

## **Do You Mean What I Mean? Cultural Factors and Vocabulary Knowledge**

**Nora Liza Abdul Kadir & Siti Rahayah Ariffin**

*Faculty of Education, Universiti Kebangsaan Malaysia  
43600 UKM Bangi, Selangor Darul Ehsan, Malaysia  
e-mail: norakadiryusoff@yahoo.com, sitira@ukm.my*

### **Abstract**

*The focus of this research is to see if one's cultural factors do influence their performance in answering items in an intelligence test done in one of the public universities in Malaysia. It is an instrument used to measure one's intelligence level and it is designed based on the Raven Standard Progressive Matrices – Plus Version and Mill Vocabulary. This is a quantitative study where the participants were selected using simple random sampling method. 773 students from one of the public universities in Malaysia, of different ethnic backgrounds (545 Malays, 228 non-Malays) from eight faculties have taken this test. The data obtained was analysed using Winsteps version 3.64.2, and in the 88 vocabulary items (open ended and multiple-choice questions), race-related DIF was found in a number of items. It could be concluded that cultural factors do influence the students' ability in answering the intelligence test, which means that elements of bias due to incompatible cultural factors between the test builders and test takers do exist. It is hoped that those involved in testing to be more aware of the possibility of unintentional bias towards one's race occurring in test papers, and to be more sensitive in constructing items in any instruments that is used to measure one's ability.*

**Keywords** *Differential Item Functioning (DIF), intelligence test, vocabulary, ethnic background, cultural background*

### **INTRODUCTION**

In the effort to identify and educate gifted students in Malaysia, one of the public universities in Malaysia has been entrusted by the Malaysian Government to carry out programmes for the gifted individuals, and also to come up with instruments that could be used to identify this group of special people. Therefore, several instruments have been constructed, and one of them was studied by the researcher. This instrument was designed based on a well-known intelligence test, which is the Raven Standard Progressive Matrices – Plus Version and Mill Vocabulary (Noriah, Rosadah and Siti Fatimah, 2009). There are three sections in this instrument, which are Part A (Symbol Search) that tests students ability to identify symbols, Part B (Open Ended Questions – Vocabulary Skills) and Part C (Multiple Choice Questions – Vocabulary Skills) test students' mastery of the vocabulary (Noriah et al., 2009).

There are several aspects that offer researchers vast research opportunities in regards with gifted students. One example is comparing the mastery of vocabulary and the proficiency in a language between gifted and ordinary students. Another aspect that could be done a study on is to compare the rate of vocabulary mastery and proficiency between gifted students in Malaysia and those in the foreign countries, and this comparison could be made based on the students' performance in this intelligence test and in Raven Standard Progressive – Plus Version and Mill Vocabulary. However, this paper would focus on finding out if differential item functioning (DIF) do exist in items in Part B and Part C. The reason to this is that the way candidates answer the items can be influenced by the demographic factors, and in this study, it is their ethnical/cultural backgrounds.

### **REVIEW OF LITERATURE**

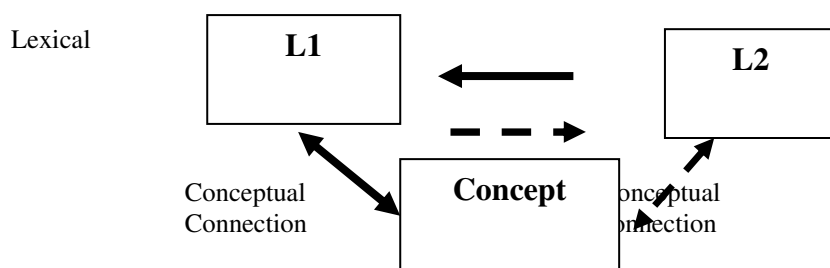
Numerous issues were raised by several parties, where they question if the tests taken by candidates are not being biased against them. An example of biasness that could occur is language bias. This was mentioned in Schon, Shaftel and Markham (2008, p. 106) where Ortiz and Ochoa (2005b, p. 156)

stated that "...tests are linguistically biased not because of any inherent psychometrics defect, but simply because of the expectations and assumptions regarding the comparability of language that are rarely met when working with diverse individuals". Another issue that may rise is when questions or items are translated from the original language into another language, as this may affect the level of difficulty of items, as some terms do not exist in a candidate's mother tongue which may differ from the language used in the questions (Schon, Shaftel & Markham, 2008).

In regards with students with diverse backgrounds and fairness in the tests or examinations, Huang (2009) wrote that many organisations, institutions and also professional individuals who are stakeholders in the education field were all in agreement saying that measurements in the field of education strongly emphasize the importance of fairness to meet the needs of candidates who come from various races, different genders and ethnic backgrounds. Fox and Chang (2007) also mentioned that the aspects of language and culture should be taken into account when drafting question papers. One example given by Lynn, Backhoff and Contreras (2005) is the problem of language bias against students of Native American descent, where their mother tongue is different from the second language learned in school (English) and this affected the evaluation of their IQ level when they answered an intelligence test that is written in English. This view was also agreed by Gray, McCallum and Bain (2009) where they did mention that there are few intelligence tests which are less sensitive to the limits of language proficiency of a candidate.

