

# **Calibration and Standardization of Happiness Measurement Through Rasch Models Based on Multiethnic Teenagers in Indonesia**

**Yenni Rizal<sup>1</sup>, Ahman<sup>2</sup>, Juntika Nurihsan<sup>3</sup>, Nurhudaya<sup>4</sup>**

<sup>1</sup>Students of the Doctoral Guidance and Counseling Program, Universitas Pendidikan Indonesia, Bandung, Indonesia

<sup>2,3,4</sup>Lecturer in Doctoral Guidance and Counseling Program, Universitas Pendidikan Indonesia, Bandung, Indonesia

[rizalyenni@yahoo.co.id](mailto:rizalyenni@yahoo.co.id), [ahman@upi.edu](mailto:ahman@upi.edu), [juntikanurihsan@upi.edu](mailto:juntikanurihsan@upi.edu), [nurhudaya@upu.edu](mailto:nurhudaya@upu.edu)

**Abstract:** The present research aims to analyze the scale of happiness measurement of multiethnic teenagers in Indonesia based on Rasch Model, the standardization of happiness scale which applies the Rasch Model analysis is assumed to produce more accurate, consistent and suitable results with the other standard instrument. The instrument is adapted from Satisfaction with Life Scale (SWLS) which developed by Diener. The determined research sample implements simple random sampling technique and Prety Random sampling generator application. The size of sample makes use of 99% confidence level and 5% margin of error. According to verification result, 606 of 610 answer sheets are capable to be processed. The result of Rasch Model based instrument analysis provides four main information, such as person-item map, item analysis, participants' ability analysis and instrument analysis. Implementing Rasch Model in the calibration and the standardization of happiness instrument have accomplished the required criteria and have produced instrument validity for each item. 11 items are not approved the standard criteria as the measuring instrument and 34 items are invalid as the formulation of happiness data. The value of Cronbach Alpha, which presents the interaction between person and overall specific item, achieves good category. Furthermore, the value of Person Reliability reaches fair category and the value of Item Reliability achieves excellent category.

**Keywords:** *happiness measurement, standardization and calibration, Rasch models*

## **INTRODUCTION**

Happiness becomes the final goal of human life through history. Aristoteles argues that all what human do is to reach happiness (Seligman : 2013). Happiness is very important topic and popular in psychology field in millennium era today, live happily means has positive feeling and comfort and doesn't have negative emotion (Demirci & Ekşi, 2018; Lee, 2017; Wilson, 2009). Happiness has three components, namely: satisfaction with life, positive influence and there is no negative influence (Diener, Sandvik, & Pavot, 1991; Tella & Adeniyi, 2009). Happiness is determined by genetic factor, relevant condition and activity related to happiness element (Lyubomksky, Sheldon, & Schkade, 2005). In the last several years, in a whole world studies on psychology instrument had been conducted to measure happiness construct such as Oxford Happiness Inventory, Satisfaction with Life Scale (SWLS), Positive and Negative Affect Schedule (PANAS Scale), etc. Various kind of data on happiness in the world based on survey result in international scope also had been collected from 2005 until this year in the form of world happiness report (John F. Helliwell, Richard Layard, 2018). This happiness report become broad data source for the researcher and it is studied to develop happiness instrument which is adjusted with various kinds of happiness indicator.

Various steps had been developed to measure happiness but not all instruments made suited with individual ability particularly teenagers in Indonesia who have diverse cultural background

or high heterogeneity level. Research mention that psychology theories actually related to culture-bound, local values and limited validity. These contextual limitations that make a psychological theory is not always strongly relevant if it is applied in another region or cultural context. The researchers conclude that each culture should be understood from its own reference frame including the existing ecological, historical, philosophical and religious contexts. Uchida in his study on happiness cultural construction found that there is difference in happiness meaning in the context of West culture ( individualistic) and East culture (collectivist) (Uchida & Norasakkunkit, 2004). Senturk argues that happiness level may keep on changing in social-demographical matter (Göksoy, 2017).

