

Differential Item Functioning of Teaching Practices in Islamic Education Instrument (TPIED)

*Sarimah Mokhtar
umimursyid@yahoo.com
Universiti Kebangsaan Malaysia*

*Mohd Kashfi Mohd Jailani
Kashfi2011@gmail.com.my
Universiti Kebangsaan Malaysia*

*Ab. Halim Tamuri
Tamuri67@gmail.com
Universiti Kebangsaan Malaysia*

*Mohd Aderi Che Noh
aderi@ukm.my
Universiti Kebangsaan Malaysia*

*Kamarulzaman Abdul Ghani
qamar68@ukm.my
Universiti Kebangsaan Malaysia*

ABSTRACT

The purpose of this paper is to detect Differential Item Functioning (DIF) in Teaching Practice in Islamic Education instrument (TPIED) based on gender and stream. The instrument was administered on 150 from form four students Malaysian secondary schools' from various groups within the Ministry of Education school in Selangor, Malaysia. The samples were clustered and randomly chosen. The TPIED used consists of 4 constructs which were (a) teaching skills, (b) personality traits, (c) Intelligence and (d) motivations and 95 items on a 5-point likert scale. Rasch model was used to examine both validity and reliability of the instrument, person's reliability and item's reliability is more robust compared to Cronbach's Alpha. It also allows item elimination based on t-value and differential measure by Winsteps version 3.68.2. The person's reliability value obtained was 0.95 while the item's reliability value was 0.93. The findings also detected seven item were distinguished as DIF based on gender and eight items were detected as DIF based stream. The final analysis has dropped 13 items and maintained 82 items that were legitimate and reliable to gauge 4 constructs in TPIED. This TPIED, free from DIF, could be used to measure the four constructs in the Teaching Practice in Islamic Education instrument (TPIED) for Malaysian Islamic education, from Malaysian secondary schools students' perceptions. Since the four constructs was recommended by al-Ghazali and Dunkin and Biddle.

Keywords: Teaching Practice in Islamic Education, Construct Validity, Gender Differential Item Functioning (GDIF), Differential Item Functioning (DIF).

INTRODUCTION

A Malaysian secondary school is the most suitable place to polish the teaching practices and teaching skills for Malaysian Islamic education teachers, from Malaysian secondary schools students' perceptions al-Ghazali, (t.t), Dunkin and Biddle (1974), Ab. Halim (2006), Aderi (2008), Misnan Jemali (2008), Kamarul Azmi Jasmi (2010). Teaching practice and teaching skills are very important in producing good students who will later develop to be good persons with good behavior, which is better known as Akhlaq, Miskawayh (1996), al-Ghazali (t.t), (2008), Nasr (2000), Ghazali (2001), Ab.Halim (2007), Azhar (2007), Kamarul Azmi Jasmi (2010). Malaysian secondary schools students should have the desire and ability to induce themselves in the process of creating human being with the first class manners of moral education by the teaching practices and teaching skills for Malaysian Islamic education teachers. Thus, the inculcation of teaching practice and teaching skills by the Malaysian Islamic education teachers should be carried out fully by the school which involves those who are dealing with the students.

One of the vital processes in schools' education system is the evaluation and quality assurance of every students and teaching practices and teaching skills by Malaysian Islamic education teachers, from Malaysian secondary schools students' perceptions that are produced. Though the process is rather difficult, the initiative to ensure the achievement of Teaching Practice in Islamic Education for Malaysian Islamic education teachers, from Malaysian secondary schools' student perceptions is crucial to evaluate if the process is in place or vice versa Dunkin & Biddle (1974), Ab. Halim (2006), Aderi (2008), Misnan Jemali (2008), Kamarul Azmi Jasmi (2010).

Competency evaluation on students' perceptions about teaching practices and teaching skills by Malaysian Islamic education teachers, from Malaysian secondary schools was debated during the formation of Malaysian Qualification Framework (MQF) in the year 2006. Competency evaluation's purpose is to gauge performance of the following teaching practices and skills by Malaysian Islamic education teachers from Malaysian secondary schools. It consists of four domains for teacher properties : (i) teaching skills (ii) personality traits (iii) Intelligence (iv) motivations al-Ghazali, (t.t), Dunkin and Biddle, (1974).

The development of *Teaching Practice in Islamic Education (TPIED)* instrument is used to gauge student's perceptions about teaching practices and teaching skills by Malaysian Islamic education teachers from Malaysian secondary schools. *(TPIED)* concentrates on ordinary element in the eye of stakeholders who have psychometric understanding and can be gauged without form as limited by the views to fulfill what is supposed to be evaluated. Ideally, the skills evaluation will be upgraded while they are at secondary schools.

Every item for evaluation should be pre-tested to see how compatible when it is used on students' perceptions about teaching practices and teaching skills by Malaysian Islamic education teachers, from Malaysian secondary schools. As stressed by Osterlind, (1989), the item analysis is a process to study the item critically with the aim of identifying and reducing the error of measurement. It is not sufficient to analyze the item which is based on expert judgement in ensuring the quality of the item. Thus, Rasch measurement Model through Differential Item Functioning (DIF) analysis was used to strengthen the authenticity of the item. According to Camilli (1993) DIF and its authenticity should be interrelated. Berk (1982) states that because the item is at the most basic criterion in content analysis and also the base for inference, DIF study is a necessity for study of analysis of item inclination. DIF is not really required because it does not show that the test does not measure the same capability of individuals in different groups Maller (2001).

OBJECTIVE

The aim of this study was to identify items to validate and examine the reliability of a self report instrument namely the *Teaching Practice in Islamic Education (TPIED)* for Malaysian Islamic education teachers, from Malaysian secondary schools student perceptions whether they function differently on students of different gender . Besides that, this paper would also like to identify items whether they function differently according to certain groups like stream.