How one's culture background may influence his or her knowledge also drew interest among researchers, and Davies (2006) claimed that culture influences one's pattern of thinking and not surprisingly this view is increasingly being accepted by psychologists. One's schema, which is a mental structure, is formed as a result from the interaction between one's brain with his or her surrounding, and interestingly, because the schemas are formed in each individual, therefore it represents the individual's experience and culture, and this is an opinion expressed by Quinn and Holland (1987) (Brooks & Karathanos, 2009). Fox and Cheng (2007) mentioned an example given by Sasaki (2000) where students can complete a cloze text well if they found that the text contains similar cultural aspects as theirs. Davis (2006) also mentioned that many researchers have discovered that there are differences in the way of thinking between those of Western and Asian origins, which is caused mainly due to the differences in the mother tongue, as well as in cultural backgrounds. Some researchers have found that the linguistic aspects in examination questions can affect students' achievement; Solano-Flores (2008) states that ESL students taking the examination of a subject written in their second language are indirectly being tested in their proficiency of the second language.

A model mentioned by Sunderman & Kroll (2006), which was adapted from Kroll and Stewart (1994), can be used to explain why bilingual students at times face difficulties in translating from one language into another, and this model is known as the Revised Hierarchical Model (RHM), which is shown in Figure 1:



**Figure 1** Revised Hierarchical

Looking at Figure 1, a bilingual's memory is compartmentalised into two; one to keep the word form at its lexical level and another is to keep the meaning of the word at the conceptual level (Ferré, Sánchez-Casas & Guasch, 2006). According to Sunderman & Kroll (2006) this model describes how concept and lexical are interrelated when a person is getting more skilled with a second language (L2). In the early stages of L2 learning, it is believed that a word in L2 is related to its meaning in the

mother tongue (L1), which means that the L1 translation of the word makes it easier for a person to know the meaning of the word in L2 (Sunderman & Kroll, 2006). Comensaña, Perea, Piñeiro & Fraga (2009) mentioned that when a person is at the initial stage in learning L2, all the concepts he knew could only be accessed through his L1, and this view is agreed by Sunderman & Kroll (2006) where they were at the opinion that at this level, the connections between words in L1 and the concepts are much stronger compared to that between the words in L2 and the concepts.

Ferré, Sánchez-Casas & Guasch (2006) stated that in the early stages of L2 learning, the learners tend to confuse in the spelling of a word in L1 and L2, because as stated earlier, the connection between L1 and the concept is stronger than that of L2 with the concept. Ferré, Sánchez-Casas & Guasch (2006) gave an example where learners of Spanish as their L2 were asked to determine whether the translation of a word in L1 (e.g. garlic) to L2 (*ajo*) is correct; they were given two words in L2, where one has the same form as *ajo*, (*ojo* – eye) and the other word has the same meaning as *ajo* (*cebolla* – onion), and it was found that these new learners tend to choose words that have similar forms, in this case they think that *ajo*=*ojo* is correct, and not *ajo*=*cebolla*. But as they become more fluent in L2, the problem mentioned will lessen, but they will be confused with the meaning (semantic) of the L2 word given. An example used by Comensaña, Perea, Piñeiro and Fraga (2009) to illustrate this is the outcome from a study done by Talamas et al. (1999), where people who are quite fluent in Spanish, which is their L2, were asked to select the correct pair of English – Spanish words, which are chair – *silla* (correct answer), table – *silla* & mouse – *silla* (incorrect answers), and it was found that it is rather difficult for them to state that table – *silla* is incorrect, due to the semantic interference.

Park, Pearson & Reckase (2005) said that according to Rudas & Zwick (1997), differential item functioning or DIF occurs when there exist a probability that someone from a small group to give a correct answer, and these estimations of probability are "...matched for ranges of a score on a related conditioning variables" (p. 81). Camilli & Shepard (1994) offer some explanations regarding how DIF could occur, where they said some items were written in a way that the wordings may cause discomfort or upset a particular group, some may even cause the candidates not to be able to display their true ability, wrongly worded items could defeat the actual purposes of the items which is to test a particular construct, and this was mentioned by (Young & Sudweeks 2005) in their article. Park, Pearson & Reckase (2005) also mentioned that DIF can also be seen as a psychometric technique that can be used to detect the difference in performances between two groups which the same ability is tested (Dorans & Holland, 1993, p. 35). Still on the performance between groups, Banks (2009) cited a study done by Holland & Wainer (1993), where they stated that when other abilities are controlled, different groups may have different probability to give the correct answer, and this is when we can say DIF has happened. According to Mendes-Barnett and Ercikan (2006), it is difficult to pinpoint the source of DIF in a mathematic test paper, but they mentioned that O'Neill & McPeck (1993) did a study and they have concluded that the content and context of the items may be the source of DIF. Numerous studies have been done, using the confirmatory approach that was suggested by Shealy & Stout (1993), to detect the occurrence of gender-related DIF in multiple choice questions in mathematic test papers, like what has been done by Ryan and Fan (1996), or a study to observe gender-related DIF in items that require high order thinking or those that contains diagrams or graphs (Ryan & Chiu 2001) (in Mendes-Barnett & Ercikan 2006). O'Neill & McPeck (1993) have also called for more research to be done on test item functioning and aspects that are involved in the issue of students of various backgrounds in a class (Park, Pearson & Reckase 2005).

