

Developing Peer-Based Behavioural Assessment Instrument For Secondary School Students

Ajmain @ Jimaain Safar¹, Wan Hassan Wan Embong², Ahmad Marzuki Mohamad³

Islamic Civilization Faculty, Universiti Teknologi Malaysia¹²³

ajmain@utm.my¹

ABSTRACT

The purpose of this study is to develop a peer-based Behavioural Assessment Instrument (BAI) for secondary school students. The constructs of the instrument was developed based on Muhammad Abdullah Darraz's behavioural classification and the syllabus on Islamic behaviour. The instrument was also developed through literature reviews and interviews with four content experts and six practicing specialists. Seven constructs were established including behaviour towards Allah S.W.T., the Messenger s.a.w., the self, the family, the community, the environment and the nation. A total of 243 items were created based on the Instrument Construction Table. 318 secondary school students were selected to be involved in this study. The Nvivo version 7.0 and Winsteps version 3.68.2 software were used to analyse the data. Items in the BAI constructs were examined using Rasch measurement model. Based on the Rasch measurement model analysis, 69 items were removed. The results showed a high level of reliability index of 0.98 and respondents' reliability index of 0.98. Assessment of the items suggested that 174 out of 243 items were suitable for peer group assessment. The item and respondent separation indexes were found to be at an acceptable level. All the items fit with the Rasch measurement model based on the test of PTMEA CORR. Findings of the study showed that instrument was unidimensional. The GDIF analysis also indicated that all items of the instrument were not significant and suitable to be used. Hence, the results of the study suggested that 174 items could be used in the Behavioural Assessment Instrument (BAI).

INTRODUCTION

The main agenda of education in Malaysia aims to develop students who possess noble character, which covers the essence of human values. The emphasis is given not only to sharing knowledge, enhancing intellect and skills but also emphasising on building character. Modest and brilliant men can be produced through a comprehensive education system, integrated and worthy. Therefore, in order to create balanced and harmonious men, the quality of the educational process must be measured and assessed. Nik Aziz (1994) recommends comprehensive assessment should be done on a continuous basis on the student. To that end, Hassan Langu-lung (1995) has put emphasis on that the best approach to assess the level of the internalization of Islamic education, which should be done through an evaluation of a person about his conduct in the performance of Islamic teachings.

RESEARCH BACKGROUND

The study of Islamic education in primary and secondary school in Malaysia aims to produce students who practice virtuous and noble values as the pillar of dignified national culture (Ministry of Education 2002). To internalize the study of Islamic Education, it must be understood and practiced in daily life that will ultimately shape the attitudes and manners to become virtuous believers (Ahmad Mohd Salleh 2008).

To ensure the internalization and development of students' attitude and moral behaviour and so on have an impact on the curriculum through, an assessment instrument can be used as a tool to measure it. Through the review of the literatures, there are still a limited number of instruments developed to assess non-cognitive aspects in the field of Islamic education especially in the aspects of affective and psychomotor in Malaysia. The available instruments are found to consist of constructs that have a limited coverage of measuring the aspects of manners and behaviour in Islam. The self-assessment instrument, which was developed by Azhar (2006), Azma (2006) and Azimi *et al.* (2006) are more to the internalization of religious values, which includes a few constructs of moral behaviour. The IPAPSM instrument developed by Azhar (2006) has only involved three constructs of moral behaviour, which are personal, social and religious moral constructs. The SPPIP-M instrument developed by Azma (2006) also measures three constructs of the internalization of moral behaviour, which are divided into three constructs; *tasawur*, feelings and attitudes. Meanwhile, the MRPI instrument developed by Azimi *et al.* (2006) measures the internalization of religion according to the constructs of personal behaviour, social behaviour and religious ceremonies. SPMI instruments produced by the SRI-SMIH Education Center (2006), involves peer assessment, and only three constructs are covered in the assessment, which are of family, trust and courtesy. In the meantime, the instruments developed by the Ministry of Education Examination Board (2005) measure the constructs of everyday manners, social manners, customs, worship, manners towards parents and families as well as studying manners. However, only 16 of the 37 sub-field manners and morals at the secondary school level were assessed.

