KEYNOTE SPEAKER

Adjunct. Prof. Dato' Dr. Ghazali bin Dato' Mohd. Yusoff, *DPTJ., DMM., DNS., PJK.*
Executive Chairman, Nusantara Technologies Sdn. Bhd Malaysia

Prof. Baharuddin bin Aris, PhD
Dean of Faculty of Education, University Teknologi Malaysia

Prof. Dr. Hadi Nur
Ibnu Sina Institute for Fundamental Science Studies, University Technologi Malaysia

Prof. Dr. Edy Suandi Hamid, M.Ec
Rector Universitas Islam Indonesia

Prof. Dr. Mashadi Said
Head English Dept.Gunadarma University Indonesia

Prof. Dr. Hamzah Upu, M.Ed.
Dean Faculty of Mathematics & Science, UNM Indonesia

Prof. Hamdan Juhannis, PhD
Universitas Islam Negeri Alauiddin Makassar ndonesia

Dr. Muhammad Yaumi, MA.
Director of Learning Center UIN Alauddin—Makassar Indonesia

ISBN 978-967-12214-2-6

The 1st Academic Symposium on Integrating Knowledge (The 1st ASIK)

PROCEEDINGS

"Integrating Knowledge with Science and Religion."
The 20th-21st of June 2014
Editor in Chief:
Prof. Dr. Hadi Nur

Ibnu Sina Institutes for Fundamental Science Studies
Universiti Teknologi Malaysia
2014

The 1st Academic Symposium on Integrating Knowledge
(The 1st ASIK)

PROCEEDINGS

"Integrating Knowledge with Science and Religion."

The 20th-21st of June 2014

Editor in Chief:

Prof. Dr. Hadi Nur

Ibnu Sina Institutes for Fundamental Science Studies
Universiti Teknologi Malaysia
2014

Published by:

Ibnu Sina Institutes for Fundamental Science Studies
Universiti Teknologi Malaysia
Skudai, Johor Bahru, Johor 81300

All the papers in this proceeding are presented at the 1st Academic Symposium on Integrating Knowledge on 20th – 21st June 2014 in Universitas Islam Negeri Alauddin Makassar, Indonesia.

Perpustakaan Negara Malaysia Data Pengkatalogan dalam Penerbitan
Cataloguing in – Publication Data

Hadi Nur
Proceeding of The 1st Academic Symposium on Integrating Knowledge (ASIK) 2014
Hadi Nur at al
ISBN: 978-967-12214-2-6

Design by:
Academic Symposium on Integrating Knowledge

PREFACE

Bismillahirrahmanirrahim.

In the name of God, the Most Gracious, the Most Merciful

Assalamualaikum warahmatullahi wabarakatuh.

Academic Symposium on Integrating Knowledge (ASIK) has successfully organized the 1st Academic Symposium on Integrating Knowledge (The 1st ASIK) on 20th – 21st June 2014 in Universitas Islam Negeri Alauddin Makassar, Indonesia. The theme of this International Symposium is "Integrating Knowledge with Science and Religion". The 1st ASIK covers many disciplines in education, science, technology, language, social sciences, health, and religion involved in the research.

This International Symposium is expected to present prospect for all academicians, scientists, and researchers to encourage, impart and share ideas in promoting research network among interdisciplinary field of studies. There are more than 50 papers presented by academicians, scientists, and researchers from Asia.

Finally, I would like to extend my gratitude to all those who are involved in the publication of the proceedings of the 1st ASIK 2014. It is hoped that this proceeding will contribute to the development on integrating knowledge with science and religion particularly in Asia and among the international academicians, scientists, and researchers in general.

Editor in Chief:
Prof. Dr. Hadi Nur—Ibnu Sina Institutes for Fundamental Science Studies

FOREWORDS

Bismillahirrahmanirrahim.

In the name of God, the Most Gracious, the Most Merciful

Assalamualaikum warahmatullahi wabarakatuh.

I would like to express praises and gratitude to Almighty Allah because it is only by His permission that I am able to convey my forewords in the proceedings of the 1st Academic Symposium on Integrating Knowledge (The 1st ASIK) 20-14 organized by Academic Symposium on Integrating Knowledge (ASIK) in collaboration with Learning Center UIN Alauddin Makassar, Faculty of Education UTM, Ibnu Sina Institute for Fundamental Science UTM.

