

# Validity and reliability of Islamic Spiritual Well-being Index for Adolescent (IKSIR) using Rasch analysis

Sakinah Ab. Rahim, Siti Aisyah Panatik, Siti Norlina Muhamad  
Faculty of Social Science and Humanities  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia

**Abstract:-** The aim of this study was to examine the degree of validity and reliability of a new developed instrument. The instrument was developed to measure the level of spiritual well-being among Muslim adolescence in Malaysia. Throughout this study, the empirical evidence of validity and reliability of this instrument by using Rasch Model was discussed. The instrument measure the level of spiritual well-being by four constructs which are faith, knowledge, worship and morals. A survey was conducted to 298 form four students from three Islamic based secondary schools in Johor. Statistical analysis using Winsteps software was performed to measure the validity and reliability of 71 items in the instrument. Results showed the reliability of the IKSIR items are between 0.93 to 0.99. The separation index of item was above 2; all items indicated positive value in polarity item analysis while six items were identified with the value of MNSQ infit and outfit beyond the standardized range of 0.5 to 1.5 should be dropped from the instrument. Thus, the degree of validity and reliability of IKSIR were high according to the Rasch model of measurement.

**Keywords:-** *Spiritual well-being, Rasch Model, validity, reliability, separation index, item fit analysis.*

## I. INTRODUCTION

There is continued interest in examining the implication of spiritual well-being on quality of life especially with health-related outcomes [1, 2, 3]; job satisfaction [4, 5]; life satisfaction [6] and much more. Despite the interest, non-generalizable samples and the lack of appropriate instruments that measure spirituality and/or religiosity compromise most studies in this field [7]. Researchers from different background of religion and culture facing difficulty in selecting a reliable spiritual instrument to suit with their purpose of study and the samples' background especially to the Muslims [8, 9] because most of the existing instruments were developed from Judeo-Christian perspective [10] totally different in terms of God's concept (aqidah) and the way of worships [11]. Thus, this instrument being developed to broaden the alternative selection of spirituality instruments for a spiritual study from Islamic perspective.

The Islamic Spiritual Well-being Index for adolescent (IKSIR) is a new developed written self-report measure in Malay language designed to measure the adolescent's level of spiritual well-being. It measures how understanding, awareness and practices of four construct which were faith

(aqidah), knowledge ('ilm), worship (ibadah) and morals (akhlaq) influenced the level of spiritual well-being which resulting a good quality of life. In the early stage of development, the instrument was consists of 81 items but after two series of pilot studies, 10 items were omitted. Thus, this survey was conducted to finalize the instrument before it can be used for public. The construct of faith was operationalized through the six pillars of faith which are having faith in God (Allah), angels (malaikat), prophets, holy books (kitab), the Day of Last Judgement (qiamah) and God's judgment (qada' and qadar). While knowledge was operationalized in general; no specification or categorization because we do not want the respondent to be confused with its philosophy which is required a higher level of understanding and thinking. Thus, this construct just measure the level of understanding, awareness and practices of the concept of general knowledge. The construct of worship (ibadah) was operationalized by two types of worship which were compulsory and non-compulsory (sunnat). We limited the compulsory ibadah only to five times prayers in a day (solah fardhu) and Ramadhan fasting while the non-compulsory ibadah are limited to sunat prayers and fasting, reciting Quran and doing charity. While the construct of morals adapted from [12] was operationalized by nine types of ethics that are honest, trustworthy, talk genuinely, speaks exquisitely, shyness, braveness, cooperation, generosity and politeness.

The main aim of this study is to establish the validity and reliability of the IKSIR items. The validity and reliability value is vital to maintain the accuracy of the instrument. The higher the value of validity and reliability, the more accurate the data would be [13, 14]. Reliability refers to the consistency of the item in measuring the constructs in different situation while validity refers to the extent of instrument measuring what it should be measures [15, 16, 17, 18]. In this study, all statistical procedures had done using the Rasch measurement model.