This study aims to determine whether language bias does occur in one intelligence test, in which it is believed that this may affect students' of multicultural backgrounds performance in answering items in Part B (Open Ended Questions - Vocabulary) and Section C (Multiple Choice Questions - Vocabulary). Part A (Symbol Search) is not examined because it only tests the candidates' ability to manipulate symbols. This study intends to provide new information and enrich the knowledge of all those involved or interested in the construction and use of intelligence test and also instruments to measure students' general ability on how the performances of students can be influenced by their backgrounds

## **METHODOLOGY**

In getting the data needed for this research, a quantitative study was done, in which the sample were selected using the random sampling method. It was decided that the sample was to be taken from eight different faculties in a public university in Malaysia, and a total of 773 samples were obtained, in which they volunteered to do the test. For this study, an instrument that was designed to measure a person's intelligence level was used to obtain the data. This instrument consists of three major sections, which are Part A (Symbol Search), which has 60 questions, Section B (Open Ended Questions - Vocabulary) with 44 questions and Section C (Multiple Choice Questions - Vocabulary) that consists of 44 questions. To answer the research questions posed in this study, only the responses of the individuals in Part B and Part C are analysed, using the Winsteps 3.64.2 software to identify DIFs in the items, or in other words, to identify items that are potentially biased to a certain group of race, where in this case, the individuals were grouped as either Malays or other ethnic groups.

## RESULTS

Table 1 contains the demographic information of the candidates who answered the intelligence test. The Chinese, Indians and other ethnic backgrounds that are non-Malays are put into one group labelled as 'other ethnic groups', and the Malays are labelled as 'Malays'. This was done so that a clearer result could be obtained. The division of the races is shown in Table 1:

**Table 1** Demographic information of candidates who answered the intelligence test

Gender	Malay	Other Ethnic Groups
Male	165	81
Female	380	147
<b>Total</b>	<b>545</b>	<b>228</b>

In regards to the reliability of the individuals in answering the items, it was found that the reliability value for the persons who have answered the items for Part B is 0.62 and for Part C is 0.43. SitiRahayah (2008) said that Bond and Fox (2001) have mentioned that the best value for reliability should be more than 0.8 ( $> 0.8$ ) for it to be accepted. This value indicated that the individuals were inconsistent in giving their answers. Whereas for the reliability of the items, the value for the items in Part B and Part C were very good, this is at 0.99. This means that these items were consistent in measuring the ability of the candidates.

With regards to the reliability of individuals who answered this instrument, the reliability for Part B is 0.62, and for Part C the value is 0.43. Bond and Fox (2001) stated that the best reliability should be more than 0.8 ( $> 0.8$ ) for it to be accepted (SitiRahayahAriffin, 2008). This indicates that individuals who took this test were inconsistent in giving their answers. But for the reliability of the items, the reliability value for items in Part B and C were very good, which is 0.99. This means that these items are consistent in measuring the ability of candidates.

Table 2 shows items that were found to have race-related DIF in Part B (Open-ended Question – Vocabulary):

**Table 2** DIF in items in Part B (Open-ended Question – Vocabulary)

Person Class	DIF Measure	DIF S.E.	Person Class	DIF Measure	DIF S.E.	DIF Contrast	Joint S.E.	t	d.f.	Name
1	-1.53	0.09	2	-0.83	0.14	-0.7	0.17	-4.2	771	q061
1	-1.4	0.09	2	-0.24	0.15	-1.16	0.17	-6.65	771	q071
1	-3.54	0.18	2	-2.24	0.16	-1.31	0.24	-5.55	771	q072
1	-0.9	0.09	2	0.09	0.16	-0.99	0.18	-5.45	771	q083
1	-1.7	0.1	2	-0.79	0.14	-0.91	0.17	-5.4	771	q070
1	-0.87	0.09	2	0.09	0.16	-0.95	0.18	-5.28	771	q091
1	-1.53	0.09	2	-0.69	0.14	-0.84	0.17	-4.97	771	q064
1	-0.78	0.09	2	0.06	0.16	-0.84	0.18	-4.69	771	q074