I would like to take this opportunity to congratulate and compliment the committee members of this International Symposium who have consistently work very hard to produce this proceedings.

The publication of this proceeding is expected to benefit as many parties as possible and become a reference for those who wish to gain further information on integrating knowledge with science and religion.

Finally, I hope that through such initiatives of knowledge integration with science and religion event and publication of symposium's proceeding, a higher quality of research and publication can be multiplied in the future.

Best regards,

Prof. Dr. Hadi Nur
Ibnu Sina Institutes for Fundamental Science Studies
Universiti Teknologi Malaysia

TECHNICAL COMMITTEE

Patron
Prof. Dr. Hadi Nur - Head of Catalytic Science and Technology Research Group, Universiti Teknologi malaysia.

Editors:

- Prof. Dr. Hadi Nur – UTM, Malaysia
- Assoc. Prof. Othman Che Puan - UTM, Malaysia
- Prof. Dr. Amran Md Rasli - UTM, Malaysia
- Prof. Dr. H. Azhar Arsyad, MA - UIN Alauddin, Indonesia
- Prof. Hamdan Juhannis, PhD. - UIN Alauddin, Indonesia
- Assoc. Prof. Dr. Yusof Bin Boon - UTM, Malaysia
- Prof. Dr. Hamzah Upu – UNM, Indonesia
- Dr. Hamimah Abu Naim – UTM, Malaysia
- Dr. Ahmad Johari B Sihes - UTM, Malaysia
- Dr. Hashim Fauzy bin Yaacob - UTM, Malaysia
- Prof. Dr. Baso Jabu, M.Hum. – UNM, Indonesia
- Dr. Sukardi Weda, S.S., M.Hum., M.Pd., M.Si., M.M. - UNM, Indonesia
- Murni Mahmud, PhD. - UNM, Indonesia
- Dr. Muhammad Yaumi, M.Hum, MA. – UIN Alauddin, Indonesia
- Prof. Dr. Mashadi Said - Universitas Gunadarma, Indonesia
- M. Furqaan Naiem, MD., M.Sc., PhD. - Universitas Hasanuddin, Indonesia
- Dr. Muslimin Kara, M.Ag. - UIN Alauddin, Makassar, Indonesia.
- Dr. Nurhayati, S.Kep.,NM., MARS, UIN Alauddin, Makassar - Indonesia
- Assoc Prof Dr.Ismail Said - Universiti Teknologi Malaysia
- Dr. Rustan Santaria, M.Hum - STAIN Palopo, Indonesia
- Dr. Irwan Said, M.Si. Universitas Tadulako- Indonesia
- Asoc. Prof. Dr. Rosman Md Yusoff, Universiti Teknologi Malaysia
- Dr. Moch. Azhar bin Abd Hamid, Universiti Teknologi Malaysia
- Dr. Hj. Kassim bin Thukiman, Universiti Teknologi Malaysia

Chairman
Dr. Muhammad Yaumi, MA., M.Hum.

Deputy Chair
Hasbullah Said

Secretary
Andi Anto Patak

Treasurer
Serly Towolioe
Lita Limpo

Divisions

- Secretariat:
  Chairperson: Andi Anto Patak – Universiti Teknologi Malaysia (Malaysia)
  Members:
  1. Ratna Sari Devi – UIN Alauddin Makassar (Indonesia)
  2. Dwi Oktaviani – UIN Alauddin Makassar (Indonesia)
  3. Abdul Muarif – UIN Alauddin Makassar (Indonesia)
  4. Ahmiranil Khaerat – UIN Alauddin Makassar (Indonesia)
  5. Anthye Aurora Hamzah – UIN Alauddin Makassar (Indonesia)

6. Andi Muhammad Syukri – UIN Alauddin Makassar (Indonesia)
7. Azhar El Marosyi – UIN Alauddin Makassar (Indonesia)
8. Andi Rusdam – STIEM Bongaya (Indonesia)
9. Dr. Hj. Sosiawan Ma'mun – Universitas 17 Agustus (Indonesia)
10. Dr. Rustan Santaria – STAIN Palopo (Indonesia)
11. Dr. Rusdiana Junaid – Universitas Cokroaminoto Palopo (Indonesia)