## II. METHOD

### A. Sample

A total number of 298 students from three Islamic based secondary schools in the state of Johor, Malaysia were involved in this study. They were selected using the stratified sampling method. Out of 298 students, 41.3% (N=123) were male while the rest that is 58.7% (N = 175) were female. All the respondents were Muslims, Malays, and aged 16.

**B. Data Collection and Analysis**

The first draft of IKSIR instrument contains 81 items consisting of 4 constructs. The back and forth processes of validating the instrument content with the experts and pilot tests, ten items were dropped. Thus, the instrument was used in this survey contains 71 items. The instrument was distributed to the respondents during their teaching and learning session. The test was administered for about half an hour where about five minutes were taken by the researcher to give instruction to the respondents. All respondents successfully completed the test within the time given. To analyze the data, all items were coded and scored. Statistical analyses of the instrument were performed using the Rasch measurement model version 3.72.3. Reliability and separation indexes were performed to determine the degree of instrument reliability. In the other hand, dimensionality, item fit, and polarity item analysis were performed to examine the validity of the instrument.

**III. RESULT AND DISCUSSION**

Using Rasch Model, researchers performed dimensionality and item fit statistics analysis to validate the developed instrument through this study.

**A. Dimensionality Analysis**

Dimensionality analysis is a test carried out to examine the dimensionality properties of the instrument. Uni-dimensional means all the items really measure the construct at a time. Table 1 shows the result for dimensionality test of the IKSIR.

	Empirical		Modeled	
Total raw variance in observations	114.7	100%	100%	
Raw variance explained by measure	45.7	38.1%	42%	
Raw variance explained by persons	15.1	13.1%	14.5%	
Raw variance explained by items	28.6	24.9%	27.5%	
Raw unexplained variance (total)	71.0	61.9%	100%	58%
Unexplained variance in 1 <sup>st</sup> contrast	4.7	4.1%	6.6%	
Unexplained variance in 2 <sup>nd</sup> contrast	4.1	3.5%	5.7%	
Unexplained variance in 3 <sup>rd</sup> contrast	2.7	2.4%	3.8%	
Unexplained variance in 4 <sup>th</sup> contrast	2.5	2.2%	3.6%	
Unexplained variance in 5 <sup>th</sup> contrast	2.2	1.9%	3.1%	

Table 1. Dimensionality Analysis

From the table, the empirical raw variance explained by measure is 38.1% while the unexplained variance in first contrast is 4.7 units or 4.1%. Linacre in [19] stated that the value of raw variance explained by measure which is more than 60% indicates that the instrument is very good, while the value less than 20% considered very bad. Hence, the value range between 20% up to 60% considered fair; means the instrument is acceptable. Whereas the value of empirical unexplained variance in first contrast is 4.7 units (4.1%) which indicates the residual items in the instrument is 4. Linacre [19]

stated that the minimum value for unexplained variance in first contrast should be 5 or less; giving the percentage of 5% to be weighted to form another dimension or construct. Thus, the result validated that the IKSIR is a unidimension instrument with four constructs as conceptualized and operationalized by the researchers.

**B. Fit Statistics**

Statistical analysis for suitability of items were carried out to identify items that should have the value of mean-square (MNSQ) in fit and outfit ranged from 0.5 to 1.5 ( $0.5 \leq MNSQ \leq 1.5$ ). Generally, MNSQ show the randomness; the amount of distortion of the measurement system. In fit means inlier sensitive fit. The in fit more sensitive to the pattern of responses to items targeted on the person, and vice-versa. While outfit means outlier sensitive fit. The outfit value shows the sensitiveness towards responses to items difficulty for a person, and vice-versa. There are some ideas that the value of MNSQ outfit must come first before MNSQ infit and vice versa in order to get better evaluation of the item fit analysis [16, 18]. In this study, we were being neutral where we look at the MNSQ value of both infit and outfit to identify either the items were fit or not with the Rasch Model to be fit in the final instrument. Through this analysis, the results for item polarity also were obtained (as shown in point-measure correlation or PTMEA Corr. column). Table 2 to Table 5 below shows the results for item fit analyses for each every item according to the constructs.