1	-1.56	0.09	2	-0.94	0.14	-0.62	0.17	-3.73	771	q080
1	-1.5	0.09	2	-1.07	0.14	-0.43	0.17	-2.58	771	q078
1	-2.35	0.11	2	-1.9	0.15	-0.45	0.19	-2.44	771	q062
1	-4.09	0.22	2	-3.38	0.22	-0.71	0.31	-2.26	771	q079
1	-0.22	0.09	2	0.11	0.16	-0.33	0.18	-1.83	771	q081
1	-0.85	0.09	2	-0.57	0.14	-0.28	0.17	-1.68	771	q099
1	-2.38	0.11	2	-2.1	0.15	-0.28	0.19	-1.47	771	q068
1	-0.44	0.09	2	-0.2	0.15	-0.24	0.17	-1.41	771	q104
1	-2.7	0.13	2	-2.42	0.16	-0.28	0.21	-1.36	771	q075
1	-0.62	0.09	2	-0.41	0.14	-0.21	0.17	-1.25	771	q093
1	-2.39	0.12	2	-2.24	0.16	-0.15	0.19	-0.78	771	q066
1	-0.28	0.09	2	-0.15	0.15	-0.13	0.18	-0.74	771	q067
1	0.88	0.12	2	1.01	0.21	-0.13	0.24	-0.53	771	q089
1	2.08	0.18	2	1.91	0.3	0.17	0.35	0.48	771	q084
1	-0.37	0.09	2	-0.47	0.14	0.1	0.17	0.61	771	q098
1	1.5	0.14	2	1.24	0.23	0.25	0.27	0.94	771	q092
1	0.25	0.1	2	0.04	0.16	0.21	0.18	1.16	771	q095
1	1.93	0.17	2	1.3	0.23	0.63	0.29	2.19	771	q102
1	4.03	0.45	2	2.64	0.42	1.39	0.61	2.27	771	q086
1	0.18	0.1	2	-0.26	0.15	0.44	0.18	2.48	771	q065
1	-0.6	0.09	2	-1.03	0.14	0.43	0.16	2.63	771	q077
1	4.55	0.58	2	2.64	0.42	1.91	0.71	2.68	771	q096
1	6.84<	1.81	2	1.91	0.3	4.93	1.83	2.69	771	q087
1	5.65	1	2	2.64	0.42	3.01	1.08	2.78	771	q088
1	6.83<	1.8	2	1.41	0.24	5.42	1.81	2.99	771	q085
1	-0.16	0.09	2	-0.79	0.14	0.63	0.17	3.8	771	q073
1	1.44	0.14	2	0.56	0.18	0.87	0.23	3.84	771	q101
1	-2.05	0.1	2	-2.9	0.19	0.84	0.21	3.94	771	q069
1	2.08	0.18	2	0.97	0.2	1.11	0.27	4.08	771	q100
1	0.19	0.1	2	-0.65	0.14	0.85	0.17	4.97	771	q063
1	0.87	0.11	2	-0.13	0.15	1	0.19	5.26	771	q082
1	1.6	0.15	2	0.24	0.16	1.36	0.22	6.15	771	q097
1	1.54	0.14	2	0.19	0.16	1.35	0.22	6.21	771	q103
1	-1.07	0.09	2	-2.31	0.16	1.24	0.18	6.82	771	q076
1	3.55	0.36	2	0.41	0.17	3.14	0.4	7.94	771	q090
1	2.78	0.25	2	0.38	0.17	2.4	0.3	7.99	771	q094

Table 3 shows race-related DIF in items in Part C (Multiple Choice Questions – Vocabulary):

**Table 3** DIF in items in Part C (Multiple Choice Questions – Vocabulary)

Person	DIF	DIF	Person	DIF	DIF	DIF	Joint			
Class	Measure	S.E.	Class	Measure	S.E.	Contrast	S.E.	t	d.f.	Name
1	-2.61	0.1	2	-1.46	0.15	-1.15	0.18	-6.33	763	q105
1	-2.2	0.09	2	-0.48	0.19	-1.71	0.21	-8.17	763	q112
1	-1.71	0.09	2	0.46	0.26	-2.16	0.27	-7.97	763	q114
1	-4.21	0.16	2	-2.84	0.15	-1.37	0.22	-6.22	763	q109
1	-3.67	0.13	2	-2.64	0.15	-1.03	0.2	-5.21	763	q107
1	-0.75	0.1	2	0.39	0.25	-1.14	0.27	-4.27	763	q136
1	-0.11	0.11	2	1.25	0.35	-1.36	0.37	-3.69	763	q140
1	-2	0.09	2	-1.48	0.15	-0.51	0.18	-2.89	763	q131
1	-1.11	0.09	2	-0.62	0.18	-0.49	0.2	-2.4	763	q116
1	-2.25	0.09	2	-1.97	0.15	-0.28	0.17	-1.61	763	q143