This happiness instrument based on Rasch model is the instrument designed for study which is aimed to reveal the happiness level among multiethnic teenagers. This instrument which is developed for data collection is adapted from Satisfaction with Life Scale (SWLS) developed by Diener. According to Diener, theoretically happiness consists of cognitive component and affective component (Diener, 2000). Adaptation of this concept is based on Diener's thinking who use subjective well being label that happiness is the combination of cognitive assessment (satisfaction with life) and affective evaluation (emotional experience) (Lyubomksky et al., 2005).

Scale to measure this happiness level require an analysis which ensure that measurement result will be more accurate, consistent and suits with another instrument standard. The standardization of this happiness scale uses Rasch model analysis. This model probes whether instrument item can be suited with individual ability range, so this instrument is believed to reflect happiness level among teenagers from various ethnics. Rasch model analysis use probabilistic model that individual with ability level which is higher than another individual must has more change to answer one item correctly. With same principle, more difficult item makes individual has less change to be able to answer that item (Sumintono, 2018). Rasch analysis shows more advantages compared to another method (Rasch, n.d.), The advantage of Rasch model is its capability to predict missing data, which is based on systematical response pattern. Besides, Rasch model capable to generate error standard measurement value for instrument used which can increase calculation accuracy. Rasch analysis provide systematical method which is tightly and properly can diagnose scale and individual item in the case of potential item bias or differential item function (DIF) and uni dimensionality (Medvedev, 2017).

## **METHOD**

### ***Participant***

Calibration and standardization of instrument to reveal happiness data involve population of 6700 students distributed in 5 SMAN (Public Senior High School) and 2 SMKN (Public Vocational Secondary School) in Pontianak City. The selection of school is based on consideration that those seven schools have students from all ethnics exist in Pontianak. Participants of study are selected by using simple random technique with generator sampling application Prety Random. The size of minimal sample is based on Krejcie-Morgan Table. Refer to this table, by using 99% confidence level of 99% and 5% margin of error, then the size of minimal sample ranged between 586-610. To ensure that the size of minimal sample is fulfilled then sample is determined as many as 610 students. Based on verification result, from 610 answer heets which are adequate to be processed are 606. Distribution of participants involved in calibration and standardization of instrument is as follow.

**Table 1.** Distribution of Participants

Ethnic	Senior High School (SMA)														Total P/L		Total
	SMK 3		SMK 8		SMA 5		SMA 4		SMA 2		SMA 8		SMA 9				
	P	L	P	L	P	L	P	L	P	L	P	L	P	L	P	L	
Java (J)	8	1	8	1	13	1	13	3	6	6	16	13	16	12	80	37	117
Madura (M)	3	0	20	5	1	2	7	4	0	1	2	1	0	3	33	16	49
Melay (U)	13	7	3	7	10	7	23	19	9	8	9	22	23	18	90	88	178
Chinese (C)	19	18	27	9	12	6	2	2	1	0	4	3	5	1	70	39	109
Bugis (B)	7	3	5	1	7	0	10	4	2	0	3	8	7	6	41	22	63
Dayak (D)	2	3	9	4	15	17	2	0	0	3	20	10	2	3	50	40	90
Total	52	32	72	27	58	33	57	32	18	18	54	57	53	43	364	242	606
	84		99		91		89		36		111		96		606		

**Instrument**

The concept of happiness scale in this study refer to a theory proposed by Diener which is adapted from Satisfaction with Life Scale (SWLS) (Carr, 2004). The adaptation of this concept is based on Diener’s thinking who use subjective well-being label in which happiness is combination of cognitive assessment (satisfaction with life) and affective evaluation (positive and negative emotional experience) (Akın, Aslan, Çelik, Kaya, & Aslan, 2015; Arias Gallegos, Huamani Cahua, & Caycho-Rodríguez, 2018; Koç & Pepe, 2018).