There are a number constructs which are left to be included in the assessment, for example the behaviour towards Allah S.W.T., the Messenger s.a.w., the Environment and the country. Accordingly, this study aims to develop a more comprehensive Peer-Based Behaviour Assessment Instrument, which covers the behaviour towards Allah S.W.T., the Messenger s.a.w., the Self, the Family, the Community, the Environment, and the Nation. This instrument hopefully is valid and reliable, and can be used to measure the level of students' behaviour.

RESEARCH OBJECTIVES

This research aims to develop an assessment instrument, which is valid and reliable, and can be used to measure the level of students' behaviour.

METHODOLOGY

This study has adopted a mixed-method research design. The qualitative approach was used during the development of the constructs and the sub-constructs, which involved interviews of 10 respondents; those who are expert teachers in the field of Islamic Education as well as the content experts of behaviour education, while the quantitative approach was used in the process of determining the validity and the reliability of the items consist in the instrumen. The study involved 318 students, who were randomly selected from two secondary schools in the state of Johore.

The finding collected through interviews were analysed using Nvivo version 7.0 to produce the categories of the constructs and the subconstructs. Meanwhile, the validity and the reliability of the instrument were analysed using Winstep 3.62.4 software (Linacre, 2007) and the validity of the instrument were also confirmed through the Rasch Model.

FINDING OF THE ANALYSIS AND DISCUSSION

Finding of the Qualitative Method of Research

A review on the literature was made in advance to develop the conceptual research framework, followed by collection of the data through panel of experts. Through the literature review, six constructs were developed, which are behaviour towards Allah S.W.T. and the Messenger s.a.w., the self, the family, the community, the environment, and the nation. Meanwhile, the first construct of the behaviour towards Allah S.W.T. and the Messenger s.a.w. should be split into two constructs according to the result from the panel of experts.

Next, 34 subconstructs of the 7 constructs were developed based on the literature review and the interview of the panel of experts. From the developed subconstructs, 275 items were listed representing the existing subconstructs. To ensure that the items were constructed to represent the constructs and the subconstructs, as well as the content validity, the items were reviewed by five experts. Each item must reach the consensus of more than three experts, and if not, it will be removed from the list of the instrument. The value of the coefficient of agreement among the examiners for all items reviewed and agreed by those five experts were subjected to the Cohen's Kappa index of reliability for the constructs and sub-constructs. Overall, the Cohen's Kappa index of reliability is very high between 0.90 to 0.95 (Staruss & Corbin, 1990). From the review of the panel of experts, 32 items were removed, while 25 items were revised. Only 243 items that had been agreed upon, and those items are; 68 items of behaviour towards Allah S.W.T., 8 items of behaviour towards the Messenger s.a.w., 72 items of behaviour towards the Self, 13 items of the behaviour towards the Family, 44 items of behaviour towards the Community, 20 items of behaviour towards the Environment, and 18 items of behaviour towards the Nation.

Finding of the Quantitative Method of Research

The study was conducted on 318 respondents from 16-year-old school student. Rasch measurement model was used to analyse the reliability of the instrument through the analysis of: (1) the index of the items and the respondent, (2) the index of separation, (3) the polarity of the items, (4) fit items, (5) dimensionality and, (6) *differential item functioning* (DIF) by gender (Rodiah, 2010).

Table 1 shows the sample of the research which consists of 318 respondents. Of the total number of respondents, 140 (44%) respondents were male and 178 (56%) respondents are female. Respondents were determined by the teacher to become the appraiser of a friend they have known and mixed around for at least six months. Every individual must be appraised by two friends of their age. Therefore, a total number of 70 male and 89 female were appraised in this study.

Table 1 Profile of the Appraiser and the Appraised Respondent

	APPRAISER		APPRAISED	
	Frequency	Percentage	Frequency	Percentage
Gender				
Male	140	44%	70	44%
Female	178	56%	89	56%
TOTAL	318	100%	159	100%

Validating Items of the Instrumen

This section reports on the analysis of the items designed based on the five-point Likert scale, which are; (5) very often (perform consistently), (4) often (perform consistently and leave occasionally), (3) quite often (perform sometime), (2) not often (perform once or twice), (1) not very often (never perform). The instrument consists of seven constructs with 243 items.

a) Item Reliability

Table 2 demonstrates the summary of the reliability of the items for seven constructs of the assessment instrument consists of 243 items. All seven constructs show high reliability index from 0.89 to 0.99.