1	0.23	0.12	2	0.67	0.28	-0.44	0.3	-1.46	763	q146
1	-1.13	0.09	2	-0.92	0.17	-0.21	0.19	-1.09	763	q124
1	2.1	0.26	2	2.84	0.72	-0.75	0.76	-0.98	763	q115
1	-2.84	0.1	2	-2.66	0.15	-0.17	0.18	-0.97	763	q108
1	-3.25	0.12	2	-3.16	0.15	-0.08	0.19	-0.44	763	q132
1	0.63	0.14	2	0.67	0.28	-0.04	0.31	-0.13	763	q139
1	-2.96	0.11	2	-2.96	0.15	0	0.19	0	763	q113
1	2.42	0.31	2	2.42	0.59	0	0.66	0	763	q119
1	-1.43	0.09	2	-1.43	0.15	0	0.18	0	763	q121
1	-1	0.09	2	-1	0.17	0	0.19	0	763	q133
1	1.95	0.25	2	1.88	0.46	0.07	0.52	0.12	763	q137
1	3.74	0.58	2	3.55	1.01	0.18	1.16	0.16	763	q145
1	1.97	0.25	2	1.69	0.42	0.28	0.49	0.57	763	q123
1	1.65	0.21	2	1.13	0.33	0.52	0.4	1.31	763	q129
1	1.23	0.18	2	0.67	0.28	0.56	0.33	1.7	763	q141
1	1.33	0.19	2	0.67	0.28	0.66	0.33	1.98	763	q128
1	-1.02	0.09	2	-1.39	0.15	0.37	0.18	2.04	763	q130
1	6.02<	1.8	2	2.12	0.51	3.9	1.87	2.09	763	q127
1	4.14	0.71	2	2.12	0.51	2.02	0.88	2.31	763	q134
1	-1.95	0.09	2	-2.35	0.14	0.4	0.17	2.35	763	q106
1	3.45	0.5	2	1.69	0.42	1.76	0.66	2.67	763	q117
1	4.84	1	2	1.88	0.46	2.95	1.1	2.68	763	q120
1	-0.82	0.09	2	-1.39	0.15	0.57	0.18	3.12	763	q118
1	0.81	0.15	2	-0.04	0.22	0.84	0.26	3.2	763	q147
1	2.17	0.27	2	0.83	0.3	1.33	0.4	3.32	763	q148
1	1.14	0.17	2	0.06	0.22	1.08	0.28	3.84	763	q144
1	1.48	0.2	2	0.27	0.24	1.21	0.31	3.88	763	q142
1	0.35	0.13	2	-0.58	0.18	0.93	0.22	4.19	763	q138
1	3.45	0.5	2	0.93	0.31	2.52	0.59	4.28	763	q125
1	-1.69	0.09	2	-2.43	0.15	0.74	0.17	4.35	763	q110
1	2.1	0.26	2	0.39	0.25	1.7	0.36	4.7	763	q126
1	0.17	0.12	2	-1.17	0.16	1.33	0.2	6.66	763	q122
1	0.35	0.13	2	-1.32	0.16	1.67	0.2	8.27	763	q111
1	-1.07	0.09	2	-2.6	0.15	1.53	0.17	8.86	763	q135

It was found that quite a number of items in both Part B and Part C contained race-related DIF. But for this paper only a few items would be selected to be analysed. Items discussed in this paper found to be biased against the Malays are items q072, q105 and q109, whereas items q106, q110 and q111 are biased against the other ethnic groups. For this paper, only items q072, q105 and q109, which are biased against the Malays, are discussed in this paper.

## DISCUSSIONS ON ITEMS

Q072 is in Part B (Open-ended Question – Vocabulary) and q105 and q109 are in Part C (Multiple Choice Question – Vocabulary Skills). The questions are as the following, and the candidates are asked to give meaning that is closest to the words from a choice of words given:

**Table 4** Definition of ‘berbalah’ for item q072

Item	Definition/ Synonym from dictionaries/thesaurus
<b>q072. Berbalah (to dispute; to quarrel)</b>	<p><i>Tesaurus bahasa Melayu Dewan</i> (2005) lists many entries, including <i>berbantah</i> (to quarrel), <i>bertekak</i> (to argue; to quarrel), <i>bertengkar</i> (to argue or quarrel), <i>bertikai</i> (to differ in opinion; to quarrel; to be different; to be divergent), <i>bertelagah</i> (to argue), <i>bertelingkah</i> (to differ in opinion), <i>bersengketa</i> (to dispute; to quarrel), <i>berselisihfaham</i> (to disagree), <i>bergaduh</i> (to cause an uproar; to fight), <i>berkelahi</i> (to quarrel; to fight), <i>bertegangurat</i> (to be steadfast), <i>bertingkuh</i> (to quarrel), <i>berperangmulut</i> (a duel of words), <i>bertikamlidah</i> (an exchange of angry words).</p> <p><i>Kamus Dewan</i> (Fourth Edition) explains the meaning of “berbalah” as <i>berbantah</i> (argue; expostulate; argument) <i>bertengkar</i> (bicker; argue; wrangle) (<i>Dewan Bahasa dan Pustaka</i>, 2010).</p>

Item q072 is an open-ended question, where students have to provide an answer that has the closest meaning with the word ‘berbalah’. The Malay candidates may have been facing some difficulties answering this question correctly. It is possible that the term is rarely used in everyday speech which made them to be less able to give a precise meaning.

For item q105, it is a question where students were required to select the answer that has the closest meaning with the word ‘berangan’. The stem and the options of answer are as follow, with their meanings:

- *Penghibur* (entertainer)
- *Fikiran* (mind; subconscious)
- *Makanan* (food; sustenance)
- *Khayal* (dreamy; stoned)
- *Pokok* (tree)
- *Mimpi* (dream)

**Table 5** Definition of ‘Berangan’ for Item q105

Item	Definition/ Synonym
<b>q105. ‘Berangan’</b>	<p><i>Kamus Pelajar Edisi Kedua</i> (2008) defined ‘berangan’ as a species of tree or its fruit (Chestnuts; <i>Quercusoidocarpa</i>); <i>melamun</i> (dreaming away).</p> <p><i>Kamus Inggeris-Melayu Dewan</i> (2013) stated that <i>berangan</i> could mean dreaming away (<i>berangan-angan</i>); dreamy (<i>berangan-angan, berkhayal</i>)</p>