Item	MNSQ		PTMEA Corr.
	Infit	Outfit	
9	3.41	4.57	0.13
12	0.83	0.72	0.46
3	0.78	1.02	0.41
13	2.90	4.20	0.15
11	1.07	1.10	0.36
7	1.10	1.05	0.31
29	1.72	5.06	0.07
6	2.36	1.76	0.30
20	0.53	0.89	0.32
24	1.27	0.73	0.27
25	0.62	0.78	0.30
14	1.05	0.95	0.24
23	1.83	2.10	0.27
28	0.67	0.80	0.26
26	0.96	0.77	0.23
17	1.19	0.77	0.23
19	1.24	0.63	0.24
16	0.83	0.78	0.20
21	0.79	0.47	0.26
15	1.04	1.21	0.25
1	0.96	0.47	0.23
Mean	1.04	0.96	
SD	0.37	0.32	

Table 2. Faith Item Fit Statistics

Table 2 shows the value of MNSQ for infit and outfit of all items that formulated the construct of faith. Also shown in the table are the values of point measure correlation (PTMEA Corr.); the result for polarity item analysis. Polarity item analysis is a test carried out to identify either the items moving in one direction; other words, strictly measuring the construct [18]. The item is measuring the construct if the PTMEA Corr.

value is in ranged of 0.20 to 0.79 ( $+0.20 \leq \text{PTMEA Corr.} \leq +0.79$ ) [16, 18]. From the table, there are five problematic items detected that were item 6, 9, 13, 24 and 29. Item 9, 13 and 29 had the value of MNSQ infit and outfit bigger than the standard range and also had the value of PTMEA Corr. less than 0.20 which mean the item does not measuring the construct so they had to be dropped. In the meantime, item 29 shown a bigger value of MNSQ infit and outfit than the standard range so it had to be dropped to ensure the instrument is reliable [18].

Item	MNSQ		PTMEA Corr.
	Infit	Outfit	
42	1.37	1.36	0.60
41	0.92	0.94	0.69
40	0.86	0.82	0.67
38	0.94	0.81	0.63
34	1.00	1.19	0.55
36	0.90	0.78	0.62
35	0.64	0.60	0.66
37	0.75	0.63	0.64
31	1.21	1.45	0.37
39	0.82	0.62	0.50
32	1.27	1.34	0.31
Mean	1.05	1.01	
SD	0.34	0.39	

Table 3 Knowledge Items Statistics

As in Table 3, the result shows no problematic item detected. All items have the value of MNSQ infit and outfit in the standardized range so do the PTMEA Corr. value. Hence, no item should be dropped from the finalized instrument as suggested by the Rasch measurement model.

Item	MNSQ		PTMEA Corr.
	Infit	Outfit	
44	1.43	1.45	0.52
53	0.90	0.99	0.57
47	2.06	2.25	0.33
48	1.10	1.06	0.53
50	0.92	1.08	0.46
49	0.71	0.84	0.53
46	1.28	1.36	0.39
52	0.89	1.00	0.46
60	0.64	0.64	0.56
57	1.37	2.37	0.28
43	1.00	1.14	0.45
56	1.40	1.49	0.36
54	0.66	0.62	0.49
55	0.71	0.54	0.46
51	0.78	0.60	0.45
45	0.98	1.00	0.38
58	0.80	0.64	0.42
59	0.68	0.53	0.45
Mean	0.99	1.02	
SD	9.37	0.42	

Table 4 Worship Item Fit Statistics

Table 4 shows the result for worship item analysis. There are two problematic item detected that are item 47 and 57. Item 47 had the value of MNSQ infit and outfit beyond the standardized range while item 57 had the value of MNSQ outfit bigger than 1.5. In the other hand, all items showed the

value of PTMEA Corr. in standardized range. Hence, the two problematic items (item 47 and 57) being considered to be dropped from the instruments as suggested by the Rasch Model. In contrast, the experts suggested to maintain those items because the items is good (the PTMEA Corr. value for each item is good) in measuring the construct.