METHODOLOGY

The study was conducted using a quantitative survey approach. One hundred and fifty students (72 male and 78 female students) participated in the study. The sample was clustered according to 5 categories of secondary school in Selangor' a state of Malaysia. Respondents' age range was from 16 years old. The data was analyzed using Winsteps version 3.68.2, a Rasch-based an analysis software. The study used *TPIED* which consisted of 95 items. The item consist of 43 teaching skills items, 31 personality traits items, 10 Intelligence items and 11 motivations items. GDIF and DIF were used in the data analysis.

RESULTS

This study was designed to provide answers to two questions: (1) Does the *TPIED* function differently on students perceptions about teaching practices or teaching skills by Malaysian Islamic education teachers from Malaysian secondary schools of different gender?; and (2) Does the *TPIED* function differently according to different groups like gender and stream? The data was analyzed using Winsteps version 3.68.2 to determine the validity and reliability of the *TPIED*. Rasch Model analysis provided item reliability and construct validity. The item reliability index range is 0 to 1 whereby 0.8 and above is strongly acceptable Bond & Fox (2001). A construct with a set of unidimensional items should display a positive *PTMEA Corr* value which means that the items are working together to measure a single underlying construct. This is the basic step in measuring the construct validity Bond & Fox (2001). Item reliability of four domains for teacher properties; teaching skills is 0.94, personality traits is 0.90, Intelligence is 0.95 and motivations is 0.81. Wright & Masters (1982) claim the value is positif because it is near to 1.0. Hence, *TPIED* item repetition prediction is also high if it is being administrated to other groups of respondents with similar capability (Wright & Masters 1982).

Table 1 shows the responses from each cluster: gender and stream. The number of female respondents is 78(52%) whereas male 72 (48%). The total number of stream is 2. Science and non-science students are the respondents in this study. The number of non science students is 78 (52%) whereas science students 72 (48%).

Table 1 Profile of Respondents

Demography Factor	N	Factor	Frequency	Percentage
Gender	150	Male (1)	72	48.0
		Female (2)	78	52.0
Streams	150	Science (1)	72	48.0
		Non-science (2)	78	52.0

Analysis was carried out to investigate the existence of GDIF in the questionnaire used. To analyze GDIF, Winsteps performs two-tailed t-test to test significance of the differences between two

difficulty indices. The confidence level was at 95% and critical t value was set at 2.0 for all DIF analyses. Besides that GDIF contrast index was used to show the gap difference of endorsement level of each item for both male and female when compared. According to Lai & Eton (2002) the value 0.5 logits of DIF Contrast is significant for Likert Scale. Wright and Panchalakesan in Tennant & Pallant (2007) however reported that GDIF with the size of less than 0.5 logits is not significant and negligible (negligible DIF). In this study negative index of GDIF contrast means item is easier to endorse by male whereas positive index means item is easier endorsed by female.

Table 2 displays results of GDIF analysis on 43 studied items. Analysis revealed that 1 item from the 43 items in teaching skills showed significant GDIF. It is item B4. The item carries the percentage of (2%) with significant GDIF of 0.58 t-value of more than 2.0 ($t \geq 2.0$). Therefore the item is suggested to be deleted.

Table 2 GDIF Analysis of Teaching Skills Construct

Group	DIF Measure (Difficulty measure)	Group	DIF Measure (Difficulty measure)	GDIF Contrast (DIF size)	t	Df	Item Label
1	0.04	2	-0.54	0.58	2.67	147	B4

Figure 1 shows the DIF plot using DIF measure on the analysis of teaching skills construct by gender

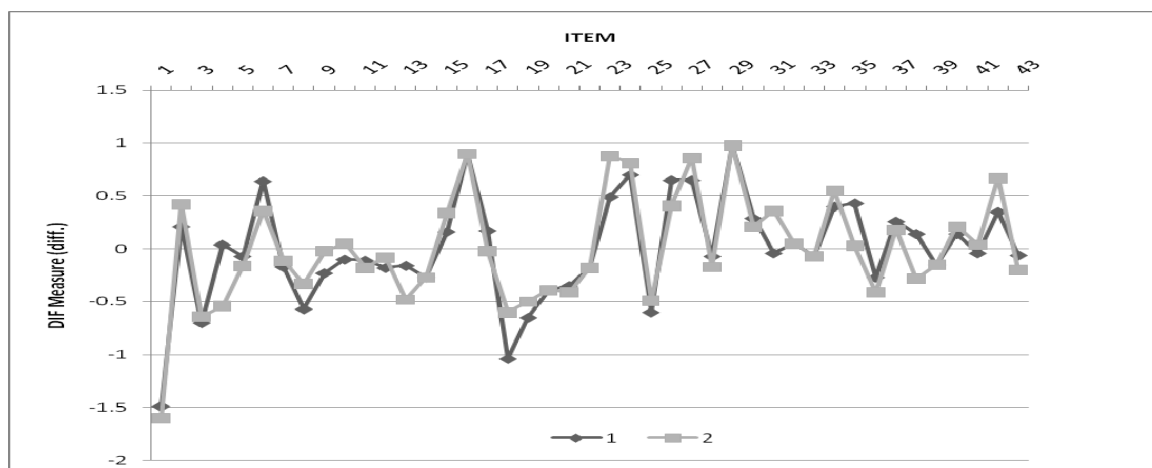


Figure 1 GDIF Plot of teaching skills Items

Table 3 shows results of GDIF analysis on Personality traits items. Analysis revealed that 3 items from the 30 items in Personality traits showed significant GDIF. They are B53, B68 and B74. These 3 items (7.5%) have significant GDIF with indices ranging from 0.55 to 0.90 logits and the t-value of more than 2 ($t \geq 2.0$). Thus these 3 items were suggested to be deleted.

