

Category	Total	%
Gender		
Male	373	78.7%
Female	101	21.3%
Age		
17 years old	41	8.6%
18 years old	33	7%
19 years old	400	84.4%
Education Level		
Senior/Vocational High School	72	15.2%
Diploma 3	20	4.2%
Diploma 4/ Bachelor	382	80.6%
Hometown		
Bandung	87	18.4%
Sukabumi	78	16.5%
Jakarta	52	11%
Tasikmalaya	46	9.7%
Bogor	22	4.6%
Bekasi	16	3.4%
Cianjur	16	3.4%
Other cities in Indonesia	157	33.1%

Research Instrument

The first instrument is the Indonesian-Adolescent Life Satisfaction Scale (I-ALSS). The I-ALSS consists of thirteen items based on five indicators of life satisfaction proposed by Diener (Diener et al., 1985). Diener proposed life satisfaction does not have aspects/dimensions/components but only consists of indicators. Therefore, the life satisfaction measurement model is unidimensional based on the indicators. The indicators measure life satisfaction, not dimensions or related variables. I-ALSS uses five answer choices from 'strongly disagree' to 'strongly agree'. The thirteen items developed are presented in table 2.

The second instrument is the Satisfaction with Life Scale (SWLS), translated into Indonesian (Novanto & Pali, 2019). The Indonesian version of SWLS is used with the developed measuring instrument as a comparative measurement tool. The five items can be seen in the research article entitled "Test of congeneric, tau-equivalent and parallel measurement models on Satisfaction with Life Scale (SWLS)" (Natanael & Novanto, 2020) or on the website https://eddiener.com/scales/7. Since SWLS has been registered on the official website of Ed Diener. On Ed Diener's official website, the instrument was used. There are three translations of the Indonesian version of SWLS, and only one result has the translator's name detected. The Indonesian version of the SWLS has been researched and proven consistent, as confirmed by Natanael and Novanto (2020), where the instrument has a reliability value of .83. The items were also valid to measure life satisfaction, indicated by all items having a loading factor value above .40, and the fit model test was fulfilled with CFI = .996; TLI = .993 and RMSEA value < .05.

Research Procedure

The selection of modern construct-oriented scale construction is based on flexibility in testing when the developed instrument will be analyzed (John & Bent-

Martinez, 2013). The stages are as follows: (1) making a research hypothesis described in the introduction, which has not found a measuring tool for life satisfaction. Therefore, the research tries to develop a measuring instrument based on the same theory as the SWLS measuring instrument, (2) making the main or alternative model suitable for constructing the questionnaire. This stage is in accordance with previous research, where the appropriate measurement model for the life satisfaction variable is a unidimensional or one-factor model.

The next stage (3) creates items based on theory/definition/aspects accompanied by evidence of content validity. This section attaches the items developed to Table 2 of the instrument section, which initially developed 13 items. Content validity was carried out using the Aiken-V calculation, and each item was tested for feasibility with the assessment of 5 experts, where each assessed a range of 1 to 5. Therefore, five experts conducted the Aiken-V validity test with five assessment ranges, resulting in content validity in the form of a score, compared with the limit listed in the Aiken table. This is continued with (4) the process of collecting and analyzing data, using social media such as Instagram and Whatsapp with Rasch analysis.

Stage (5) tests the hypothesized model, one of the conditions that should be fulfilled in the analysis. Validation testing using Rasch analysis is expected to fulfil the proof of the unidimensionality model assumption, where the initial I-ALSS model is unidimensional. Finally, stage (6) produces a "good enough" model and describes the valid items. It reports the results of the analysis of the ten I-ALSS items and the testing of other measuring instruments. In modern construct-oriented scale construction, it is also important to discuss the limitations of the measuring instrument development process.