- Paper Proceedings:
  Chairperson: Hj. Dahlia Nur
  Members:
  1. Siti Dahlia Said
  2. Muhammad Natsir Ede

- Promotions:
  Chairperson: Hurriah Ali Hassan
  Members:
  1. Alham Syahruna

- Symposium Program:
  Chairperson: Bernadeth Tongli
  Members:
  1. Arti Manikam
  2. Andi Parianti

- Public Relations
  Chairperson: Erwin Akib
  Members:
  1. Bakry Liwang

- Registration
  Chairperson: Andiana
  Members:
  1. Widya Puspita Nur
  2. Nurhayati
  3. Muhammad Rusdi
  4. Nurbaya

- Logistics
  Chairperson: Andi Ernawati
  Members:
  1. Usman
  2. Musdawati
  3. St. Hamsina

- Transportation and Accommodation
  Chairperson: Muhammad Asri

Symposium Advisor:
Nur Asik
Rahimuddin Samad, ST.,MT.,Ph.D
Drs.Syahruddin, S.Pd.,M.Pd.,Ph.D

General Advisory:

1. Religion Minister of Indonesia
2. Indonesian Ambasador of Malaysia, Kuala Lumpur
3. Indonesian Consul General in Johor Bahru, Malaysia
4. Prof. Rusdi, M.Ed., PhD. (Education Attache at Indonesian Embassy Kuala Lumpur)
5. Djudjur Hutagalung (Information, Social and Culture Consult at Indonesian Consulate General Johor Bahru)

20 June, 2014

Chairman	Secretary
Dr. Muhammad Yaumi, M.A., M.Hum.	**Andi Anto Patak**
Director of Learning Center	Measurement and Evaluation
Universitas Islam Negeri Alauddin	Faculty of Education, UTM

CONTENT PAGE

ANALYSING ITEMS USING RASCH MODEL TO ASSES STUDENTS' PERFORMANCE IN TIMMS 2007

Muhammad Tahir
State University of Makassar

ABSTRACK

This study uses the Rasch model to assess students' performance of large scale survey conducted by TIMMS 2007. Data analyzed to find out the fit and misfit of questionnaire. As a secondary data, this study used Book 7 TIMMS 2007 with Australia data set. The survey was administered to 285 students of Australia who participated in TIMMS 2007 as the base line assessment for math test. The Quest software was used to assess model fit, item difficulty and bias gender. The result of data indicates that all items are fitted the Rasch model and there is no item lies outside the threshold range of 0.77 to 1.30. In addition, the study also found that there is no evidence of gender bias or systematic guessing from the students.

KEYWORDS: Item Fit, Rasch model, item difficulty and Gender bias

1 INTRODUCTION

Like other countries Australia has been participated in large scale international survey such as TIMMS, PIRLS and PISA. According to TIMMS 2007 technical report, students were participated in TIMMS at grade eight was 4069 while at fourth grade was 4109.
The basic sample design used in TIMSS 2007 is known as a two-stage stratified cluster design, with the first stage consisting of a sample of schools, and the second stage having a sample of intact classrooms (usually mathematics classes) from the target grades in the sampled schools.
For countries participating in TIMSS 2007, school stratification was used to enhance the precision of the survey results. Australia divided its sampling frame into eight states and territories to ensure equal precision in the survey results between states and between the two territories. Australia employed implicit stratification by school type (Government, Catholic, Independent) and school location (metropolitan area or elsewhere) within each explicit stratum. In addition, Since 1999 Australia only has tested fourth grade students because there will be a transition time between seven and eight grade when the survey was conducted.
TIMSS 2007 covers two main subjects, math and science test. Test is constructed into multiple choice test and questionnaire. The question applied in this test discussing about content and cognitive domain.
Traditionally classical test analysis suggests that item fit in any measurement is a Cronbach alpha .80 and Rasch model requires three items fitted a coherence scale quite well (Curtis 2004, p. 124). Furthermore Rasch model seeks for assumption namely unidimensionality where all items are expected to measure single latent trait or construct (Fox, 1999 p. 340). He then articulates that the model also assumes that the students are not guessing the answer and the item should well discriminate between low achievers and high achievers and lastly, students are absolutely independent (p. 341)

Fox (1999) argued that Rasch model functions are to help assessor and test developer to estimate the requirement measurement and help them to tailor the result of data with their original purposes. This paper explores the application of Rasch model to analyze data derived from TIMMS 2007 that measures the students' achievement in math. The literature review then will support the analysis of data from multiple choice items. Emphasis is then placed to how Rasch calibration implemented to scale calibration and analysis. The advantages of Rasch model is not only an interval scale, but also natural metric, with the scale unit referred to as logit (Keeves & Master, 1999).