Table 2 Reliability of the 7 Constructs

No.	Constructs	Reliability		INFIT MNSQ		OUTFIT MNSQ	
		Item	Separation	Max	Min	Max	Min
1.	BTA	0.99	9.11	1.67	0.82	1.78	0.81
2.	BTM	0.98	7.04	1.51	0.98	1.66	0.98
3.	BTS	0.97	6.15	1.39	0.73	1.76	0.71
4.	BTF	0.89	2.88	0.96	0.78	0.96	0.71
5.	BTC	0.93	3.78	1.51	0.69	1.65	0.66
6.	BTE	0.97	5.42	1.30	0.75	1.33	0.72
7.	BTN	0.94	4.00	1.25	0.84	1.35	0.78

Based on Table 2, the reliability of the items shows excellent result as all items have value near to 1.0. In fact, according to Wright dan Masters (1982), the expected repetition of the item reliability for the seven constructs is very high when it is administered to a group of respondents who have similar characteristics. However, the reliability of the items can be increased if the misfit items are given due attention.

Table 2 also shows the four constructs of BTF (Behaviour towards the Family), BTN (Behaviour towards the Nation), BTE (Behaviour towards the Environment) and BTS (Behaviour towards the Self) which achieve maximum value below 1.4 logit MNSQ *Outfit*. The BTF construct of 0.96 logit, BTN construct of 1.25 logit, BTE construct of 1.30 logit, and BTS construct of 1.39 logit, are fit to Rasch model. While the constructs of BTA (Behaviour towards Allah S.W.T.), BTM (Behaviour towards the Messenger s.a.w.) and BTC (Behaviour towards the Community) should be given special attention because there are several misfit items based on the *Outfit and Infit* MNSQ as it shows that there is at least one misfit item identified. Those constructs that should be given special attention are; Behaviour towards Allah S.W.T. (*Infit* MNSQ 1.67 logit / *Outfit* MNSQ 1.78 logit); Behaviour towards the Messenger s.a.w. (*Infit* MNSQ 1.51 logit / *Outfit* MNSQ 1.66 logit); and Behaviour towards the Community (*Infit* MNSQ 1.51 logit / *Outfit* MNSQ 1.65 logit). The analysis of the misfit order must be further carried out to identify the items that do not fit with the Rasch measurement model.

Table 2 also shows the index of separation of the constructs, which range from 2.88 to 9.11. According to Linacre (2006); Bond and Fox (2007), the acceptable value must be higher than 2.0. The highest separation index are those constructs of Behaviour towards Allah S.W.T. with 9 levels of separation, followed by Behaviour towards the Messenger s.a.w. with 7 levels of separation, Behaviour towards the Self with 6 levels of separation, Behaviour towards the Environment with 5 levels of separation, Behaviour towards the Community and Nation with the

same 4 levels of separation, and finally, Behaviour towards the Family with 3 levels of separation. These statistics indicate that the items are 3 to 9 times more dispersed from the square root of the error. The item separation index is divided into 3 to 9 strata or levels of agreement. This situation also shows that these items are 3 to 9 times more dispersed from the square root of error. The separation index will increase if the reliability is increased and the misfit items are detected and removed from the analysis.

b) Identifying the Polarity of the Items of the Constructs

Further analysis was made to identify the polarity of the items. Table 3 shows that 243 items of the 7 constructs demonstrates positive value of point measure correlation (PTMEA CORR) for every construct. The most minimum PTMEA CORR is 0.27 for item K14 in the construct of Behaviour towards Allah S.W.T., while the most maximum PTMEA CORR is 0.71 for item K502 in the construct of Behaviour towards the Community. According to Linacre (2002), acceptable value of point measure correlation (PTMEA CORR) ranges between 0.20 to 0.79. Any negative values and any value below 0.20 should be removed as it does not measure any construct. The analysis of the item polarity shows that all constructs have statistically demonstrated that the items move parallel in one direction to measure the constructs that are intended to be measured.