Data Analysis Technique

The analysis used the Rating Scale Model (RSM), suitable for polytomy data, such as data in the form of a Likert scale (Andrich, 1978). The steps are as follows: (a) data cleaning to detect outliers, (b) testing the

Table 2
Item Development

Indicator	Item	Initial Item Number	Aiken Item Number	Final Item Number
Assessment of life, in general, is ideal/good	All the needs in my life are fulfilled	Item 1	X1	F1
	My life is going well	Item 2	X2	
	I feel comfortable with my current daily routine	Item 3	X3	F2
Confidence to get a good life	I am optimistic that I can achieve a good life	Item 4	X4	F3
	My life will be full in the future	Item 5	X5	F4
	I spend a lot of time learning new things	Item 6		
Feeling satisfied with life	I enjoy the life	Item 7	X6	F5
	The life I live gives me a sense of comfort	Item 8		
	I am grateful for the life I have	Item 9	X7	F6
Feeling satisfied with important achievements in life	I can make my dream come true	Item 10	X8	
	I praise myself for my current achievements	Item 11	X9	F7
Desire to not want to change anything in life	The life I live is according to my wishes	Item 12	X10	F8
	I like the life I am living right now	Item 13		

unidimensionality assumption and local independence assumption, and (c) summary of fit statistics, which contains descriptions and explanations of person and item reliability, person and item separations, and model fit, (d) display an overview of the level of difficulty on items with a Wright Map, (e) diagnostic rating scale, (f) calibration items, and (g) differential item functioning (DIF) analysis to determine the level of bias in the instrument tested.

Results and Discussion

Result

I-ALSS Validity Evidence

Testing an instrument developed or reused in research, especially in Psychology and Education, has been agreed upon by *American Educational Research Association*, *American Psychological Association*, and *Nasional Council on Measurement in Education* (2014). Therefore, the instrument being tested fulfils the evidentiary criteria. The I-ALSS measurement tool has fulfilled three of the five proofs suggested by the AERA, APA, and NCME institutions.

The proof of the content validity of the I-ALSS uses the Aiken-V calculation obtained from the assessment of five experts. An item fulfils content validity based on the Aiken table when it has an Aiken value of .79 (Aiken, 1985). Only ten of the thirteen developed items fulfilled the content validity requirements as seen in Table 3.

Table 3
Calculation of Aiken 13 I-ALSS Initial Items

Item	Aiken Value	Result
Item 1	.791	Feasible
Item 2	.875	Feasible
Item 3	.833	Feasible
Item 4	.916	Feasible
Item 5	.958	Feasible
Item 6	.722	Delete
Item 7	.833	Feasible
Item 8	.667	Delete
Item 9	.875	Feasible
Item 10	.791	Feasible
Item 11	.916	Feasible
Item 12	.958	Feasible
Item 13	.708	Delete

Table 4
Comparison of SWLS with I-ALSS Based on Convergent Validity

Index	SWLS	I-ALSS
Lowest Loading factor	.532	.308
Highest Loading factor	.810	.748
Number of model modifications	0	7
Number of valid items	5	10

In subsequent analysis, only ten items were tested or proven. Item numbering is recycled from 1 to 13, numbering items before testing the validity of Aiken, to item numbers X1 to X10 (numbering items that pass the validity of Aiken), as observed in table 2.

The next proof is the internal structure of I-ALSS, and the correlation value of the item shows this evidence with the measured variable. Testing the I-ALSS internal structure was conducted by analyzing the 474 initial respondents. Correlation analysis shows that the value between items and the total variable of 10 I-ALSS is in the range of .503 - .695, meaning the items have good discriminating power values.

Furthermore, the instrument tested shows a relationship with other variables or instruments in the 474 initial respondents. This was proven through a convergent validity test using the Multi-Trait Multi-Method on the I-ALSS measuring instrument with SWLS. The analysis shows the fit test model with a value of 2 (82) = 304,939, RMSEA = .076, CFI = .962 and TLI = .951. The convergent validity test in Figure 1 shows that I-ALSS and SWLS have a very large relationship, indicated by the value of r = .949. Meanwhile, I-ALSS and SWLS measure the same indicators, namely life satisfaction.

The comparison of SWLS and I-ALSS is summarized in Table 4. SWLS does not have a single measurement error, however the 10 I-ALSS items were detected seven times measurement errors indicating by the model modification on the I-ALSS between items. However, this is not a significant problem because none of the items in the I-ALSS