Researches on item calibration always invite researchers to investigate. A study on literacy and numeracy is carried out by Hungi (1997 & 2003). Hungi finds that the growth of literacy and numeracy achievement between year 3 and year 5 in South Australia is about 0.50 logits per year. Likewise Walstead & Robson (1997) argue that female students of the equal ability do not perform as well as males on multiple choice tests. Further they mentioned that female students were difficult if the test contains numerical computational, spatial or high thinking skills (reasoning skills). Keats & Lord (cited in Barret, 2001 p.124) also articulates that students' performance does not show the true ability of students as the result of guessing. Furthermore Barret (2001) in his study about differential item functioning also found a little evidence of gender bias or systematic guessing and many questions did not adequately discriminate students' ability.

2 RESEARCH QUESTION AND STUDY PURPOSES

This study analyses the 16 questions that were adapted from TIMMS 2007 for Australian data set. The study purpose is to find the answer to three research questions. First, Do these question discriminate well the students in the basis of their ability? Second, Do these question in TIMMS 2007 reflecting gender bias exclusively? Third, Do the items fit the Rasch model?

3 STUDY PURPOSES

This study is expected to give information about the item measurement in TIMMS 2007 particularly math achievement. By measuring the item difficulty and students response to the question, it will then give information whether all items (16 items) are well fitted model or not. Regarding with gender variable, then this study also investigates the gender bias of students regarding with math achievement. It is also expected that the result of item analysis can give significant information for those who intends to do the same research in the future.

4 PERTINENTS IDEAS

In measurement, there are two elements which are usually used to determine how well data met the requirement of model. Rasch analysis measures and reports statistic data in terms of infit and outfit which is called two chi-square ratio (Wright, 1984; Wrigth & Master, 1981 Fox, 1999).
Historically, Rasch model is first used in educational measurement to solve the problem in scoring item. Thompson argues (2003) that the main focus of Rasch model is to implement unidimensional scale that functions as measuring students performance on test and judging item difficulty on each items difficulty in question in the test. Then this scale is marked in logits. This logits provide information regarding with item calibration
In marking process, Linacre cited in Thompson (2007)) argued that there are three should be involved in judging system such as each rater, each candidate and each assessment item. According to Bond & Fox (2007, P. 238) the term outfit is dealing with conventional sum of squared standardized residuals while infit means an information-weighted sum. They articulated that infit and outfit is a part of report that indicates the chi square devided by their degrees of freedom. In addition to these, we expect the ratio between infit and outfit is +1 or close to 0 with positive and negative infinity (Bond & Fox, 2007, Blackman et al., 2006, and Keeves & Master, 1999,).

Joyce & Yates (2007) in his study about self-concept analysis argued that data gathered from measurement should meet the fit criteria and allow three elements of requirements: "1). Equal difference should rely on two set item difficulties and two measurement of scale, 2) the omitting process should not affected the item scale and 3) There is a free intervention both from students or opinions after deciding final grade" (p. 234).
Barrett (2001) reveals in his study that multiple choice is a latent trait which based on the assumption that the performance of the students is taken by underlying or trait (p. 124). According to Barrett (2001, p. 124) another function of Rasch model is to determine students correct answer on a multiple choice test in terms of two parameters, one refers to item difficulty and the other deals with students ability. In addition he claims that high level students may have chance to answer particular question

than low level students. Otherwise a students with different range of ability may also have chance to answer less difficult question than on a more difficult question. The use Rasch model and item respond theory in measurement can mediate the ability of students to answer question in a test and the degree of item difficulty.