**Table 2.** Components Of Subjective Well-Being

Domain	Cognitive Component		Affective Component	
	Satisfaction		Positive Affect	Negative Affect
Self	Significant others’ view of one’s life		Happiness	Depression
Family	Satisfaction with current life		Elation	Sadness
Peer Group	Significant others’ view of one’s life		Ecstasy	Envy
Health	Satisfaction with past		Pride	Anger
Finances	Satisfaction with future		Affection	Stress
Work	Desire to change life		Joy	Guilt or shame
Leisure	Satisfaction with current life		Contentment	Anxiety

Source: Adapted Form Diener *et al.*

This happiness instrument consists of two components, namely cognitive and affective with total items are 34 statements. Each component has statement item which consists of 25 item of cognitive component and 9 items of affective component. Item score is adjusted with positive statement and negative statement with alternative of answer consists of 5 choices namely very fitted, less fitted, unfitted and very unfitted.

Calibration in this Rasch modeling has calibration requirement which should be fulfilled concurrently in instrument measurement in three things, namely: measurement scale, respondent (person), and item. Calibration of happiness instrument which had been designed will produce valid data and can make item analysis activity or research conducted generates what is desired. The final product of Rasch modeling is making happiness measurement scale with same interval, because raw score doesn't have interval property which makes the score is not used directly to interpret student ability. Rasch modeling concurrently use score data based on personal or score data per item. Both scores become basis to estimate true score which shows individual ability level and item difficulty level. (Bambang & Wahyu ; 2015).

### ***Procedure***

Questionnaire of this study filled by students from seven multiethnic Senior High School distributed in Pontianak, West Kalimantan. Data from sample in Senior High School is obtained through class representative which has students with different ethnic background. All students from all ethnics are asked to fill questionnaire and after data had been collected, six ethnics are selected to become study sample. They are Malay, Dayak, Chinese, Bugis, Madura and Javanese. The numbers of data which are collected to become study sample and meet the criteria to be processed are 606 instrument sheets.

### ***Data Analysis***

Rasch model at least give four main information in instrument calibration and standardization, namely person-item map, item analysis, participant ability analysis, and instrument analysis. The information obtained on these things is presented in following explanation, This Wright Map analysis aims to obtain description of construct map and distribution of participant ability and distribution of item difficulty by using same scale namely logarithm odd unit which is usually abbreviated as *logit*. Rasch modeling estimate probability that each response category of each individual item is supported by considering item difficulty (location) and one's ability in latent property measured ( that is, happiness), whereas non-parametric method does not discriminate between difficulty and ability (Medvedev, 2017). This Wright Map analysis can be done based on outlier information in Table 1 Variable Maps from Rasch Model menu with Winstep Program.

To find out item difficulty level, Table 13 Item Measure is used. From that table, it obtains mean of logit item and standard deviation. With analysis result of Person-Item Map, it will obtain information on items which are too difficult and items which are too easy. Data on ability level or participant ability can be obtained from Output Table, namely Tabel 17 Person Measure. From that table, it is known that mean value of logit person and Deviation Standard. Person Fit detect whether there is or there is no participant whose response pattern is not fitted or different. Different response pattern is because of the answer given is not fitted with the ability compared to ideal model. It also tests participant consistency in thinking. To know this Person Fit then Output Tables, that is Table 6 Person Fir Order is used. The criteria used is as in item fit level analysis.

Research instrument analysis is presented in Rasch Model in Output Tables, namely Tabel 3 Summary Statistics.