Table 3 Item Polarity of the 7 Constructs

No.	Constructs	PTMEA		CORR		Total Item
		Min	Item	Max	Item	
1.	BTA	0.27	K14	0.66	K154	68
2.	BTM	0.35	K207	0.57	K206	8
3.	BTS	0.38	K330	0.71	K352	72
4.	BTF	0.51	K412	0.64	K402	13
5.	BTC	0.33	K504	0.72	K502	44
6.	BTE	0.46	K613	0.68	K608	20
7.	BTN	0.43	K715	0.62	K702	18
TOTAL						243

c) Identifying the Item Fit of the Constructs

To identify the fit of the items, the analysis of item fit for every construct must be done. Table 4 shows the statistical finding for the items of the 7 constructs of behaviour. Outfit is a t standardized outlier-sensitive towards the mean square fit statistic, more sensitive to the attitude that is unexpected by the respondents. Linacre (2007) has outlined four groups based on the value that is fit for measuring items in a construct, which are (1) value > 2.0 means the unnecessary variables are more that the useful ones, reduces the accuracy of the measurement; (2) value > 1.5 means does not meet the measurement variable, however does not reduce the accuracy of the measurement; (3) value 0.5 – 1.5 means productive or good measurement; (4) value < 0.5 means very easy to be predicted and less productive in the measurement, producing confusing measurement towards the expected and reality.

According to Bond dan Fox (2007), the acceptable range for grading scale or Likert scale is between 0.6 logit to 1.4 logit. An item values more than 1.4 indicates that the item is non-homogen with other items within a construct, while item values less than 0.6 indicates the redundancy among the items.

Table 4 Misfit Items of the 7 Constructs

Constructs	Measure	Model	INFIT		OUTFIT		PTMEA CORR	ITEM
			S.E	MNSQ	ZSTD	MNSQ		
BTA	2.23	0.07	1.41	4.2	1.49	5.0	0.29	K15
BTA	1.90	0.06	1.43	5.2	1.53	5.9	0.37	K112
BTA	1.77	0.06	1.49	5.2	1.71	7.0	0.31	K122
BTA	0.94	0.06	1.59	7.0	1.65	7.2	0.34	K137
BTA	0.83	0.06	1.67	7.9	1.78	8.4	0.27	K114
BTM	0.79	0.06	1.51	6.1	1.66	7.3	0.39	K208
BTC	0.89	0.06	1.51	6.2	1.65	7.5	0.33	K504

Table 4 shows the size of the misfit items with the Rasch measurement model for all 7 constructs. A total of 7 items are identified as misfit for being outside from the fixed logit value between 0.6 to 1.4. These include item K15, K112, K122, K137, K14 from the BTA construct; item K208 from the BTM construct, and item K504 from the BTC construct. Those seven items were proposed to be separated from the instrument.

d) Identifying Items Measuring Single or Unidimensional Construct

Principal Components Analysis (PCA) was conducted to determine the items in the instruments that only measure a single construct. To ensure that items are produced only to measure a single construct, Linacre (2005) suggests, the variance explained by measures preferably >40%. While the unexplained variance in 1st contrast <3.0 is good, and <5% is well accepted. This shows that the second dimension does not clearly existed.

Table 5. Unidimensional: Standardized Residual Variance of the 7 Constructs

No.	Constructs	<i>variance explained by Measures (%)</i>	<i>unexplained variance in 1st contrast</i>	
1.	BTA	56.1	6.9	4.4%
2.	BTM	56.1	3.0	16.2%
3.	BTS	56.1	5.5	3.3%
4.	BTF	56.1	4.3	14.4%
5.	BTC	56.1	6.7	6.7%
6.	BTE	56.1	4.0	8.8%
7.	BTN	56.1	3.6	8.9%
	OVERALL	56.1	14.1	2.5%

The findings based on the PCA (principal component analysis) as shown in Table 5 found that variance explained by measures (%) indicates that all constructs have a value of 56.1% above the value of a good variance which is 40%. In total, the unexplained variance in 1st contrast shows there is no constructs that is <3. This shows the possibility of the existence of one or more misfit items that tend to form the second dimension. Although the total value of the unexplained variance in 1st contrast is <2.5%, which is generally a well-received value. Table 5 is an index reading before the misfit items or items that do not fit the Rasch measurement model are removed.

e) Identifying Differential Item Functioning (DIF) by Gender

To identify the presence of differential item functioning, Winstep software has run two-tailed t-test to test the significant differences between the two difficulty indexes. According to Bond and Fox (2007), significant DIF is based on three indicators of the value of t, the DIF contrast and p-value. The measure of the t-value is at ± 2.0 or $t \geq + 2.0 \leq -2.0$, DIF index contrast is at the value of ± 0.5 or $\geq +0.5 \leq -0.5$ and the significant p-value is $p < 0.05$. Lai and Eton (2002) suggest, the index DIF contrast is important for the Likert scale. While Tennant and Pallant (2007) report that GDIF size less than $+0.5$ logits are not significant (negligible DIF). Negative index of GDIF means an item is easier confirmed by male or female, while positive index means the item is easier confirmed by male or female.