For item q105, the candidates were asked to provide a word that has the closest meaning to the word given from a choice provided. Item q105 is also a close-ended question, and it is biased against the Malay candidates. The possible explanation to why they cannot answer this question correctly is because of the stem itself, which is ‘Berangan’. It is possible that the stem is confusing, where the item writers intended the answer to be *berkhayal* (dreamy). But the students may think that the

word 'berangan' here actually means the chestnut tree (*pokokberangan*), thus giving *pokok* (tree) as the answer rather than what is expected by the examiners and item developers. *KamusPelajarEdisiKedua* (DewanBahasadanPustaka, 2010) gave the meaning of 'berangan' as a name of a species of tree or a fruit. On the other hand, it is also possible that the Malay candidates think of 'berangan' as chestnuts that are widely sold for consumption, and this caused them to select the option 'makanan' (food) as the answer. In this case, there may be Malay candidates confused the term 'berangan' with 'berangan-angan' (daydream) which can cause them to decide on 'mimpi' (dream) as the answer. *KamusPelajarEdisiKedua* (DewanBahasadanPustaka, 2010) defined 'berkhayal' as 'berangan-angan' and 'mimpi' as 'berangan-angan' or 'berkhayal'.

For item q109, the students were required to select the answer that has the closest meaning with the word 'kejam'. The correct answer for this item is 'bengis'. The stem and the options of answer are as follow, with their meanings:

q109Kejam (cruel)

- Celik(open for eyes; realize; aware)
- Bunuh(kill; murder)
- Belasah(to beat)
- Bengis(stern; angry; shrewish;)
- Keras(hard; loud and shrill; strong; firmly)
- Menderita(to suffer)

**Table 6** Definition and synonym of 'kejam' for Item q109

Item	Definition/ Synonym
q109. <b>Kejam (cruel)</b>	<p><i>KamusPelajarEdisiKedua</i> (DewanBahasadanPustaka 2010) the definition is as follows: "...berlaku<b>kejam</b>kpd (terhadap) (acting cruelly/ brutally on); berbuat<b>keras</b>kpd (acting severely on): <i>Merekadiusir, dikejamidanhartamerekadirampas</i> (They were deported, being cruelly acted on and their belongings dispossessed)".</p> <p>TesaurusBahasaMelayuDewan (2005) gives the word as synonymous with brutal as '1. <i>Zalim</i> (cruel; tyrannical): <i>bengis</i> (cruel, heartless, merciless), <i>ganas</i> (fierce), <i>tidakberperikemanusiaan</i> (inhuman), <i>tiadabelaskasih</i> (cruel), <i>butahati</i> (insensibility), <i>tidakberhatiperut</i> (cruel), <i>tidakbertimbang rasa, tidakadil</i> (unfair)'. Candidates may have selected 'bunuh' or 'belasah' as an answer to this question, because they see the use of 'kejam' (cruel) as in the following sentence:</p> <ol style="list-style-type: none"> <li>1. <i>Mangsatelahdibunuhdengan<b>kejam</b></i> (Victims were killed brutally).</li> <li>2. <i>Budakberkenaatelahdibelasahdengan<b>kejam</b></i> (The boy was brutally beaten).</li> </ol>

For q109 item, the candidate must find the word that is closest to the word 'kejam' (cruel). 'Kejam' is an adjective and in the *KamusPelajarEdisiKedua* (DewanBahasadanPustaka, 2010) the definition is as follows berikut "...berlaku**kejam**kepada(terhadap) (acting cruelly/ brutally on); berbuat**keras**kepada (acting severely on): *Merekadiusir, dikejamidanhartamerekadirampas* (They were deported, being cruelly acted on and their belongings dispossessed)". TesaurusBahasaMelayuDewan (2005) gives the word as synonymous with brutal as '1. *Zalim* (cruel; tyrannical): *bengis* (cruel, heartless, merciless), *ganas* (fierce), *tidakberperikemanusiaan* (inhuman),



*tiadabelaskasihan* (cruel), *butahati* (insensibility), *tidakberhatiperut* (cruel), *tidakbertimbang rasa, tidakadil* (unfair)'. Candidates may have selected 'bunuh' (killed) or 'belasah' (beaten) as an answer to this question, because they see the use of 'kejam' (cruel) as in the following sentence:

1. *Mangsatelahdibunuhdengankejam* (Victims were killed brutally).
2. *Budakberkenaatelahdibelasahdengankejam* (The boy was brutally beaten).

It is also possible that the candidates did not select 'bengis', which is the answer expected by the item writers. The possible explanation to this is because the candidates may think that 'bengis' means someone who is bad-tempered, stern or shrewish. The example of sentences that has the word 'bengis' in them are as follow:

1. *Sititidaktahanlagidengansifatbengissuaminya* (Siti could no longer put up with her husband's shrewishness).
2. *Malik adalahseorangbapa yang bengisdantidakpenyayang* (Malik is a stern and unloving father).