Table 3 GDIF Analysis of Personality traits Construct

Group	DIF Measure (Difficulty measure)	Group	DIF Measure (Difficulty measure)	GDIF Contrast (DIF size)	t	Df	Item Label
1	0.40	2	1.15	-0.75	-2.54	146	B53
1	-0.53	2	0.38	-0.90	-3.13	144	B68
1	1.87	2	1.33	0.55	2.49	147	B74

Figure 2 shows the DIF plot using DIF measure on the analysis of Personality traits construct by gender

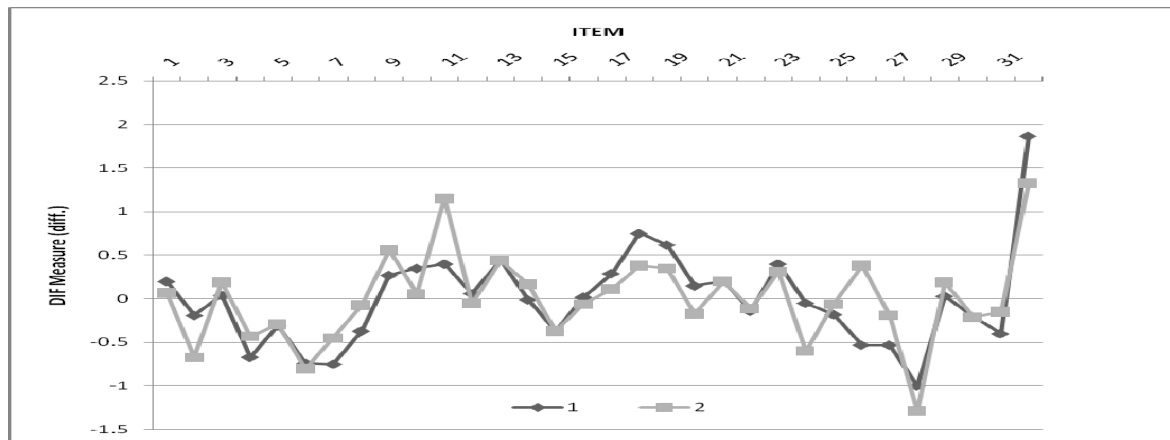


Figure 2 GDIF Plot of Personality traits Items

Table 4 shows results of GDIF analysis on Intelligence items (B75 to B84). Analysis revealed that 2 out of 10 items showed significant GDIF. The items are B78 and B 84. These 2 items (20%) with significant GDIF with indices 0.75 logits indicated t-value of more than 2 ($t \geq 2.0$). Thus, these 2 items were suggested to be deleted because they function differently on students of different gender.

Table 4 GDIF Analysis of Intelligence Construct

Group	DIF Measure (Difficulty measure)	Group	DIF Measure (Difficulty measure)	GDIF Contrast (DIF size)	t	Df	Item Label
1	0.14	2	-0.61	0.75	3.31	143	B78
1	-0.34	2	0.41	-0.75	-2.85	142	B84

Figure 3 shows the DIF plot using DIF measure on the analysis of Intelligence construct by gender

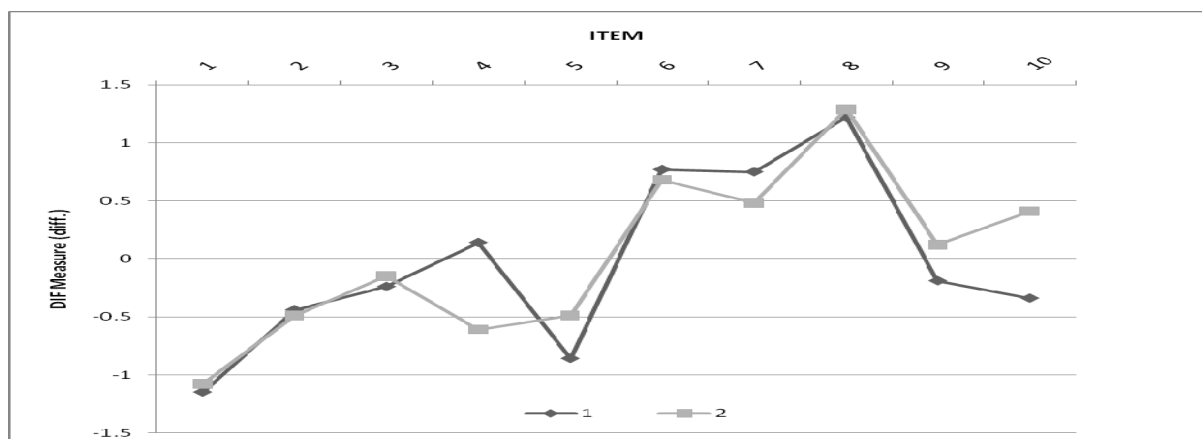


Figure 3 GDIF Plot of Intelligence Items

Table 5 shows results DIF analysis according to gender on 11 studied items of motivations construct. Analysis revealed that 1 out of 8 items in motivations showed significant DIF. It is item B91. This item with significant DIF which indices 0.53 logits indicated the t-value of more than 2 ($t \geq 2.0$). Therefore this item was suggested to be deleted.

Table 5 DIF Analysis of Motivations Skills Construct

Group	DIF Measure (Difficulty measure)	Group	DIF Measure (Difficulty measure)	DIF Contrast (DIF size)	t	Df	Item Label
1	-0.43	2	0.10	-0.53	-2.21	143	B91

Figure 4 shows the DIF plot using DIF measure on the analysis of motivations construct by gender