Table 6 Differential Item Functioning Analysis by Gender

Gender	DIF Measure	DIF S.E	Gender	DIF Measure	DIF S.E	DIF Contrast	Joint S.E	t	p	Item	Construct
1	-0.39	0.10	2	-0.91	0.09	-0.53	0.13	-2.56	000.	K13	BTA
1	1.88	0.11	2	2.45	0.10	-0.57	0.15	-3.93	000.	K15	BTA
1	0.12	0.10	2	0.90	0.08	-0.78	0.13	-6.10	000.	K16	BTA
1	2.49	0.12	2	3.01	0.11	-0.52	0.17	-3.16	000.	K17	BTA
1	-0.38	0.09	2	-0.91	0.13	0.54	0.16	3.35	000.	K18	BTA
1	0.38	0.08	2	-0.40	0.10	0.79	0.13	6.07	000.	K111	BTA
1	0.47	0.09	2	1.07	0.08	-0.59	0.12	-4.87	0.02	K112	BTA
1	0.42	0.09	2	1.02	0.08	-0.54	0.12	-2.06	0.04	K114	BTA
1	-0.07	0.09	2	0.44	0.08	-0.51	0.12	-3.81	0.02	K115	BTA
1	0.44	0.09	2	1.06	0.08	-0.55	0.12	-2.04	0.04	K117	BTA
1	-0.18	0.10	2	0.42	0.09	-0.60	0.13	-4.50	0.00	K118	BTA
1	-0.30	0.10	2	0.26	0.09	-0.56	0.13	-4.22	0.00	K119	BTA
1	-0.44	0.09	2	0.22	0.08	-0.66	0.12	-2.93	0.00	K120	BTA
1	1.27	0.09	2	2.13	0.09	-0.86	0.13	-6.65	0.00	K122	BTA
1	-1.10	0.11	2	-0.56	0.10	-0.54	0.15	-2.93	0.00	K124	BTA
1	-0.73	0.11	2	-0.17	0.10	-0.56	0.14	-2.51	0.01	K125	BTA
1	-0.36	0.10	2	-0.91	0.09	-0.53	0.13	-2.50	0.01	K127	BTA
1	1.49	0.12	2	3.01	0.11	-0.52	0.17	-3.11	0.02	K129	BTA
1	0.68	0.08	2	-0.40	0.10	0.61	0.13	4.07	0.00	K131	BTA
1	0.10	0.10	2	0.61	0.08	-0.51	0.13	-3.61	0.00	K133	BTA
1	0.27	0.09	2	1.07	0.08	-0.53	0.12	-4.17	0.02	K136	BTA
1	0.08	0.10	2	0.59	0.09	-0.51	0.14	-3.48	0.00	K139	BTA
1	2.19	0.12	2	2.01	0.11	-0.52	0.17	-3.01	0.00	K140	BTA
1	0.42	0.09	2	1.02	0.08	-0.54	0.12	-2.06	0.02	K143	ATA
1	0.10	0.10	2	0.51	0.08	-0.51	0.13	-3.76	0.00	K146	ATA
1	0.32	0.09	2	1.02	0.08	-0.52	0.12	-2.06	0.01	K157	BTA
1	0.24	0.08	2	-0.44	0.10	0.64	0.13	4.50	0.04	K159	BTA
1	-0.11	0.09	2	0.41	0.08	-0.51	0.12	-3.21	0.03	K161	BTA
1	-0.45	0.10	2	-0.96	0.12	0.51	0.16	3.04	0.00	K164	BTA
1	-0.16	0.09	2	0.41	0.08	-0.54	0.12	-4.20	0.00	K165	BTA
1	2.11	0.12	2	2.01	0.11	-0.51	0.17	-3.12	0.01	K206	BTM
1	0.10	0.10	2	0.61	0.08	-0.51	0.13	-3.28	0.00	K207	BTM
1	-0.07	0.09	2	0.54	0.08	-0.51	0.12	-3.21	0.02	K208	BTM
1	-0.34	0.10	2	-0.91	0.09	-0.53	0.13	-2.51	0.00	K303	BTS