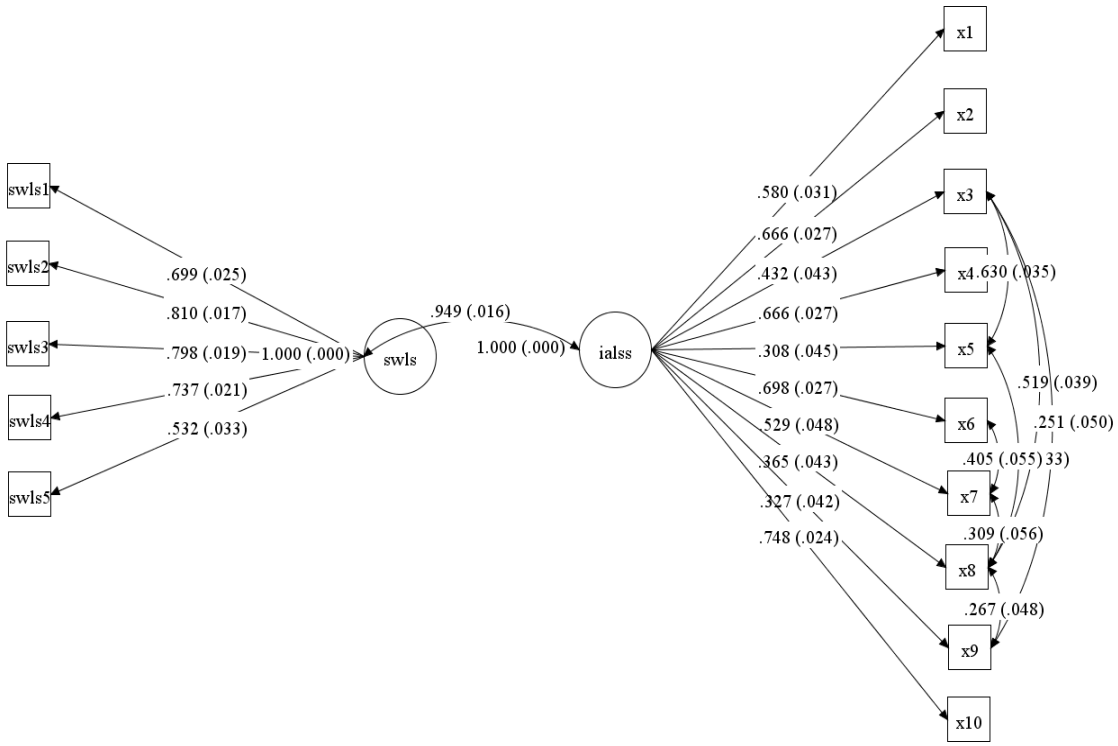

Figure 1. Convergent validity test of I-ALSS with SWLS

correlated more than three times with the others, and each had a good loading factor value of> .30 (Salsabila et al., 2019).

Evidence of validity has been presented before entering into the analysis of the Rasch model. The next step is to test the model, as the main test in this research

Rasch Model Test

Data Cleaning. The first step in testing the Rasch Model is removing the outliers of the research. In the data obtained, 54 out of 474 participants were detected as outliers. Generally, the ideal limit for detection is when the participants are outside the MNSQ outfit limit value of .50 to 1.50 (Boroel et al., 2017). In the Winstep program, there is person outfit limit > 2.00. For further analysis, only 420 participants were used for the Rasch Model.

Unidimensionality and Local Independence Assumptions. Data analysis using the Rasch Model requires two assumptions; namely, the measurement model is unidimensional and local independence. Holster and Lake (2016) stated that one way to verify the unidimensional model is to determine the value of "Variance Explained by Measure > 40%". The results on the I-ALSS instrument obtained a Variance Explained by Measure value of 46.5%, indicating that the I-ALSS is unidimensional. The first assumption is fulfilled in this research.

The second assumption is local independence, which can be interpreted as the level of relationship between the residuals and items. Christensen et al. (2017) revealed that the limit for items that did not experience local independence was the residual correlation value < .30. The residual correlation value has the term, namely the critical value Q3, and the limitations of the local value of independence are also fulfilled. The critical value of the ten I-ALSS items is fulfilled because the smallest to the largest critical values are -.25 to .26. Therefore, it is certain that the items fulfil the two initial assumptions to enter the Rasch Model analysis.

Fit Statistic and Reliability. Descriptively, the values in Table 5 show

several interpretations. First, the achievement test indicates that students have difficulty answering the questions when the person's mean value exceeds the item (Othman et al., 2015). The person means of 1.51 is greater than the item means = .00, indicating that the adolescent participants felt life satisfaction when tested on the attitude scale, especially in the developed I-ALSS. It can be further explained as follows: groups of people with high abilities are measured by a difficulty limit of a test. The person mean is said to have a high ability when the value is higher than the difficulty of the test. Similar to the attitude scale, when the value of life satisfaction is higher than the difficulty level of the instrument, the people tested have high life satisfaction.