## RESULTS

### 1) Wright Map Analysis: Person-Item Map

Based on map information in Variable Maps Table, it is seen that mean of logit person ( $M=0.60$ ) is higher than mean of item logit ( $M=0.00$ ). It means that in average, participant ability is above item standard difficulty level. There are 16 participants included in outlier category (more than  $+2SD$  or less than  $-2SD$ ), with detail as follow: 15 participants who have high ability with logit value is above  $+2SD$  (T), namely 255JP, 573CP, 213CL, 309JP, 172DL, 313JP, 415BP, 063UL, 008JL, 055UL, 115UL, 256JP, 397UP, 420BP, 602CP, and 1 participant has logit value under  $-2SD$  (T), namely 197ML. Whereas the others namely 560 students have logit value  $-2SD$  until  $+2SD$ . Therefore, those 16 participants will not involved in study data processing.

Based on map information in Table Variable Maps, it is seen also that logit value for item difficulty level distributed for each level. It means that for each group of very easy item ( $0.0$  logit  $-1SD$ ), easy item ( $0.0$  logit  $-1SD$ ), difficult item ( $0.0 + 1SD$ ) and very difficult item (more than  $0.0 + 1SD$ ). Only three items which are outlier, namely two items which have logit value under  $-2SD$  (T) namely 10005 and 10015 (logit value of each are  $-1.59$  and  $-1.62$ ) and one item which has logit value above  $+2SD$  namely item 10002 with logit value is  $1.47$ . Therefore, those three items will not involved in study data processing. Item 10002 and 10005 are cognitive dimension, self domain, description of satisfaction with self condition based on others' view. Whereas, Item 10015 is cognitive dimension, peer domain, description of satisfaction with peer relationship.

### 2) Item Difficulty Level (Item Measure)

Table 13 Item Measure shows that mean of logit item is  $0.00$  and deviation standard is  $0.72$ . In accord with analysis item of Person-Item Map in item 4.1 above, the information is obtained that there is one item which is too difficult (Item 1002) and two items which are too easy (Item 10006 and 10015).

Refer to four categories in interpreting item difficulty level, it is known that there are 7 items included in very difficult category (including item 10002 which is outlier), 18 items included in difficult category, 13 items included in easy category, and 7 items included in very easy category (including 2 items, namely 10005 and 100015 which are outlier).

Item fit level is quality of item fit with model. The meaning of item fit explains whether or not an item functions normally in doing measurement. An item which is not fit indicates that there is misconception in participant toward that item. Table which presents it from Menu Output is Table 10 Item Fit Order. The criteria used to see item fit level is: (1) outfit mean-square value (MNSQ) should more than  $0.5$  and less than  $1.5$ , (2) outfit Z-standard (ZSTD) should more than  $-2.0$  and less than  $2.0$  and (3) point measure correlation (Pt Mean Corr) should more than  $0.4$  and less than  $0.85$ . It is better than ZSTD value is not made to become reference if sample size used more than 500 sampling unit. An item is considered fit with model or has function in doing measurement if it minimally meets one of three criteria above. Therefore, the instrument calibration in this study use sample size of 606 and MSSQ and Pt Mean Corr become criteria to interpret item fit.

Refer to Table 10, it is known that Item 10015 did not meet the standard of item fit with model because it has MNSQ value =  $1.64$  and Pt Mean Corr= $0.13$ . Detection of item bias detection in instrument developed in this study is seen from two things, namely gender and ethnic. An item contains bias if one participant with certain characteristic is more benefited compared to participant with another characteristic. It is through detection of differential item

functioning (DIF) in Table 30 Item DIFF, particularly Table 30.4. An item contain bias if its item probability value less than 0.01.

Refer to Table 30.4 for bias analysis based on gender, ethnic and combination of gender and ethnic, it is known that there is item which has probability less than 0.01, namely 11 item bias due to gender, 3 bias due to ethnic, and 7 bias due to combination of gender and ethnic. The number of item which contain bias is presented in following table.