Referring to students' response, Masters and Keeves (1999, p. 25) clearly describe that " if the ability the person exceeded the difficulty of the item, then response would be expected to be correct or favorable and if the person were less than the difficulty level of the item, then the response would be expected to be incorrect or favorable. According to Wrigth & Master cited in Curtis (2004, p. 126) measurement involves four process which allows any researcher to be investigated, such as one dimensional abstraction, comparing people and test result; a linear magnitude inherent in positioning objects along line; and a unit determined by a process which can be repeated without modification on the variable.

The computer application provides information about Rasch model analysis which generates outcome of test whether the items fit the model, refined, or removed from the test (Keeves & Lawson, 2002). Regarding with the applicability of data to the model, Rasch suggested using chi-square fit statistic (Linacre & Wright, 1994). They argued that chi-square is a means squares, where it divided by their degree of freedom. The chi-square is commonly used as OUTFIT and INFIT. Another calibration that is discussed in this study is OUTFIT. OUTFIT is recognized as unexpected outlier, off-target, less information response while INFIT on the other hand, refers to on target-observation, unexpected inlaying pattern among informative and so inliers-sensitive (Linacre & Wright, 1994 p. 4). Accordingly, Adam & Khoo (1997) define that the value of item fit fall in a range between 0.77 to 1.31, conversely when item fit statistic the INFIT is above 1.3, then the item is under fit.

5 METHOD

The data were analyzed using Rasch (1980) measurement, which can measure the students' achievement and item difficulties with the same scale. Rasch calibration is used to examine the data form fit and outfit concept. Book 7 was chosen as secondary data reflecting Australian students' performance. This study only focuses on students' math achievement as a basis for measurement.

Baseline data of participants obtained from TIMMS 2007 however the researcher limit the sample data which only consisted 285 students, however, this sample was originally 4069 total participants. Those students should answer all multiple choice items that have designed prior to study. This survey research only consisted 30 items (1-16 was math questions and 17-30 was science questions). Data was analyzed using software QUEST (Adam & Khoo, 1997). According Keeves & Master (1999) appropriate sample survey that is used in QUEST is 50 person, then for dichotomous items at least 100 person and for trichotomus attitude scales at least 150 person.

Rash model provides important information about the characteristic of item and students response on survey. Students respond the survey with four options. Option a, b, c, and d. each option is given score 1 for correct answer and 0 for incorrect answer. Each answer of item is examined to determine the item difficulty and then it calibrated into infit and outfit model measurement. Before the data enters the calibration the researcher needs to recode data to ascertain that the data was valid or not.

6 RESULTS

This research used two main categories of goodness of fit indices, *INFIT* and *OUTFIT*. Outfit refers to unweighted or the outfit means square index and infit on the other hand is weighted index or infit means square (Blackman, et al. 2006). They articulated that infit and outfit is derived from chi-square ratios which generate result from discrepancies between predicted and observed score. The result of fit ranges from positive and negative and it is around zero to one and depending on how the observed values create variation in response rather than we expect (Blackman, et al. 2006 p. 249).

Figure 1 provides, for example the item -analysis map for the math test. In this figure 16 item had calibrated on the scale which is located on the top left side of the scale to item analysis (infit and outfit) were on the scale located on left on the scale).

```
-----------------------------------------------------------------------------------------------------
Item Fit                                                                    27/10/2009 12:10
all on mathematics (N = 285 L = 16 Probability Level=0.50)
-----------------------------------------------------------------------------------------------------
INFIT
 MNSQ      0.56      0.63      0.71      0.83      1.00      1.20      1.40      1.60      1.80
--------------+---------+---------+---------+---------+---------+---------+---------+---------+----
   1 M022043                          *     |
   2 M022049                                |*
   3 M022050                          *     |
   4 M022057                                |           *
   5 M022257                             *  |
   6 M022062                                |           *
   7 M022066                                |*
   8 M042003                                |     *
   9 M042079                                |  *
  10 M042055                           *    |
  11 M042039                           *    |
  12 M042199                           *    |
  13 M042265                     *          |
  14 M042137                                |*
  15 M042148                        *       |
  16 M042254                                *
```

Figure 1.the item fit of math achievement

In figure 1 infit means square value showed all items is fit because their infit means square values were not greater than 1.30 or less than 0.77 (Fox and Bond, 2007). In other words, no item lies outside of the threshold range of 0.77 to 1.30.