Banks (2006) has given an explanation on how culture can influence the choice of answers in multiple choice questions by candidates, where one would be more inclined to choose an answer that has characteristics similar to their cultural backgrounds even if that answer is incorrect, or as mentioned by Veale and Foreman (1983) where they said "...different cultural groups may be attracted to different response options on multiple-choice tests, especially if the option contains culturally relevant stimuli that draw the groups' attention."

Similar opinion was voiced by Davies (2006) regarding one's cognition and culture, where he believed that how a person thinks is strongly influenced by several cultural aspects, such as the mother tongue and cultural background and he believed that the differences in how Western and Asian people think exist is due to their cultural factors. The same idea could be applied to the Malays, Chinese, Indians and other races/ethnic groups: these groups have their own mother tongue and have different ways in how they live their everyday lives. A study conducted by Beiser and Gotoweic (2000), mentioned by Lynn, Backhoff and Contreras (2005), was done where they distributed an intelligence-measuring instrument called WISC-R, that is written in English, to Native American children in parts of Canada. It was found that their verbal performance in the test was significantly lower than the average score obtained by children of other ethnic group besides native American, where they only managed to obtain 79.3, and these researchers were of the opinion that this happened because English is only their second language. Therefore developers of tests should always take into consideration about elements of culture and language every time they construct a new test paper (Fox and Chang, 2007).

From the discussion above, it can be concluded that in this intelligence test, items do function differently to students who have the same or different ethnic groups with the test developer. There is a possibility that they have different interpretations of certain BahasaMelayu words due to some factors even if the test developers and candidates are from the same ethnic group. Candidates' understanding or definitions of certain words may differ from what the test makers had in mind, and from here it can be seen that items that requires students to give definitions to words or select words that have the closest meaning to the words given is highly potential to be biased to students.

In this study, it can be seen that language bias occurred not only to the ethnic group other than Malay, but it also occurred to the Malays, even when the instrument is written in BahasaMelayu, a language that is considered as the mother tongue for most of the Malays. It was seen that the Malay students were unable to answer some of the items correctly. There are several possible explanations to this matter. One of the possibilities is the influence of the dialects that the Malay students use in their daily conversations with their families or friends who come from the same geographical area as they do. The dialects may not use or have different meanings of words in BahasaMelayu. Another possible explanation is the Malay students were not familiar to some of the words in this test, due to lack of exposure (not being taught in school, for example) or not using the word in their daily conversation.

Therefore, it is hoped that test developers and also stakeholders be more aware of the impact of one's culture and mother tongue, and also dialect, when answering a test or instrument that is written in a language, in this case here in the national language, which could be different from one's mother tongue. It is suggested that extra care should be taken in developing the instruments. Another

suggestion is to have different versions of the same test written in different languages be made so that this type of bias will be thoroughly reduced, and each student will have the same chances to perform well in a test.

\*Note: This article was written when the researcher was a graduate student in Universiti Kebangsaan Malaysia, Selangor, Malaysia. The researcher is now working in Universiti Pendidikan Sultan Idris, Perak, Malaysia.