**Table 3. Item Bias**

<b>No. Item Bias</b>			
<b>Gender (A)</b>	<b>Ethnic (B)</b>	<b>Combination (AB)</b>	<b>Decision</b>
2		2	Not Used
5		5	Not Used
11		11	Not Used
15		15	Not Used
17	17	17	Not Used
18	18	18	Not Used
21	21	21	Not Used
29			Not Used
33			Not Used
34			Not Used
42			Not Used

3) Participant Ability Analysis

The data of participant ability analysis can be obtained from Output Table, namely Table 17 Person Measure. From that table, it is known that mean value of logit person is 0.6 logit and SD is 0.35. If it is related to Person-Item Map as said in earlier explanation, it seems that mean of logit person is higher compared to mean of logit item, namely 0.00. It indicates that in general, participant ability is higher compared to item difficulty level. As consequence, it is supposed that many participants tend to have high value. This supposition is strengthened by the form of person ability curve in Person-Item Map with frequency of most participants is above the mean of logit item.

Refer to Table 6 Person Fit Order with its interpretation criteria, it obtains information that there are 176 participants who give answer which is not fitted with their ability compared to ideal model. The detail is 99 participants who have outfit MCNSQ value is more than 1.5 and 77 people who have outfit MCNSQ value less than 0.5.

In study data analysis process, all participants included in this group are not made to become study sample. Supporting Olsen study which concluded that the level of students success in doing items very depends on their ability level and item difficulty level (Lina Wøhlk Olsen, n.d.)

4) Instrument Analysis

Rasch Model facilitates research instrument analysis as presented in Output Tables, namely Table 3.1 Summary Statistics. Based on Table 3.1, it knows the information as follow:

**Table 4. Summary Statistic**

	<b>Mean</b>	<b>SD</b>	<b>Separation</b>	<b>Reliability</b>	<b>Cronbach Alpha</b>
Person	0,60	0,35	1,63	0,73	0,76
Item	0,00	0,72	14,44	1,00	

Person measure = 0.60 logit shows score mean of all participants in doing the items of happiness data instrument. The mean score of person which is bigger than mean score of item (mean score of item is 0.000 logit) shows that in general participant ability is higher compared to item difficulty of instrument.

Cronbach Alpha value, which represent interaction between person and items totally is 0.76 which is included in good category. Next, Person Reliability value of 0.73 is included in medium category whereas for Item Reliability is 1.00 which is included in special category.

Another data in Table 3.1 which can be used is INFIT MNSQ and OUTFIT MNSQ both in Person Table and Item Table. Based on Person Table, it is known that mean score of INFIT MNSQ is 1.02 and for OUTFIT MNSQ is 1.03. The criteria is that the more it approaches number 1 then it is better, because the ideal value is 1. Therefore, both mean of person and mean of item approach the ideal criteria. Meanwhile, in the case of INFIT ZSTD and OUTFIT ZSTD, mean score for person and item is similar, namely -0.2. The ideal value is 0, so the more it approaches 0 then it is better. Therefore, it can be said that the quality of person and item is good.

The last is related to separation or grouping of person and item. From output of Table 3.1, it is known that separation for person is 1.63 and for item is 14.44. The bigger of separation value, the better of person and instrument quality in a whole. Separation value is calculated more thoroughly by formula:  $H = \{(4 \times \text{separation}) + 1\} / 3$ . Thus, separation value for person is 2.507 rounded to become 3, whereas separation for item is 19.59 rounded to become 20. It means that the participants of study are diverse in their ability which can be categorized into three groups. Whereas, item difficulty level is distributed in 13 groups started from the group of easiest items until the group of most difficult items.

## CONCLUSION

Calibration and standardization of happiness instrument by using Rasch model had met standardized criteria and generate instrument validity in each item. There are 11 items which are less adequate to meet standard criteria as measurement tool and there are 34 items which are adequate to be used in the instrument to reveal happiness data. Cronbach Alpha value which represents interaction between person and items in a whole, included in good category. Next, Person Reliability value is included in enough category whereas Item Reliability is included in special category.