Item	MNSQ		PTMEA Corr.
	Infit	Outfit	
67	1.46	1.29	0.32
68	1.07	1.19	0.58
69	0.81	0.92	0.64
61	1.09	1.15	0.56
71	1.43	1.48	0.40
64	1.01	1.22	0.54
65	0.77	0.79	0.62
70	0.75	0.74	0.66
77	0.69	0.69	0.66
75	0.79	0.82	0.61
76	0.81	0.74	0.61
80	0.85	0.85	0.57
78	0.92	0.84	0.57
63	0.91	0.88	0.54
72	0.71	0.67	0.61
62	0.84	0.76	0.58
79	0.77	0.66	0.56
66	1.14	0.96	0.45
74	0.75	0.66	0.56
81	0.96	0.89	0.44
73	0.66	0.54	0.56
Mean	0.98	1.00	
SD	0.42	0.57	

Table 4 Morals Item Fit Statistics

Table 5 shows the result for morals item analysis. According to the Rasch Model, all items are suit with the measurement model; measuring the construct. No problematic item detected and all items have a good value of PTMEA Corr. Hence, no item that formulated the construct of morals to be dropped from the instrument.

C. Reliability and Separation Index of Items

Reliability is not a test to prove the quality of the data but it is a test to prove the quality of reproducibility of relative measure location. High reliability (of persons or items) means that there is a high probability that persons (or items) estimated with high measures or in other word the higher the reliability value, the sample should be bigger and/or less measurement error. Separation index is an index to classify the person abilities (person separation index) or item difficulty (item separation index). The instrument is considered good if the separation index for person and item are more than 2 because the person ability and item difficulty are varied [18]. Table 7 shows the value of reliability and separation index of item according to the instrument constructs. Reliability value for the construct of faith was 0.93 while for knowledge was 0.98. In the meantime, worship and morals constructs had reliability value of 0.99 respectively. The separation index for faith items was 3.56; knowledge items were 6.54; worship items was 8.11 and morals items was 8.40 respectively.

Construct	Reliability	Separation Index
Faith	0.93	3.56
Knowledge	0.98	6.54
Worship	0.99	8.11
Morals	0.99	8.40

Table 6. Reliability and Separation Index of Items

Results show that items for each construct have a very good value of item reliability [16, 18]. These values indicate that every item were reliable in measuring the respondents' spiritual well-being according to the constructs respectively. The reliability values range from 0.93 to 0.99 which are close to 1 as suggested by the model. Thus, all 71 items in IKSIR are reliable. In addition, the separation values means that the items are varied in difficulty [18]. The result shows separation indexes are range from 3.56 to 8.40 respectively; indicates that three to eight groups of item exist in this test. Therefore, the instrument has good distribution of item difficulty.

#### IV. CONCLUSION

Analyzing instrument by every single item is the best method to determine the quality of a new or developed instrument. Using the Rasch Measurement Model, the developer gets a lot of information about the items which is sufficient to improve the quality of the instrument. A quality instrument is an instrument that only measure what it is supposed to measure without redundancy or other factor that lead to misunderstanding to the respondents. As a conclusion, IKSIR is a new developed instrument to measure the level of spiritual well-being from Islamic perspective for adolescent. IKSIR shows the unidimension property which was formulated by four constructs that were faith, knowledge, worship and morals. From 71 items, 5 items were suggested by Rasch measurement model to be omitted from the instrument so that the items have a good internal consistency which leads to a valid and reliable instrument to be used in the future. Thus the final IKSIR to be gazette consists of 66 items measuring four constructs.