## REFERENCES

- Banks, K. (2006). A comprehensive framework for evaluating hypotheses about cultural bias in educational testing. *Applied Measurement in Education*, 19 (2), 115-132.
- Beiser, M. & Gotowiec, A. (2000) Accounting for native/non-native differences in IQ scores. *Psychology in the Schools*, 37, 237-252.
- Brooks, K., & Karathanos, K.. (2009). Building on the Cultural and Linguistic Capital of English Learner (EL) Students. (online journal) *Multicultural Education*, 16(4), 47-51.  
<http://proquest.umi.com/pqdweb?index=0&did=1878338371&SrchMode=2&sid=4&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1268645253&clientId=60729>.
- Comesaña M., Perea, M., Piñeiro, A. & Isabel Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners (online journal). *Journal of Experimental Child Psychology* 104, 22-33. <http://www.uv.es/~mperea/MontseJECp.pdf>.
- Davies, W. M. (2007). Cognitive contours: recent work on cross-cultural psychology and its relevance to education. *Studies in Philosophy and Education*, 26, 13-42.  
<http://www.springerlink.com/content/a02127833056787j/fulltext.pdf>.
- Dewan Bahasa dan Pustaka. (2005). *Tesaurus Bahasa Melayu Dewan* (edisi baharu). Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Dewan Bahasa dan Pustaka. (2010). *Carian PRPM@DBP Malaysia* (kamus online). <http://prpm.dbp.gov.my/>.
- Ferré, P., Sánchez-Casas, R., & Guasch, M. (2006). Can a horse be a donkey? Semantic and form interference effects in translation recognition in early and late proficient and nonproficient Spanish-Catalan bilinguals (online journal). *Language Learning*, 56(4), 571-608.  
<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=5&hid=112&sid=8cf25cc4-8c28-42ed-a248-41254be63172%40sessionmgr114>.
- Fox, J. & Cheng, Liying. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers (jurnal online). *Assessment in Education* 14 (1), 9-26. <http://web.ebscohost.com/ehost/pdf?vid=12&hid=4&sid=64db9371-8ee6-4eee-a8da0c145abd1685%40sessionmgr113>.
- Gray, R., McCallum, R., & Bain, S. 2009. Language-Reduced Screening for Giftedness (online journal). *Journal for the Education of the Gifted* 33(1), 38-64.  
<http://web.ebscohost.com/ehost/pdf?vid=13&hid=111&sid=64db9371-8ee6-4eee-a8da0c145abd1685%40sessionmgr11>.
- Huang, J. (2009). Factors Affecting the Assessment of ESL Students' Writing (jurnal online). *International Journal of Applied Educational Studies* 5(1), 1-17.  
<http://web.ebscohost.com/ehost/pdf?vid=19&hid=112&sid=2c5bf1fc-ea39-4724-b1dcd2dfed7ae2b4%40sessionmgr114>.
- Hye-Sook, Park, Pearson, P. D. & Reckase, M. D. (2005). Assessing the effect of cohort, gender, and race on Differential Item Functioning (DIF) in an adaptive test designed for multi-aged groups. *Reading Psychology* 26 (1), 81-101.  
<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=1407b1df-10c2-47ab-99c3-ddc30b715afe%40sessionmgr114&vid=8&hid=115>.
- Lynn, R., Backhoff, E. & Contreras, L. A. (2005). Ethnic and racial differences on the Standard Progressive Matrices in Mexico (jurnal online). *Journal of Biosocial Science* 37, 107-113.  
<http://proquest.umi.com/pqdweb?index=0&did=793460861&SrchMode=2&sid=1&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1266822409&clientId=60729>.
- Marbley, A. F., Bonner, F. & Berg, R. (2008). Measurement and assessment: Conversations with professional people of color in the field of education (online journal). *Multicultural Education* 16(1), 12-20.  
[http://vnweb.hwwilsonweb.com/hww/results/results\\_single.jhtml;hwwilsonid=512AQGASO55EJQA3DILSFGGADUNGIIV0](http://vnweb.hwwilsonweb.com/hww/results/results_single.jhtml;hwwilsonid=512AQGASO55EJQA3DILSFGGADUNGIIV0).
- Mendes-Barnett, S. & Ericikan, K. (2006). Examining sources of gender DIF in Mathematics assessments using a Confirmatory Multidimensional Model Approach. *Applied Measurement in Education* 19 (4), 289-304.

- <http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=1407b1df-10c2-47ab-99c3-ddc30b715afe%40sessionmgr114&vid=12&hid=115>.
- NoriahMohd. Ishak, RosadahAbd, Majid danSiti Fatimah MohdYassin (2009). *PERMATApintar: Pengalaman UKM*.Bangi: PusatPERMATApintar Negara, UniversitiKebangsaan Malaysia.
- Ortiz, S. O., & Ochoa, S. H. (2005b). Conceptual measurement and methodological issues in cognitive assessment of culturally and linguistically diverse individuals. In R. L. Rhodes, S.H. Ochoa & S. O. Ortiz (Eds.), *Assessing Culturally and Linguistically Diverse Students: A Practical Guide* (pp. 153–167). New York: Guilford Press.
- Rudas, T., &Zwick, R. (1997). Estimating the importance of differential item functioning. *Journal of Educational and Behavioral Statistics*, 22(1), 31–45.
- Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze test items: a multiple data source approach. *Language Testing*, 17(1), 85-114.
- Schon, J., Shaftel, J. & Markham, P. (2008). Contemporary issues in the assessment of culturally and linguistically diverse learners (online journal). *Journal of Applied School Psychology* 24(2), 163-189. <http://web.ebscohost.com/ehost/pdf?vid=25&hid=112&sid=64db9371-8ee6-4eee-a8da-0c145abd1685%40sessionmgr113>.
- Schwieter, J. &Sunderman, G. 2009. Concept selection and developmental effects in bilingual speech production (online journal). *Language Learning* 59(4): 897-927. <http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=2&hid=112&sid=c3b96cc6-c520-41a8-9008-8084e8384d82%40sessionmgr110>.
- SitiRahayahAriffin (2008). *Inovasi dalam Pengukuran dan Penilaian Pendidikan*. Bangi, Selangor: Fakulti Pendidikan, Universiti Kebangsaan Malaysia.
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English Language learners. (online journal). *Educational Researcher*, 37(4), 189-199. [http://www.aera.net/uploadedFiles/Publications/Journals/Educational\\_Researcher/3704/ER%20May%20Feature%201.pdf](http://www.aera.net/uploadedFiles/Publications/Journals/Educational_Researcher/3704/ER%20May%20Feature%201.pdf).
- Sunderman, G. & Kroll, J. F. (2006). First language activation during second language lexical processing: An Investigation of Lexical Form, Meaning, and Grammatical Class (online journal). *Studies in Second Language Acquisition* 28, 387-422. <http://mailer.fsu.edu/~gsunderm/Research/ssl%202006.pdf>.
- Tonzar, C., Lotto, L., & Job, R. (2009). L2 Vocabulary acquisition in children: Effects of learning method and cognate status (online journal). *Language Learning*, 59(3), 623-646. <http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=2&hid=112&sid=7a83f10e-bb64-4ca4-b763-7feae53eed53%40sessionmgr114>.