The person's standard deviation value of 1.25 indicates that the level of life satisfaction varies. Meanwhile, the standard deviation item value of .76 indicates that the answer pattern spread can be quite varied. Diverse participants implies that life satisfaction levels vary from low to high. For the distribution of the pattern of answers that is quite varied, some participants choose from the option of 'strongly disagree' to 'strongly agree'. A positive value on the standard deviation item indicates that the answer pattern is more inclined to agree.

Table 5
Summary of Fit Statistics Index

	Person	*Item*
N	420	10
Measure		
Mean	1.51	.00
Standard Deviation	1.25	.76
Standard Error	.56	.08
Outfit Mean Square		
Mean	.99	.99
Standard Deviation	.46	.16
Separation	2.00	9.50
Reliability	.80	.99
Alpha Cronbach	.82	
Chi-Square	7646.64	

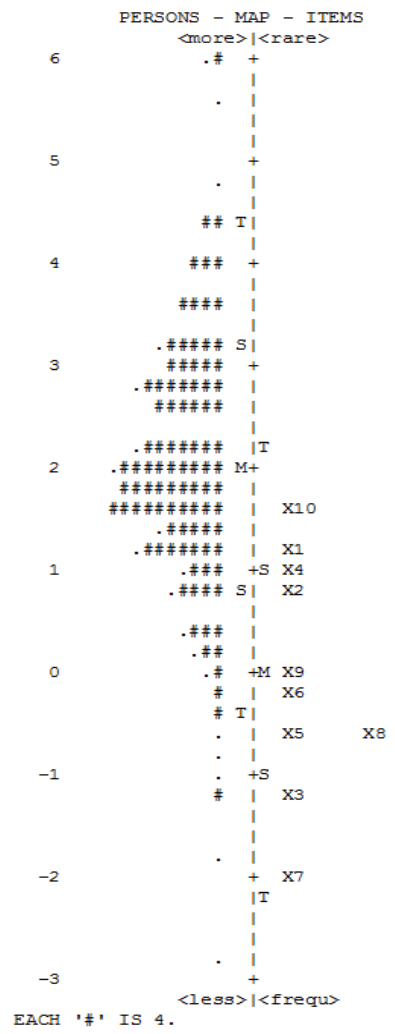

Figure 2. Wright map I-ALSS

Person separation of 2.00 indicates that this participant is homogeneous, and the item separation value of 9.50 explains the accuracy of item measurement on the variable. The resulting model fit index can also be seen from the Chi-Square value of 7646.64 and significant p-value < .000. Moreover, the Cronbach-Alpha value obtained was > .80, indicating a good I-ALSS consistency.

Wright Map. The Wright Map on the achievement test is used to indicate the level of difficulty of the questions. Meanwhile, the attitude scale shows easy or difficult items for participants. Figure 2 shows that item number X7 which reads: "I am grateful for the life I have", is easily agreed upon by the overall participants. Meanwhile, the item that is difficult to accept is item number X10 ("The life I live is in accordance with my wishes").

Easy to agree and hard to agree in this case is neither a criterion nor a group approach, but the participants' general assessment of the life satisfaction items answered.

Rating Scale Diagnosis. I-ALSS uses a Likert scale in its development process, meaning that five categories of answer choices will be described by considering the threshold value. Based on Van Zile-Tamsen (2017), the ideal threshold range between answer choices is 1.4 – 5.0 logit. The choice of answers on the items developed is appropriate or correct when the distance between the answers is 1.4 – 5.0 logit. The I-ALSS instrument has a threshold distance between 1.4 and 5.0 for each answer choice. A summary of the I-ALSS rating scale diagnosis can be seen in Table 6. For example, in the answer choices from "Disagree" to "Hesitating", the resulting threshold range is 2.04 logit (Value of -3.19 to -1.15 logit has a range of 2.04 logit). Another example is from the choice of "Agree" to "Strongly Agree" when carefully calculated, the resulting threshold range is 1.66 logit. This limitation is the evidence to check the appropriateness of five ranges of answer choices for the instrument tested. Based on the test results, the five answer choices are suitable for use in the I-ALSS.