#### REFERENCES

- D.B Bekelman, S.M. Dy, D.M. Becker, I.S. Wittstein, D.E. Hendricks, T.E. Yamashita, and S.H. Gottlieb, "Spiritual well-being and depression in patients with heart failure," *Society of General Internal Medicine*, vol. 22, 2007, pp. 470–477.
- G.H. Boswell, E. Kahana, and P.D. Anderson, "Spirituality and healthy lifestyle behaviors: stress counter-balancing effects on the well-being of older adults," *Journal of Religions and Health*, vol. 45 issue 4, 2006, pp. 587–602.
- S. Cotton, E. Larkin, A. Hoopes, B.A. Cromer, and S.L. Rosenthal, "The impact of adolescent spirituality on depressive symptoms and health risk behaviors," *Journal of Adolescent Health*, vol. 36, 2005, pp. 529–529.
- D.P. Ashmos and D. Duchon, "Spiritualitu at workplace: a conceptualization and measure," *Journal of Management Inquiry*, vol. 9, 2000, pp. 134–145.
- Sarooj Noor and Sajjad Arif (2011), "Achieving job satisfaction via workplace spirituality: Pakistani doctors in focus," *European Journal of Social Sciences*, vol. 19 issue 4, 2011, pp. 507–515.
- J. Field, "Well-being and Happiness: IFLL Thematic Paper 4", 2009, retrieves from <http://www.niace.org.uk/life-long learning inquiry/docs/IFLLwellbeing.pdf>
- T.P. Daaleman and B.B. Frey, "The spirituality index of well-being: a new instrument for health-related quality-of-life research," *Annals of Family Medicine*, vol. 2 issue 5, 2004, pp. 499–502.
- Khalim Zainal and Wan Zulkifli Wan Hassan, "Pendekatan Islam dalam menangani masalah disiplin tegar dalam kalangan pelajar sekolah: satu kajian kes," *Journal of Islamic and Arabic Education*, vol. 1 issue 2, 2009, pp. 1–4.
- Fariza Md. Sham, "Tekanan emosi remaja Islam," *Islamiyyat*, vol. 27 issue 1, 2005, pp. 3-23.
- L.G. Underwood and J.A. Teresi, "The daily spiritual experience scale: development, theoretical, description, reliability, exploratory factor analysis and preliminary construct validity using health-related data," *The Society of Behaviour Medicine*, vol. 24 issue 1, 2002, pp. 22–33.
- Khurshid Ahmad, *Islam: Prinsip-prinsip dasar dan karekteristik-karekteristiknya in Pesan Islam*, Bandung: Penerbit Pustaka, 1983, pp. 12-36.
- Syaikh Muhammad Said Mursi, "Seni Mendidik Anak 2," Jakarta:Pustaka Al-Kautsar, 2001.
- Bambang Sumintono and Wahyu Widhiarso, "Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial", Cimahi: Trim Komunikata Publishing House, 2014.
- M.N. Rashidi, R.A. Begum, M. Mokhtar and J.J. Pereira, "Pelaksanaan analisis data menggunakan model pengukuran Rasch bagi menentukan wajaran item," *Journal of Advanced Research Design*, vol. 2 issue 1, 2014, pp. 1–9.
- N.J. Salkind, *Tests and Measurement for People Who (Think They) Hate Tests and Measurement*, Thousand Oaks: Sage Publications, Inc., 2006.
- Azrilah Abdul Aziz, *Asas Model Rasch: Pembentukan Skala dan Struktur Pengukuran*, Bangi: Penerbit UKM, 2012.
- Azrilah Abdul Aziz, Azlinah Mohamed, Noor Habibah Arshad, Sohaimi Zakaria and Mohd Saidfudin Mosadi, "Appraisal of course learning outcomes using Rasch measurement: a case study in information technology education", *International Journal of Systems Applications, Engineering and Development*, vol. 1, 2007, pp. 164–172.
- T.G. Bond, and C.M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Science*, 2<sup>nd</sup> ed., New Jersey: Lawrence Erlbaum Associates Publishers.
- Mohd. Kashfi Mohd Jailani, *Manual pengenalan pengukuran Rasch dan winsteps*, unpublished.
- R.J. Cohen, and M.E. Swerdilk, *Psychological Testing and Assessment: An Introduction to Test and Measurement*, 7<sup>th</sup> ed., New York: McGraw-Hill Higher Education, 2009.