1	0.10	0.10	2	0.31	0.08	-0.51	0.13	-3.88	0.00	K305 BTS
1	-0.10	0.10	2	0.77	0.08	-0.87	0.13	-6.74	0.00	K308 BTS
1	0.08	0.10	2	0.50	0.09	-0.51	0.14	-3.11	0.00	K319 BTS
1	0.37	0.09	2	1.07	0.08	-0.51	0.12	-3.87	0.02	K323 BTS
1	0.20	0.08	2	-0.41	0.10	0.61	0.13	4.10	0.04	K329 BTS
1	0.10	0.10	2	0.17	0.08	-0.51	0.13	-3.01	0.03	K330 BTS
1	-0.45	0.10	2	-0.93	0.12	0.51	0.16	3.01	0.01	K332 BTS
1	-0.16	0.09	2	0.49	0.08	-0.64	0.12	-5.20	0.00	K333 BTS
1	0.38	0.08	2	-0.40	0.10	0.54	0.13	3.07	0.03	K340 BTS
1	0.15	0.10	2	0.67	0.09	-0.52	0.14	-3.79	0.00	K342 BTS
1	0.25	0.08	2	-0.44	0.10	0.64	0.13	4.60	0.01	K347 BTS
1	0.12	0.10	2	0.80	0.08	-0.68	0.13	-4.10	0.00	K355 BTS
1	-0.19	0.10	2	-0.76	0.11	0.56	0.14	3.20	0.00	K362 BTS
1	-0.42	0.10	2	-0.91	0.12	0.51	0.16	3.01	0.00	K366 BTS
1	0.32	0.09	2	1.02	0.08	-0.51	0.12	-2.06	0.00	K369 BTS
1	0.10	0.10	2	0.71	0.08	-0.51	0.13	-2.88	0.00	K371 BTS
1	0.48	0.08	2	-0.40	0.10	0.52	0.13	3.05	0.02	K403 BTF
1	-0.07	0.09	2	0.44	0.08	-0.51	0.12	-3.11	0.01	K406 BTF
1	-0.50	0.10	2	-1.04	0.12	0.54	0.16	3.31	0.00	K407 BTF
1	-0.03	0.10	2	0.49	0.09	-0.51	0.13	-3.81	0.00	K503 BTC
1	-0.31	0.09	2	-0.51	0.13	0.52	0.16	3.01	0.00	K507 BTC
1	0.05	0.10	2	-0.46	0.10	0.51	0.14	3.27	0.00	K509 BTC
1	-0.07	0.09	2	0.44	0.08	-0.51	0.12	-3.31	0.02	K511 BTC
1	0.37	0.09	2	0.77	0.08	-0.53	0.12	-3.87	0.02	K517 BTC
1	0.24	0.09	2	-0.27	0.09	0.51	0.13	3.61	0.00	K530 BTC
1	0.08	0.10	2	-0.44	0.10	0.52	0.14	3.29	0.00	K531 BTC
1	0.48	0.08	2	-0.40	0.10	0.53	0.13	4.07	0.03	K532 BTC
1	-0.38	0.09	2	-0.71	0.13	0.54	0.16	3.31	0.01	K535 BTC
1	0.20	0.08	2	-0.44	0.10	0.64	0.13	4.80	0.00	K601 BTE
1	-0.30	0.10	2	-0.91	0.09	-0.52	0.13	-2.52	0.00	K603 BTE
1	2.19	0.12	2	2.01	0.11	-0.52	0.17	-2.16	0.04	K606 BTE
1	0.10	0.10	2	0.61	0.08	-0.51	0.13	-3.44	0.00	K711 BTN
1	0.12	0.10	2	0.80	0.08	-0.68	0.13	-5.10	0.03	K716 BTN

Table 6 shows the summary of the result of the GDIF analysis over 243 items to measure the achievement of students' behaviour. Respondents of group 1 are males, while 2 are females. The result shows that a total of 67 items are GDIF significant at $t \pm 2.0$ or $t \geq + 2.0 \leq -2.0$, with DIF contrast value $\geq +0.5 \leq -0.5$ and $p < 0.05$ with 30 items from the BTA construct, 3 items from the BTM construct, 17 items from the BTS construct, 3 items from the BTF construct, 9 items from the BTC construct, 3 items from the BTE construct, and 2 items from the BTN construct. The result also shows 32 items are simply agreed by male peers and 33 items are simply agreed by female peers. There are 35 items which are hardly agreed by their male peers and 34 items are hardly agreed by female peers. Therefore these items should be removed in the next analysis.