Table 6
Rating Scale Analysis

Answer Options	*Observed Count*	*Percentage (%)*	*Observed Average*	*Rating Scale Threshold*	*Standard Error*
Strongly Disagree	29	1%	-1.59	None	None
Disagree	273	6.5%	-.17	-3.19	.21
Doubtful	1234	30%	.94	-1.15	.07
Agree	1471	35%	2.03	1.34	.04
Strongly agree	1143	27.5%	3.49	3.00	.04

Table 7
Calibration Item

Category	Item Number	*Logit Value*	*Standard Error*	*Outfit MNSQ*	*Point Mass. Corr*
Items are difficult to approve	X10	1.38	.08	.86	.71
	X4	.66	.07	1.02	.63
	X1	.57	.08	1.16	.53
Items in the moderate category (not difficult and also not easy)	X5	.22	.07	1.09	.56
	X2	.07	.08	1.06	.60
	X9	.05	.07	1.31	.53
	X3	-.24	.08	.81	.61
Items are easy to approve	X6	-.39	.07	.78	.69
	X8	-.84	.07	.94	.62
	X7	-1.47	.07	.85	.53

Table 8
DIF by Gender

Item Number	DIF		DIF *Contrast*	*t*	*Prob*
	Male	Female			
X2	-.05	.61	.66	-.325	.0001
X8	-1.21	-.77	.44	2.13	.0033

The resulting threshold value is also graded from negative to positive, which indicates that the answer choices from 'strongly disagree' to 'strongly agree' are appropriate. Meanwhile, five answer choices are appropriate and in the lowest to highest range.

Calibration Item. After obtaining the rating scale criteria regarding the range of appropriate answer choices, the research also grouped the ten I-ALSS items into 3 groups. The distribution limit follows the opinion by Wicaksono et al. (2021), which categorizes items based on a logit value greater than .31. This indicates difficult to approve, a logit value less than -.31 easy to approve, and logit -.31 to logit 3.1 is with a moderate level of difficulty. Table 7 shows that the three most difficult items to agree on are item numbers X10, X4, and X1.

Similarly, in the item calibration section, the ten I-ALSS items are good and measure the dimensions. This assessment is based on the statement by Boroel et al. (2017), which proposed that a good "item fit" range can be seen from the MNSQ outfit value from .5 to 1.5.

DIF Analysis. The DIF analysis on the I-ALSS instrument is the last step to determine the presence of gender bias in the developed instrument. The results show that two items, namely the numbers X2 and X8, are biased as indicated in Table 8. Item X2 and X8 have a DIF contrast value of .66 (male group -.05, female group .61) and .44 (male group -1.21, female group -.77). According to Rogers and Swaminathan (1990), bias items have limitations; namely, moderate DIF occurs when the difference contrast value is between .40 - .60 and high DIF when the difference contrast value is > .60 with a probability value of < .05 which proves the item can distinguish correctly or is biased in the tested group. Therefore, item numbers X2 and X8 were detected as biased items at high and moderate DIF. Table 8 shows the difference in DIF values from the male and female groups on item numbers X2 and X8. Therefore, the final total of testing items that are good and not experiencing gender bias in this analysis is 8 items with item number coding from F1 to F8 which can be seen in Table 2.

Discussion

The initial process was data cleaning, and 54 participants were detected as outliers. The outliers detection process is carried out by determining the MNSQ outfit value > 2.00, and 54 participants had a value from 2.33 to 4.08. The serial number detected as outliers by the Winstep program from the highest to the lowest are 108, 231, 366, 400, 93, 248, 336, 447, 472, 409, 253, 323, 227, 252, 232, 192, 392, 321, 328, 60, 363, 30, 388, 287, 458, 201, 234, 52, 44, 307, 297, 437, 419, 74, 103, 374, 176, 39, 29, 428, 408, 38, 341, 381, 416, 164, 122, 95, 217, 183, 344, 316, 110, and 58. There were no outliers with low scores below the limit; all participants were detected with high scores or exceeding the recommendation.