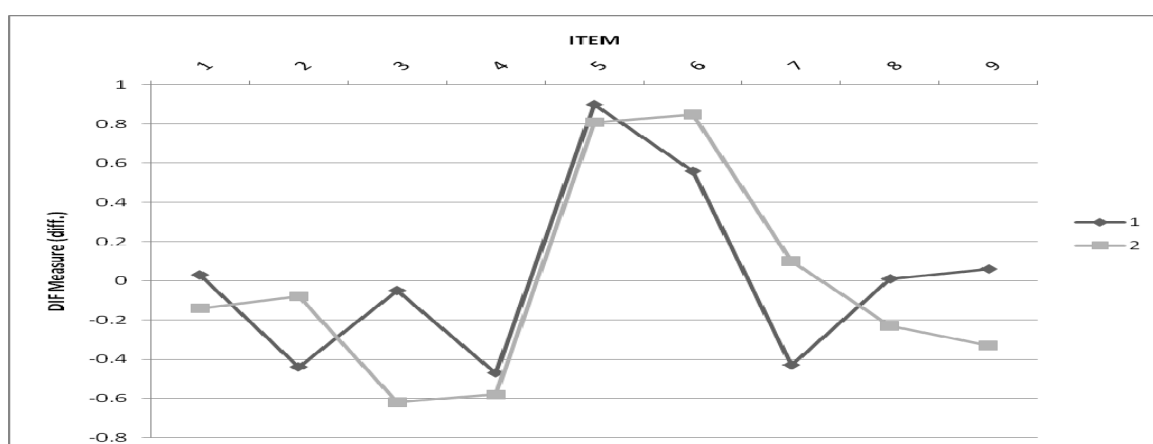


Figure 4 GDIF Plot of motivation Items

Table 6 shows results of DIF analysis on teaching skills items according to stream. Analysis revealed that 3 out of 43 items in teaching skills showed significant DIF. The items were B11, B12 and B18. These 3 items (7 %) with significant DIF indices ranging from 0.45 to -0.95 logits indicated t-value of more than 2 ($t \geq 2.0$). Thus these 3 items were suggested to be deleted.

Table 6 DIF Analysis of Teaching skills Construct According to Stream

Group	DIF Measure (Difficulty measure)	Group	DIF Measure (Difficulty measure)	DIF Contrast (DIF size)	t	Df	Item Label
1	-0.40	2	0.07	-0.47	1.12	146	B11
1	-0.37	2	0.08	-0.45	2.01	146	B12
1	-1.33	2	-0.38	-0.95	2.58	143	B18

Figure 5 shows the DIF plot using DIF measure on the analysis of Teaching skills construct by stream

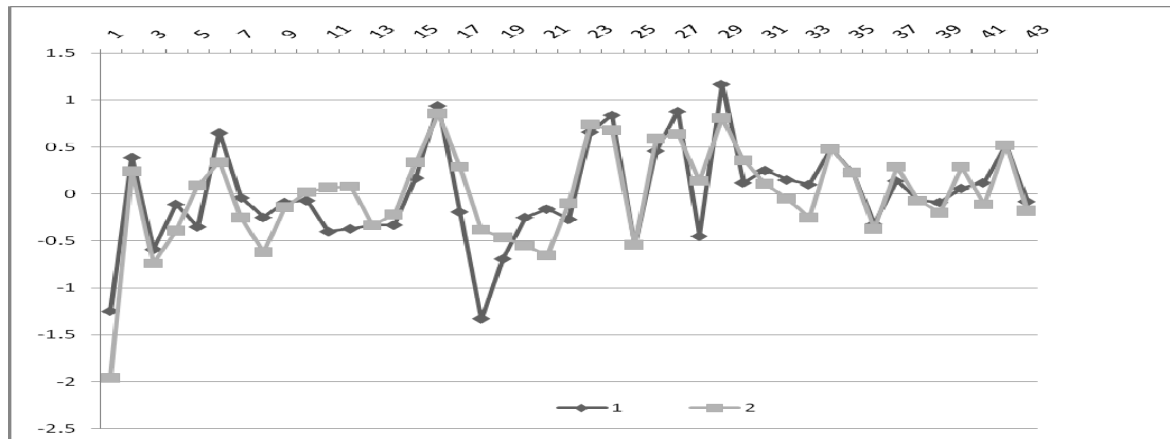


Figure 5 DIF Analysis of Teaching Skill Construct According to stream

Table 6 shows results of DIF analysis on Personality traits items according to stream. Analysis revealed that 3 items from the 30 items in Personality traits showed significant DIF. The items were B44, B61 and B71. These 3 items (10%) with significant DIF with the indices ranging from 0.62 to 0.79 logits indicated t-value of more than 2 ($t \geq 2.0$). Thus these 3 items were deleted because they function differently on students of different stream.

Table 7 DIF Analysis of Personality traits Construct According to Stream

Group	DIF Measure (Difficulty measure)	Group	DIF Measure (Difficulty measure)	DIF Contrast (DIF size)	t	Df	Item Label
1	-0.88	2	-0.09	-0.79	-2.36	142	B44
1	0.77	2	0.16	0.62	2.36	147	B61
1	-0.19	2	0.42	-0.62	-2.25	147	B71

Figure 6 shows the DIF plot using DIF measure on the analysis of personality traits construct by stream