## REFERENCES

- Akın, A., Aslan, S., Çelik, E., Kaya, Ç., & Aslan, N. (2015). Student Academic Support As a Predictor of Life Satisfaction in University Students. *Eurasian Academy of Sciences Social Sciences Journal*, 2(1), 38–49. <https://doi.org/10.17740/eas.soc.2015-v2-04>
- Arias Gallegos, W. L., Huamani Cahua, J. C., & Caycho-Rodríguez, T. (2018). Satisfacción con la vida en escolares de la ciudad de Arequipa. *Propósitos y Representaciones*, 6(1), 381–407. <https://doi.org/10.20511/pyr2018.v6n1.206>
- Demirci, İ. , & Ekşi, H. (2018). Keep Calm and Be Happy: A Mixed Method Study from Character Strengths to Well-being. *Educational Sciences: Theory & Practice*, 18(2), 279–330. <https://doi.org/10.12738/estp.2018.2.0799>
- Diener, E. (2000). Subjective Well-Being. *American Psychologist*, 55(1), 34–43. <https://doi.org/10.1037//0003-066X.55.1.34>

- Diener, E., Sandvik, E., & Pavot, W. (1991). Happier is frequency, not the intensity, of positive versus negative affect. In *American Psychologist* (Vol. 55). <https://doi.org/10.1037/0003-066X.55.1.34>
- Göksoy, S. (2017). Situations that Make Students Happy and Unhappy in Schools. *Universal Journal of Educational Research*, 5(12A), 77–83. <https://doi.org/10.13189/ujer.2017.051312>
- John F. Helliwell, Richard Layard, and J. D. S. (2018). *World Happiness Report* (R. L. and J. D. S. John F. Helliwell & This, eds.). Retrieved from <http://wordhappiness.report/>
- Koç, K., & Pepe, O. (2018). The Investigation of the Relationship between Happiness Levels of the Faculty of Sports Sciences and the Levels of Life Satisfaction and Optimism. *World Journal of Education*, 8(6), 74. <https://doi.org/10.5430/wje.v8n6p74>
- Lee, J. (2017). *Happiness and Ethical Values in Higher Education*. 1–16.
- Lina Wøhlk Olsen. (n.d.). *Essays on Georg Rasch and his*.
- Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology*, 9(2), 111–131. <https://doi.org/10.1037/1089-2680.9.2.111>
- Medvedev, O. (2017). *Measuring Mindfulness and Health-Related Outcomes : Applying Rasch Analysis and Generalisability Theory to Improve Reliability and Validity of Scales Measuring Mindfulness and Health-Related Outcomes : Applying Rasch Analysis and Generalisability Theory t*.
- Rasch, G. (n.d.). *ON GENERAL LAWS AND THE MEANING OF MEASUREMENT IN I i Iti*. (v).
- Sumintono, B. (2018). Rasch Model Measurement as Tools in Assessment for Learning Rasch Model Measurement as Tools in Assessment for Learning. *PROFESSIONAL DEVELOPMENT FOR SECONDARY SCHOOL PHYSICS TEACHERS THROUGH ACTION RESEARCH View*, (October 2017). <https://doi.org/10.2991/icei-17.2018.11>
- Tella, A., & Adeniyi, O. (2009). Cypriot Journal of Educational. *Happiness in High School Students: Autonomy, Relatedness, Competence and Meaning in Life*, 4(3), 168–182. Retrieved from [www.cjes.eu](http://www.cjes.eu)
- Uchida, Y., & Norasakkunkit, V. (2004). Cultural constructions of happiness: theory and empirical evidence. *Journal of Happiness Studies*, 223–239.
- Wilson, T. D. (2009). TARGET ARTICLE : The Message Is the Method : Celebrating and Exporting the Experimental Approach The Message Is the Method : Celebrating and Exporting the Experimental Approach. *Psychological Inquiry*, 16(911768352), 185–193. <https://doi.org/10.1207/s15327965pli1604>
- Seligman, M. (2013). *Authentic happiness*. Bandung: Mizan. Carr, A. (2004). *Positive psychology*. New York: Brunner-Routledge.