The number of outliers indicates that collecting data using online questionnaires can create a tendency to provide answers or fill out instruments randomly and not be serious about their work. Furthermore, the answers from 54 participants were not in accordance with the conditions experienced. The problem of outliers is in line with previous research that used online questionnaires, where in terms of quantity, the number was greater when compared to taking data directly (Natanael, 2021). Even though there are outliers, the fit model testing using Rasch is fulfilled after being deleted.

The most difficult item to approve was number X10 (out of 10 items after Aiken validation) or F8 (from the final item results), which reads, "The life I live is according to my wishes". Moksnes et al. (2016) reported that at the adolescent stage, education is fundamental, where they need a comfortable place to study in their psychological development. One of the problems experienced by all educators is the pandemic, which causes all adolescents to study at home. Since they study from home, it is difficult to answer item number X10 due to the alteration of living conditions. Learning at home may

cause adolescents to feel that their lives are not in accordance with their wishes. Besides, changing situations and pressures are one of the toughest stressors. However, this research is also regrettable because this has not been conducted in a situation outside the pandemic. It is highly recommended to conduct the subsequent analysis in a different situation.

The rating scale analysis on the I-ALSS shows a Likert scale that uses five suitable answer choices for the I-ALSS. Evidence related to the suitability of the range of answer choices allows participants to provide answers or fill out the I-ALSS more easily. It moves on to psychometric research on multiple answer choices. There is research comparing two (forced-choice), five (Likert scale), and seven answer choices (differential semantics), where the instrument uses a Likert scale (Fauzia, 2012). Therefore, this directly agrees with previous research, where the five answer choices (Likert scale) are suitable for the developed instrument.

In the I-ALSS, two items were detected that experienced gender bias. They favour the female group more, which is one of the limitations of this research. Scientific evidence suggests that testing for bias on items or instruments using the DIF technique should consist of 200 participants (Rogers & Swaminathan, 1990). Some studies also argued that testing using the Rasch Model could be conducted with 100 participants. According to Rogers and Swaminathan, the number of participants, especially the male group, is a limitation. This is due to the use of quota sampling, which collects as much data as possible regardless of the size of any group.

Demographic data showed that female participants were three times more than males. This implies that the proportion also allows the two biased items to favor the female group, as is the case with item number X2 “My life is going better” is more favourable for the female group. The score for the female group is higher, indicating increased life satisfaction for female adolescents when viewed from the DIF value. This proves the truth of research in the United Kingdom, where life satisfaction in females is higher than in males (Giusta et al., 2011).

Research on life satisfaction with the I-ALSS is also good for further analysis, focusing on testing cultural bias. Even with the differences in the participants’ regions of origin, it is very important to research to detect the level of bias in I-ALSS on culture. It is recommended for further analysis to control the culture of the participants, unlike in this research.

External validity testing was also carried out on the I-ALSS analysis as one of the five pieces of evidence suggested by the AERA, APA, and NCME institutions. External validity testing links the instrument to measure life satisfaction, namely SWLS. Internal validity testing using the Rating Scale Model describes the psychometric properties of only one instrument being analyzed. Meanwhile, the Rasch model analysis is focused on proving the instrument’s strength or weakness, namely I-ALSS. Testing with the Rating Scale Model shows that the I-ALSS is good in psychometric properties, as evidenced by the fulfilment of the analysis assumption test, good reliability value, item calibration of all items, the suitability of using a Likert scale for I-ALSS, and unbiased on the eight I-ALSS items.

In addition to external and internal validity, the I-ALSS proved to be feasible based on expert judgment. However, some factors need to be underlined since there is still a bias possibility in assessed items. Therefore, the items detected by this bias should be discarded in future research. Revising SWLS or the newly developed I-ALSS is recommended for analysis to test or develop the instrument. In terms of benefit, the newly developed items are not necessarily better than the old, and it is necessary to consider the principle of justice in the instruments used. Multiple groups should be tested and analyzed to get confidence in the developed instrument.

Additional suggestions are to control the culture and look for techniques to overcome it since online research has fewer outliers. It is

improbable that this will happen in research using quantitative data when no outliers are expected. Further investigation is hoped to find the best method to minimize outliers in online research.

Conclusion

Based on the analysis, it can be concluded that the development of the I-ALSS produces a fairly good instrument and has fulfilled several internal and external validity tests. The eight final items of I-ALSS ascertained that no gender bias is detected, meaning they are applied equally to the male and female groups. Therefore, the eight final I-ALSS items should be used to measure adolescents' life satisfaction in the field of Psychology.