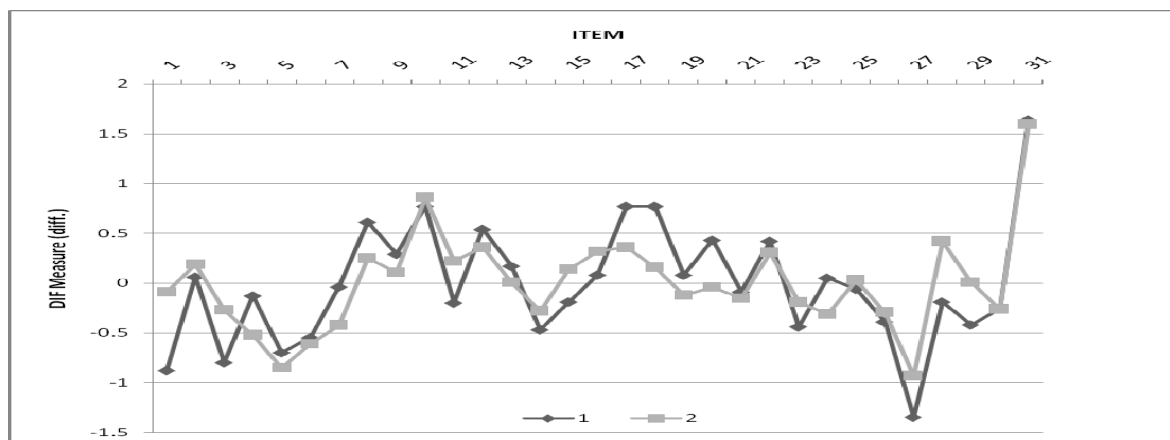


Figure 6 DIF Analysis of Personality traits Construct According to Stream

Table 7 shows results of DIF analysis on Intelligence items according to stream. Analysis revealed that 2 items from the 10 items in Intelligence showed significant DIF. The items are B78 and B84. These 2 items (20%) with significant DIF indices 0.75 logits showed t-value of more than 2 ($t \geq 2.0$). Thus these 2 items were deleted because they function differently on students of different stream.

Table 8 DIF Analysis of Intelligence Construct According to Stream

Group	DIF Measure (Difficulty measure)	Group	DIF Measure (Difficulty measure)	DIF Contrast (DIF size)	t	Df	Item Label
1	0.14	2	-0.61	0.75	3.31	143	B78
1	0.34	2	0.41	-0.75	2.85-	142	B84

Figure 7 shows the DIF plot using DIF measure on the analysis of Intelligence construct by stream.

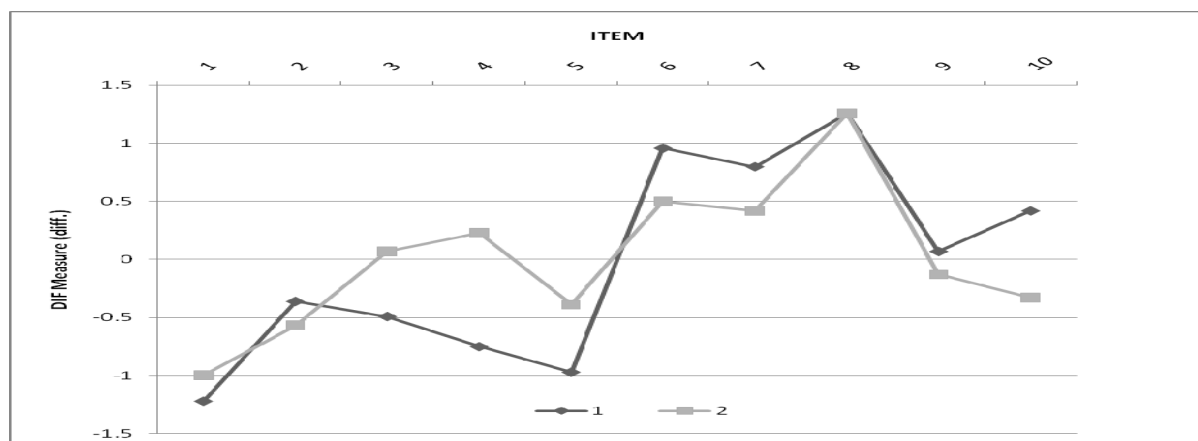


Figure 8 DIF Analysis of Intelligence Construct According to Stream

Table 9 Analysis of Teaching Practice in Islamic Education (TPIED) instrument

No	construct	Items	Items suggested to be dropped based on DIF	
			Gender	Stream
1	Teaching Skills	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18, B19, B20, B21, B22, B23, B24, B25, B26, B27, B28, B29, B30, B31, B32, B33, B34, B35, B36, B37, B38, B39, B40, B41, B42, B43	B4	B11, B12, B18

Continue Table 9

2	Personality Traits	B44, B45, B46, B47, B48, B49, B50, B51, B52, B53, B54, B55, B56, B57, B58, B59, B60, B61, B62, B63, B64, B65, B66, B67, B68, B69, B70, B71, B72, B73, B74	B53, B68, B74	B44, B61, B71
3	Intelligence	B75, B76, B77, B78, B79, B80, B81, B82, B83, B84	B78, B84	B78, B84
4	Motivations	B85, B86, B87, B88, B89, B90, B91, B92, B93, B94, B95	B91	
Total Items		95	7	8

DISCUSSION

The research findings revealed that there are 95 items consist of 43 teaching skills items, 31 personality traits items, 10 Intelligence items and 11 motivations items that are functioning as DIF according to gender and stream. New version of TPIED was established after rigorous process of DIF. This research finding is in accordance with the study by Hambleton & Rogers (2000), Lamprianao et al. (2002), Snider & Styles (2003), Stobart et al. (1992) which stated the presence of DIF is correlated with the factors of individual background such as gender, race, family status, location, facility, teacher, native language and culture. Among frequent factors which become the focus of DIF study is gender Cole, (1997), Linn & Hyde (1989), Stobart & Gipps (1997), Willingham & Cole, (1997) and race Hsin Huang Li & Stout, (1995), Stoneberg (2004). In a comparative study to trace DIF pertaining to gender, Lamprianou et al. (2002) states that gender and language factors are frequently mentioned as the factors that might cause DIF. In DIF study, frequently, the demographics of students become the variable to categorize the study groups being compared.

According to Luppescu (1993), DIF occurs when a certain item shows the different level of difficulty towards certain groups if compared with another group of the same capability. In other words if one group finds the item is much harder, then DIF persists. Nunnaly & Bernstein (1994) clarify that DIF exists when an item which is relatively much harder and discriminated or can easily be conjectured as a group compared to the others. Impartiality is an important aspect for authenticity, one technical aspect which becomes the barometer for standardization for quality test Messick in William & Cole (1997). All matters that affect impartiality will affect the authenticity of the test. According to Shealy & Stout (1993) it is found that an imbalanced study will affect the items produced cumulatively by the bias individuals. Imbalanced test means it is not implemented using the same construction when it is administered to the different subgroups. This clearly shows that the test does not show the same authenticity to the different subgroups.