Summary of Items Inspection

The findings of the inspection on 243 items based on analysis using Rasch measurement model found 7 misfit items and 62 DIF items based on gender. Thus a total of 69 items are suggested to be removed from the instrument.

Table 7 Summary of Item Functioning Inspection Analysis

NO	CONSTRUCT	ITEMS SUGGESTED TO BE REMOVED				RASCH REMOVE	RASCH REMAIN
		TOTAL	PTMEA	INFIT	GENDER		
1	BTA	68		K15 K112 K122 K137 K114	K13, K15, K16, K17, K18, K111, K112, K114, K115, K117, K118, K119, K120, K122, K124, K125, K127, K129, K131, K133, K136, K139, K140, K143, K146, K157, K159, K161,	31	37
2	BTM	8		K208	K206, K207, K208	3	5
3	BTS	72			K303, K305, K308, K319, K323, K329, K330, K332, K333, K340, K342, K347, K355, K362, K366,	17	55
4	BTF	13			K403, K406, K407	3	10
5	BTC	44		K504	K503, K507, K509, K511, K517, K530,	10	34
6	BTE	20			K601, K603, K606	3	17
7	BTN	18			K711, K716	2	16
	TOTAL	243		7	62	69	174

Table 7 shows the number of items of the constructs after the analysis. A total of 69 items are proposed to be removed. Another 174 items are remained in the instrument. Of these, the construct of Behaviour towards Allah S.W.T. remains 37 items, the construct of Behaviour towards the Messenger s.a.w. remains 5 items, the construct of Behaviour towards the Self remains 55 items, the construct of Behaviour towards Family remains 10 items, the construct of Behaviour towards the Community remains 34 items, the construct of Behaviour towards the Environment remains 17 items, and the construct of Behaviour towards the Nation remains 16 items.

Reliability Value of the Research Instrument

The reliability index of 174 items of the instrument measuring students' behaviour is 0.98, as shown in Table 8. This indicates the reliability of the items is high with the total items are sufficient enough to measure students' behaviour. It also means that the items are stable and consistent when the instrument is repeatedly used to measure students' behaviour of the same or almost the same characteristic (Wright & Masters, 1982). High Cronbach alpha value of 0.99 also indicates that the instrument is highly reliable to measure students' behaviour. According to Othman (2001) and Palant (2001), the value of 0.70 and above is a good value and can be accepted.

Table 8: Items Reliability of the Instrument

INPUT: 318 PERSONS 174 ITEMS MEASURED: 318 PERSONS 174 ITEMS 1207
CATS 3.64.2

SUMMARY OF 174 MEASURED ITEMS								
	RAW SCORE	COUN T	MEAS- URE	MOD- EL ER- ROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	1217.8	318.0	.00	.07	1.00	-.1	1.01	.0
S.D	134.5	.0	.51	.01	.19	2.3	.22	2.5
MAX	1491.0	318.0	2.81	.09	1.67	7.9	1.78	8.4
MIN	559.0	318.0	-1.06	.06	.69	-4.5	.66	-4.7
REAL RMSE	.07	ADJ.SD	.51	SEPARATION	7.08	ITEM	RELIABILITY	.98
MODEL RMSE	.07	ADJ.SD	.51	SEPARATION	7.33	ITEM	RELIABILITY	.98
S.E. OF ITEM MEAN = .03								

UMEAN=.000 USCALE=1.000

ITEM RAW SCORE-TO-MEASURE CORRELATION= -.96 (approximate due to missing data)

77273 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 172045.14

CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY= .99 (approximate due to missing data)

CONCLUSION

The development of secondary school students' behaviour assessment has achieved high validity and reliability which can be used by the students to measure behaviour among them. This instrument is suitable to be used at secondary schools in Malaysia, especially in the making of a holistic assessment system. It is comprised of 7 constructs of behaviour, which are; Behaviour towards Allah S.W.T., Behaviour towards the Messenger s.a.w., Behaviour towards the Self, Behaviour towards Family, Behaviour towards the Community, Behaviour towards the Environment, and Behaviour towards the Nation, that can be utilized by educational administration to identify the level of achievement of students' behaviour, and subsequently, through the pathway to implement the intervention programmes to increase students' achievement in behaviour. This matter fulfils the aim and philosophy of national education in the academic as well as the non-academic field, especially in the field of character building and behaviour development.