References

Aiken, L. R. (1985). Three coefficients for analyzing reliability and validity of rating. *Educational and Psychological Measurement*, *45*, 131–142. https://doi.org/10.1177/0739986387009 2005

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Education Research Association.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Arifin, Z. (2017). Kriteria instrumen dalam suatu penelitian. *Jurnal Theorems (The Original Research of Mathematics)*, *2*(1), 28–36.

Arikan, Ç. A., & Zorbaz, S. D. (2020). Measurement invariance of the satisfaction with life scale across gender and time. *Turkish Journal of Education*, *9*(4), 260–272. https://doi.org/10.19128/turje.774452

Boroel, B., Aramburo, V., & Gonzalez, M. (2017). Development of a scale to measure attitudes toward professional values: An analysis of dimensionality using rasch measurement. *Procedia - Social and Behavioral Sciences*, *237*, 292–298. https://doi.org/10.1016/j.sbspro.2017.02.079

Borualogo, I. S., & Casas, F. (2022). Subjective well-being of children and adolescents during the COVID-19 pandemic in Indonesia: Two data collections. *Current Psychology*. https://doi.org/10.1007/s12144-022-03346-x

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the rasch model using residual correlations. *Applied Psychological Measurement*, *41*(3), 178–194. https://doi.org/10.1177/0146621616677520

Diener, E., & Diener, M. (1995). Cross-cultural correlates of life satisfaction and self-esteem. *Journal of Personality and Social Psychology*, *68*(4), 653–663. https://doi.org/10.1037/0022-3514.68.4.653

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75.

Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, *5*(6), 215–217. https://doi.org/10.15406/bbij.2017.05.00149

Fauzia, F. (2012). Perbandingan sosial desirabilitas bentuk skala likert, semantik diferensial, dan forced-choice dalam pengukuran prosocial tendencies. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, *I*(4), 263–282. https://doi.org/10.15408/jp3i.v1i4.10727

Fernández-Portero, C., Alarcón, D., & Padura, Á. B. (2017). Dwelling conditions and life satisfaction of older

people through residential satisfaction. *Journal of Environmental Psychology*, *49*, 1–7. https://doi.org/10.1016/j.jenvp.2016.11.003

Giusta, M. Della, Jewell, S. L., & Kambhampati, U. S. (2011). Gender and life satisfaction in the UK. *Feminist Economics*, *17*(3), 1–34. https://doi.org/10.1080/13545701.2011.582028

Google Scholar. (2022). *The satisfaction with life scale diener 1985*. https://scholar.google.com/scholar?hl=id&as_sdt=0%2C5&q=the+satisfaction+with+life+scale+diener+1985&btnG=

Holster, T. A., & Lake, J. (2016). Guessing and the rasch model. *Language Assessment Quarterly*, *13*(2), 124–141. https://doi.org/10.1080/15434303.2016.1160096

John, O. P., & Bent-Martinez, V. (2013). Measurement reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds), *Handbook of Research Methods in Social and Personality Psychology Measurement* (hal. 473–503). https://doi.org/10.1017/CBO9780511996481.023

Karnik, S., & Kanekar, A. (2012). Childhood obesity: A global public health crisis. *International Journal of Preventive Medicine*, *3*(1), 1–7. https://doi.org/10.1201/b18227-3

Moksnes, U. K., Løhre, A., Byrne, D. G., & Haugan, G. (2014). Satisfaction with life scale in adolescents: Evaluation of factor structure and gender invariance in a Norwegian sample. *Social Indicators Research*, *118*(2), 657–671. https://doi.org/10.1007/s11205-013-0451-3

Moksnes, U. K., Løhre, A., Lillefjell, M., Byrne, D. G., & Haugan, G. (2016). The association between school stress, life satisfaction and depressive symptoms in adolescents: Life satisfaction as a potential mediator. *Social Indicators Research*, *125*(1), 339–357. https://doi.org/10.1007/s11205-014-0842-0

Natanael, Y. (2021). Analisis rasch model Indonesia problematic internet use scale (IPIUS). *Persona: Jurnal Psikologi Indonesia*, *10*(1), 167–186. https://doi.org/10.30996/persona.v10i1.4827