GDIF and DIF analysis carried out on TPIED is an effort to ascertain the test carried out to the students involved is equally formed and carried out. DIF analysis is widely applied for instruments building involving cognitive area, academic subjects evaluation and the survey items are meant for qualifying test, promotion exercise, licensing and publication Dodeen (2004), Zieky (2002). According to Dorans & Holland (1993), Dorans & Kulick (1986), Holland & Thyer (1988) there are differences on the ability to respond to each item between groups who sit for the evaluation in various areas.

DIF study in education instruments evaluation building is an important procedure which is aimed at identifying the test which does not show the same function or when administered to student groups of the same capability but different in their background. From the item aspect itself based on previous studies generally it conforms to the main cause of DIF that is present in multi-dimensionality which is measured in the item Shealy & Stout (1993). This means the item actually is measured in more than one dimension other than main dimension which should be measured. The unplanned dimension may be presented with the type content and item presentation methodology in the test used.

CONCLUSION AND RECOMMENDATIONS

Based on plural school population and with variety of students, the Teaching Practice skills evaluation should be carried out as partially as possible. Therefore the GDIF and DIF analysis is the most significant method in the study. This finding is supported by Dodeen & Johanson (2001), Dodeen (2004) that DIF is found in instrument research and other tests. Therefore DIF analysis is suggested to all instruments items which are administered to all individual of different background if the researcher values impartiality or fairness in the built instruments. Impartiality in the test is one aspect of instrument evaluation effectiveness. Since the item is the basic unit in instrument evaluation the resulting item should be stable.

The methodology for DIF in item analysis is initiated with the background of two theories of main measurement which becomes the basis for building and item analysis test that is Classical Test Theory (CTT) and Item Response Theory (IRT). CTT and IRT are two theories that explained the performance of item. Both are usually used in studies which involve measurement issues such as devising test, equating of exam score and identification of biased item. A quantitative item analysis is carried out to see the performance of each item by scrutinizing the item's function on students taking the test. The analysis result enables the test maker to identify the item with less quality or which does not function as expected. Although the designer has evaluated the item at writing stage, the item which is analyzed empirically is important to harmonize every item to achieve the authentic standard and reliability to the maximum. The TPIED item acquired is significant at t-value exceeding 2.0 ($t \geq 2.0$) is preferably dropped out or referred to the expert, because there significant differences from difficult item perspective and the groups being tested.

The DIF analysis is a procedure in item analysis that can be carried out using the CTT, IRT or both combinations. DIF is the last resort in the process of quality study of item (Linacre, 2005) to strengthen the instrument. The principle used which is also an assumption is that all DIF tests on students of the same capability, should show almost the same performance when answering the test without considering the gender, race, stream, the number of years at school and other demographic factors.

REFERENCES

Al-Quran

- Ab.Halim Tamuri & Yusoff, N. M. (2010). *Kaedah Pengajaran dan Pembelajaran Pendidikan Islam*. Bangi: Penerbit Universiti Kebangsaan Malaysia.
- Ab.Halim Tamuri, Yusopp, A. Osman, K. Razak, K. A. Awaludin, S., & Rahim, Z. A. (2004). Keberkesanan pengajaran dan pembelajaran Pendidikan Islam ke atas pembangunan diri pelajar. Bangi: Laporan Penyelidikan Fakulti Pendidikan UKM.
- Ab.Halim Tamuri. (1 April 2006). *Model Guru Pendidikan Islam:Konsep 5 Mim*. Dlm. *Seminar Amalan Pengajaran Guru Pendidikan Islam*. Bangi: Fakulti Pendidikan UKM.
- Ab.Halim Tamuri. (2007). Islamic Education teachers' perceptions of the teaching of akhlaq in Malaysian secondary schools. *Journal of Moral Education*, 36 : 371-386.
- Abdul Aziz Mohd Zin. (1995). Masalah Matapelajaran Pendidikan Islam di sekolah-sekolah menengah kerajaan di Kuala Lumpur. *Jurnal Pendidikan Islam*, 7: 165-176.
- Abdul Halim Hj Mat Diah. (1989). *Pendidikan Islam Di Malaysia*. Kuala Lumpur: Angkatan Belia Islam Malaysia.
- Abdul Salam Yusoff. (2003). *Imam Al-Ghazali Pendidikan Berkesan*. Kuala Lumpur: Utusan Publications & Distributors Sdn.Bhd.
- Abdull Sukor Shaari. (2008). *Guru Berkesan Petua dan Panduan*. Sintok: Penerbit Universiti Utara Malaysia.
- Abdullah al-Amin al-Na'my, & Terj.Mohd Romzi Omar. (1994). *Kaedah dan teknik pengajaran menurut Ibn Khaldun dan Al-Qabisi*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Abdullah Ishak. (1995). *Pendidikan Islam dan pengaruhnya di Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Ahmad Mohd. Salleh. 1997. *Pendidikan Islam Falsafah, Pedagogi dan Metodologi*. Kuala Lumpur: Penerbit Fajar Bakti Sdn.Bhd.
- Al-Attas, S. M. (1980). *The concept of education in Islam*. Kuala Lumpur: Angkatan Belia Islam Malaysia (ABIM).
- Allan C.Ornstein. (1995). *Teaching Theory into Practice*. United States of America: Allyn and Bacon.
- Alphen, A.V; Halfens, R.; Hasman, A; and Imbos, T. (1994). Likert Or Rasch? Nothing is More Applicable Than Good Theory. *Journal Of Advanced Nursing*, 20:196-201.
- Ang Huat Bin. (1999). *Konsep Dan Kaedah Pengajaran dengan Penekanan Pendidikan Inklusif*. Kuala Lumpur: Utusan Publications & Distributors Sdn.Bhd.
- Anisah Bahiyah Haji Ahmad. (2006). *Sejarah Pendidikan Pada Zaman Rasulullah S.A.W*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Asmawati Suhid. Januari. (2006). Adab dan Akhlak Islam dalam Pendidikan Islam: Satu Kajian Kes di Selangor. *Jurnal CITU*, 2: 53-66.
- Azhar Ahmad, & Ab.Halim Tamuri. (2007). Penghayatan Akhlak Pelajar Sekolah Menengah Masa Kini. *Jurnal Yadim*, 5: 16-29.
- Azizi Ahmad. (2010). *Pentaksiran Pembelajaran*. Ampang, Selangor: Dawama Sdn. Bhd.
- Azma Mahmood. (2005). Pengukuran Tahap Penghayatan Pendidikan Islam Pelajar-pelajar Sekolah Menengah di Malaysia. Tesis PhD. UKM Bangi.
- Azrilah Abdul Aziz. (2010). *Rash Model Fundamentals: Scale Construct and Measurement Structure*. Kuala Lumpur: UTM.
- Bath, D., Smith, C., Stein, S. & Swann, R. (2004). Beyond Mapping and Embedding Graduate Attributes: Bringing Together Quality Assurance and Action Learning to Create A Validated and Living Curriculum. *Higher Education Research and Development*, 23(3): 313 – 328.