REFERENCES

- Ahmad Mohd Salleh. 2008. *Pengajian Agama Islam dan j-QAF Metodologi dan Pedagogi Pendidikan*. Shah Alam Selangor : Oxford Fajar Sdn. Bhd.
- Azhar Ahmad. 2006. *Strategi Pembelajaran Pengaturan Kendiri Pendidikan Islam dan Penghayatan Akhlak Pelajar Sekolah Menengah*. Tesis Ph.D Universiti Kebangsaan Malaysia.
- Azimi Hamzah, Steven Eric Krauss @ Abdul Lateef Abdullah, Sidek Mohd Noah, Turiman Suandi, Rumaya Juhari, Jamiah Manap, Khairul Anwar Mastor, Azma Mahmoud & Hasnan Kassin. 2007. *Muslim Religiosity Personality Assessment: Prototype for Nation*

- Building*. Serdang: Institute for Social Science Studies Universiti Putra Malaysia.
- Azimi Hamzah, Steven Eric Krauss @ Abdul Lateef Abdullah, Sidek Mohd Noah, Turiman Suandi, Rumaya Juhari, Jamiah Manap, Khairul Anwar Mastor, Azma Mahmood & Hasnan Kassan. 2006. *Muslim Religiosity and Personality Indexing: Implications for Nation Building*. Serdang: Institute for Social Science Studies Universiti Putra Malaysia.
- Azma Mahmood. 2006. Pengukuran Tahap Penghayatan Pendidikan Islam Pelajar-Pelajar Sekolah Menengah di Malaysia. Tesis Ph.D Universiti Kebangsaan Malaysia.
- Bond, T.G. & Fox, C.M. 2007. *Applying the Rasch model: Fundamental measurement in the human sciences*. Ed. Ke-2. New Jersey: Lawrence Erlbaum Associates Publishers London.
- Hassan Langgulang. 1995. *Pendidikan Islam dan peralihan paradigma*. Shah Alam: Hizbi.
- Nik Azis Nik Pa. 1994. *Penghayatan Wawasan Pembangunan Negara*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Kementerian Pendidikan Malaysia. 2002a. *Sukatan Pelajaran Pendidikan Islam KBSR*.
- Kementerian Pendidikan Malaysia. 2002b. *Sukatan Pelajaran Pendidikan Islam KBSM*
- Lai, J.S. & Eton, D.T. 2002. Clinically meaningful gaps. *Rasch Measurement Transactions* 15(4): 850.
- Linacre, J.M. 2002. *Differential item and test functioning*.
<http://www.rasch.org/rmt/rmt163g.htm>
- Linacre, J.M. 2005. *Test validity and Rasch measurement: construct, content*. Rasch measurement transactions. <http://www.rasch.org/rmt/rmt181h.htm>[2 September 2010]
- Linacre, J.M. 2006. *Winsteps Rasch measurement computer program*. Chicago: Winsteps.com.
- Linacre, J.M. 2007. *A user's guide to FACETS: Rasch-model computer programs*. Chicago: MESA Press.
- Othman Mohamed. 2001. *Penulisan Tesis Dalam Bidang Sains Sosial Terapan*. Serdang Selangor: Penerbit Universiti Putra Malaysia.
- Palant, J. 2001. *SPSS Survival Manual: A step by step guide to data analysis using SPSS for windows*. Ver. 10. Illinois: Allen & Unwin.
- Rodiah Idris. 2010. *Pemeriksaan Kefungsian Item dan Pembentukan Penanda Aras Pentaksiran Kemahiran Generik*. Tesis Ph.D Universiti Kebangsaan Malaysia Bangi.
- Strauss, A. & Corbin, J. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage Publications.
- Tennant, A & Pallant, J.F. 2007. DIF matters: A practical approach to test if Differential Item Functioning (DIF) makes a difference. *Rasch Measurement Transactions* 20:1082-1084.
- Wright, B.D. & Master, G.N. 1982. *Rating scale analysis*. Chicago: MESA Press.