Natanael, Y., & Novanto, Y. (2020). Pengujian model pengukuran congeneric, tau-Equivalent dan parallel pada satisfaction with life scale (SWLS). *Psympathic: Jurnal Ilmiah Psikologi*, *7*(2), 285–298. https://doi.org/10.15575/psy.v7i2.6405

Novanto, Y., & Pali, M. (2019). Teachers's life satisfaction in Palopo and Toraja: A descriptive study. *Jurnal Sains Psikologi*, *8*(2), 207–217. https://doi.org/10.17977/um023v8i22019p207

Othman, H., Ismail, N. A., Asshaari, I., Hamzah, F. M., & Nopiah, Z. M. (2015). Application of rasch measurement model for reliability measurement instrument in vector calculus course. *Journal of Engineering Science and Technology*, *10*(2), 77–83.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*(3), 517–529. https://doi.org/10.1037/0021-9010.87.3.517

Ridho, A. (2014). *Differential item functioning pada tes multidimensi* (Disertasi tidak diterbitkan). Universitas Gajah Mada, Fakultas Psikologi.

Rissanen, T. (2015). *Studies on life satisfaction in samples of the general population and depressive patients* (Dissertation unpublished). Faculty of Health Sciences, University of Eastern Finland.

Rogers, H. J., & Swaminathan, H. (1990). A comparison of logistic regression and

differential item functioning. *Applied Psychological Measurement*, *17*(2), 105–116. https://doi.org/10.1177/014662169301700201

Salsabila, D. F., Rofifah, R., Natanael, Y., & Ramdani, Z. (2019). Uji validitas konstruk Indonesian-psychological measurement of islamic religiousness (I-PMIR). *JPIB: Jurnal Psikologi Islam dan Budaya*, *2*(2), 1–10. https://doi.org/10.15575/jpib.v2i2.5494

Schnettler, B., Miranda-Zapata, E., Lobos, G., del Carmen Lapo, M., Adasme-Berríos, C., & Hueche, C. (2017). Measurement invariance in the satisfaction with life scale in Chilean and Ecuadorian older adults. *Personality and Individual Differences*, *110*, 96–101. https://doi.org/10.1016/j.paid.2017.01.036

Sufren, & Natanael, Y. (2014). *Belajar otodidak SPSS pasti bisa*. Elex Media Komputindo.

Tomás, J. M., Gutiérrez, M., Sancho, P., & Romero, I. (2015). Measurement invariance of the satisfaction with life scale (SWLS) by gender and age in Angola. *Personality and Individual Differences*, *85*, 182–186. https://doi.org/10.1016/j.paid.2015.05.008

Tresnawaty, Y. (2013). *Pendeteksian differential item functioning pada item dikotomus dengan menggunakan pendekatan item response theory (IRT), logistic regression (LR) dan confirmatory factor analysis (CFA)* (Tesis tidak diterbitkan). UIN Syarif Hidayatullah Jakarta, Fakultas Psikologi.

van Beuningen, J. (2012). *The satisfaction with life scale examining construct validity*. Statistics Netherlands.

van der Laan, S. E. I., Finkenauer, C., Lenters, V. C., van Harmelen, A. L., van der Ent, C. K., & Nijhof, S. L. (2021). Gender-specific changes in life satisfaction after the COVID-19–related lockdown in Dutch adolescents: A longitudinal study. *Journal of Adolescent Health*, *69*(5), 737–745. https://doi.org/10.1016/j.jadohealth.2021.07.013

Van Zile-Tamsen, C. (2017). Using rasch analysis to inform rating scale development. *Research in Higher Education*, *58*(8), 922–933. https://doi.org/10.1007/s11162-017-9448-0

Wicaksono, D. A., Roebianto, A., & Sumintono, B. (2021). Internal validation of the warwick-edinburgh mental well-being scale: Rasch analysis in the Indonesian context. *Journal of Educational, Health and Community Psychology*, *10*(2), 229–248. https://doi.org/10.12928/jehcp.v10i2.20260

Zanon, C., Bardagi, M. P., Layous, K., & Hutz, C. S. (2014). Validation of the satisfaction with life scale to Brazilians: Evidences of measurement noninvariance across Brazil and US. *Social Indicators Research*, *119*(1), 443–453. https://doi.org/10.1007/s11205-013-0478-5