- Benjamin D. Wright Geoffrey N. Masters. (1982). *Rating Scale analysis Rasch Measurement*. Chicago: Mesa Press.
- Benjamin D. Wright Mark H. Stone. (1979). *Best Test design Rasch Measurement*. Chicago: Mesa Press.
- Berk, R.A. 1982. *Handbook of Methods for Detecting Test Bias*. Baltimore: Johns Hopkins University Press.
- Biggs, J. (2003). *Teaching for quality learning at university*. Maidenhead: Society for Research into Higher Education & Open University Press.
- Bond, T.G. & Fox, C.M. (2001). *Applying the Rasch Model: Fundamental Measurement in The Human Sciences*. New Jersey: Lawrence Erlbaum Associates Publishers, London.
- Brians, H. (1998). *Teacher and classroom communication*. Australia: Harcourt Brace Jovanovich Publishers.
- Bryan Coombs, & terj.Siti Aishah Mohd Elias. (2009). *Mengajar secara efektif*. Kuala Lumpur: Institut Terjemahan Negara Malaysia Berhad.
- Camilli, G. 1993. *The case against item bias detection techniques based on internal criteria: Do item bias procedures obscure test fairness issues*. In *Differential item functioning*. Holland, P.W. & Wainer, H. New Jersey: Lawrence Erlbaum Associates Publishers London.
- Chris Kyriacou. (1992). *Effective Teaching In Schools*. Great Britain: T J Press (Padstow) Ltd.
- Cole, N.S. (1997). The EST gender study: How females and males perform in educational settings. *EST technical Report*.
- Dodeen, H.(2004). Stability of differential item functioning over a single population in survey data. *The journal of Experimental* 72(3) : 181-194.
- Dodeen. H & Johanson. G. (2001). *The prevalence of gender DIF in survey data*. *Annual Meeting Proposal of the American Educational Research Association*. April 2001, Seattle: 10-14
- Dorans, N.J. & Holland, P.W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In *Differential Item Functioning*. Edited by Holland, P.W. & Wainer, H. Hillsdale, N.J: Lawrence Erlbaum Associates.
- Dorans, N.J & Kulick, E. (1998). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement* 23(4): 355-368.
- Earl Babbie. (1999). *The Basics of Sosial Research*. Boston: Wadsworth Publishing Company.
- Eunson, B. (2007). *Communication in the Workplace*, First Published, John Wiley & Sons Australia, Ltd. (MPHMV).
- Gamal Abdul Nasir Zakaria. (2003). *Prinsip-prinsip Pendidikan Islam*. Pahang: PTS Publication.
- Al-Ghazali, & Terj.Ahmad Abdurraziq al-Bakri. (2009). *Ringkasan Ihya' Ulumuddin*. PT.Sahara Intisains.
- Al-Ghazali, a.-I. A.-G. (t.t). *Ihya' 'ulum al-din. Jil. 1-5*. Misr: al-Maktabah al=Taufiqiyyah.
- Ghazali Darusalam. (2001). *Pedagogi Pendidikan Islam*. Kuala Lumpur: Utusan Publications & Distributors Sdn. Bhd.
- Halstead, J.M. (2004). *An Islamic concept of education, Comparative Education*, 40 (4), 517- 530.
- Hambleton, R.K. & Rogers, H.J. (2000). *Developing an item bias review form*. *University of Massachusetts at Amherst*. <http://ericae.net/fit/tamu/biaspub2.Htm>, 22/09/2011
- Hamid Fahmy Zarkasyi. (1990). *Pemikiran al-Ghazali tentang pendidikan*. Kuala Lumpur; Dewan Bahasa dan Pustaka.
- Hasan Langgulung. (1991). *Pendidikan Islam menghadapi Abad ke-21*. Shah Alam: HIZBI.
- Hsin-Huang Li & Stout, W. (1995). A new procedure for detection of crossing DIF: *Research Report*. University of Illinois at Urbana-Champaign, Illinois.

- James Calderhead and Susan B. Shorrock. (1997). *Understanding Teacher Education Case Studies in the Professional Development of Beginning Teachers*. London: The Falmer Press.
- Kamarul Azmi Jasmi & Ab.Halim Tamuri. (2007). *Pendidikan Islam kaedah Pengajaran & Pembelajaran*. Johor: Penerbit Universiti Teknologi Malaysia.
- Kamarul Azmi Jasmi. (2010). Guru Cemerlang Pendidikan Islam Sekolah Menengah Di Malaysia: Satu Kajian Kes. Tesis Doktor Falsafah, Universiti Kebangsaan Malaysia.
- Lai, J.S. & Eton, D.T. (2002). Clinically Meaningful Gaps. *Rasch Measurement Transactions* 15 (4) : 850.
- Lamprianou, I., Boyle, B. & Nelson, N. (2002). Using optimal appropriateness measurement to detect examines mostly affected by differential item functioning in the year 2000. Key stage 2, science, national curriculum tests in England. *Educational Measurement, Psychometrics and Assessment*, 20: 200-215.
- Linacre, J.M. (2005). *A User's Guide To WINSTEPS®MINISTEPS: Rasch-Model Compute Programs*.
- Maller, S.J.(2001). Differential item functioning in the WISC-III. Item parametres for boys and girls in the national standardization sample. *Educational and Psychological Measurement*, 61(5): 793-817.
- Michael J. Dunkin & Bruce J. Biddle. (1974). *The Study of Teaching*. United States of America: Holt, Rinehart and Winston, Inc.
- Misnan Jemali. (2008). Amalan Pengajaran Guru dalam Pengajaran dan Pembelajaran Tilawah Al-Quran sekolah-sekolah Menengah. Tesis PhD. UKM Bangi.
- Mohd Aderi Che Noh. (2008). Hubungan Antara Amalan Pengajaran Guru dan Pencapaian Tilawah al-Quran pelajar tingkatan dua di Malaysia. Tesis PhD. UKM Bangi.
- Mohd.Zuhairi Safuan. September (2000). Pandangan Fazlur Rahman Terhadap beberapa isu dalam Pendidikan Islam. *Tasawwur Islam* , 4: 163-172.
- Monica J. Taylor, J.Mark Halstead. (2007). Islami Values and Moral Education..*Journal of Moral Education*, 36: 283-296.
- Morris L.Bigge. (1964). *Learning Theories for Teachers*. United States Of America: Harper & Row, Publishers.
- Mujibul Hasan Siddiqui. (2008). *Models of Teaching*. New Delhi: A P H Publishing Corporation.
- Nunnally, J.C. & Bernstein, I. (1994). *Psychometrics Theory*. Ed.3. New York: McGraw-Hill.
- Osterlind, S.J. (1989). *Constructing Test Items*. Boston: Kluwer Academic Publishers.
- Pallant, J.F. & Tennant, A.(2007). An introduction to Rasch measurement model: An example using the hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 46: 1-18
- Rahil Mahyuddin, Habibah Elias, & Bakar, K. A. (2009). *Amalan Pengajaran Berkesan*. Shah Alam: Karisma Publications Sdn.Bhd.
- Roland, B., Adler & Elmhurst, J.M. (2008). *Communicating at Work: Principles and Practices for Business and the Professions*, McGraw-Hill Companies, Inc. (KINO).
- Rosenbusch, K. & Townsend, C. (2004). The Relationship of Gender and Organizational Setting to Transformational and Transactional Leadership Skills of Selected College Student Leaders, *Journal of Leadership Education*, 3 : 245 -260
- Shahril @ Charil Marzuki. (2002). Ciri-ciri guru yang unggul:satu kajian kes di beberapa buah sekolah menengah di Malaysia. *Jurnal Pendidikan* , 22: 97-110.
- Shealy, R. & Stout, W.F. (1993). A model-based standardization approach that separates true bias/ DTF as well as item bias/DIF. *Psychometrika* 58: 159-194.
- Smith, R.M. (1991). The Distributional Properties of Rasch Item Fit Statistics. *Educational Psychological Measurement*, 51, 541-565.

- Stobart, G & Gipps, C. (1997). *Assessment : A Teacher's Guide the Issues. Ed. 3. London: Hodder & Stoughton. Ning (DIF) in the spring 2003 Idaho Standards Achievement rtest Applying the Simultaneous Bias Test (SIBTEST) and the Mantel Haenszel Chi-squaretest.*(onlinereport) <http://www.sde...state.id.us/admin/docs/ayp/ISATDIFStudy.pdf> 14/08/2011
- Stoneberg, Jr. B.D. (2004). A Study of Gender-Based and Ethnic-Based Differential item Functioning (DIF) in the spring 2003 Idaho Standards Achievement Test Applying
- Syed M. Naquib Al-Attas, & Terj. Wan Mohd Nor Wan Daud. 2005. *Falsafah dan Amalan Pendidikan Islam Satu Huraian Konsep Asli Islamisasi*. Kuala Lumpur: Penerbit Universiti Malaya.
- Wright B. & Linacre J. (1992). Combining and splitting categories. *Rasch Measurement Transactions*, 6: 233-235.
- Wright, B.D. & Masters, G.N. (1982). *Rating Scale Analysis*. Rasch Measurement Mesa Press Chicago.
- Wright, B.D. & Linacre, J.M. (1992). *A User's Guide To Bigsteps, Rasch Model Compute Program Version 2.2*. MESA Press.
- Wright, B., & Stone, M.H. (1979). *Best Test Design*. Chicago: MESA Press.
- Wright, B.D. & Masters, G.N. (1982). *Rating Scale Analysis*, Chicago: MESA Press
- Zawawi Ahmad. (1998). *Strategi pengajaran Pendidikan Islam*. Selangor: Penerbit Fajar Bakti Sdn Bhd.

Biodata:

Sarimah Mokhtar, Faculty of Education, Universiti Kebangsaan Malaysia email: umimur-syid@yahoo.com (Student PhD in Islamic Education)

Mohd Kashfi Mohd Jailani, Faculty of Education Universiti Kebangsaan Malaysia. email: kashfi2011@gmail.com (Student PhD in Measurement and Evaluation)

Prof. Madya Dato' Dr. Ab. Halim Tamuri, Faculty in Education Universiti Kebangsaan Malaysia. email: tamuri67@gmail.com (Senior Lecturer)

Dr. Mohd Aderi Che Noh, Faculty of Education Universiti Kebangsaan Malaysia. email: aderi@ukm.my (Lecturer)

Dr. Kamarulzaman Abdul Ghani, Faculty of Education Universiti Kebangsaan Malaysia. email: qamar68@ukm.my (Senior Lecturer)