Those items do not need to be refined or discarded from the tests and should remain in the examination. This also suggests that probably Australian students understand the concept of math test offered in TIMMS 2007. Therefore, the results of item fit may answer the research question 1 where all items are fit the model.

In the kidmap, the items are located in the left side of map are those who successfully answered the items, and the items on the right side provides information about the child did not complete the items successfully.

```
Trial1 2009_10_20
------------------------------- K  I  D  M  A  P--------------------------------
 Candidate: 50                                ability:  -1.32
 group:     all                                fit:      0.94
 scale:     mathematics                          % score:  25.00

------------Harder Achieved ----------------------Harder Not Achieved ----------
                                | |
                                | |   1(2)
                                | |
                                | |
                                | |   2(2)
                                | |
                                | |   6(4)
                                | |
                                | |
                                | |
```

```
                    | |   9(4)
                    | |
                    | |
                    | |  13(4)
                    | |  15(4)
                 8  | |   5(2)
                    | |  12(2)
                    | |   7(1)
                    | |   4(3)
.........................................
                10  | |
                    | |  14(3)
                11  | |
                16  | |
                    |XXX|
                    | |
                    | |   3(4)
                    | |
                    | |
------------Easier Achieved ----------------------Easier Not Achieved ----------
=================================================================
```

Figure 2. KIDMAP for candidate 50 showing good fit to the model (infit mean square =.94)

In Figure 2, the child's ability (candidate 50) estimates of -1.32 logits is plotted in the center column with the dotted line on the left indicating the upper bound of the ability estimate (ability estimates plus one standard error: $b_n + s_n$) and the dotted line on the right indicates the ability of students in its lower bound (ability estimates minus one standard error: $b_n - s_n$). Furthermore this KIDMAP also showed that the fit index (the fit means square value 0.94) indicates a pattern of performance (candidate 50) that closely related to the Rasch predicted model based on the child ability, the expected fit value is +1.0. This figure also depicts that item 8 is correct despite a less than 50% probability of success. The 0.94 infit mean square on the figure 2 of KIDMAP shows a performance close to that predicted by Rasch model (1.0) and the KIDMAP describes that visually. According to the Rasch model expected candidate 50 was unsuccessful because he guessed (item 8) which is far beyond his knowledge.

Meanwhile in Figure 3, the infit mean square value of -1.73 for students 150 in math achievement isc not fit the Rasch model for an estimated ability -1.73. It can also be detected from figure 3 that the unexpected responses made by the candidate 150 are item 10, 7, and 8. However, these scores may be the source of unfit, estimated at 73% (1.73-1.0 x 100%) more variation than the Rasch model.

Candidate 150 makes the pattern response unpredictably where he made unexpected correct answer: item 8, 7, and 10. This surely convinces us that candidate 150 has something missing knowledge to those items and they are only making guessing that cause misfit on results.

```
Trial1 2009_10_20
------------------------------- K  I  D  M  A  P--------------------------------
 Candidate: 150                                ability:  -1.73
 group:     all                               fit:       1.19
 scale:     mathematics                          % score:  18.75

------------Harder Achieved ----------------------Harder Not Achieved ----------
                    | |
                    | |   1(1)
                    | |
                    | |
                    | |   2(2)
                    | |
```

```
                        |  |   6(2)
                        |  |
                        |  |
                        |  |
                        |  |   9(4)
                        |  |
                        |  |
                        |  |  13(3)
                        |  |  15(2)
                    8   |  |   5(2)
                        |  |  12(4)
                    7   |  |
                        |  |   4(1)
                        |  |
                   10   |  |
.......................................  14(2)
                        |  |  11(1)
                        |  |  16(3)
                        |  |
                        |  |
                        |  |   3(1)
                        |XXX|
                        |  |
 .......................................
                        |  |
                        |  |
                        |  |
                        |  |
------------Easier Achieved ----------------------Easier Not Achieved ----------
```

Figure 3. KIDMAP for candidate 150 showing inadequate fit to the model (infit mean square =1.19)

When the data shows fit model, then the t values have a mean near 0 and standard deviation near 1. However, Bond & Fox (2007) articulate that if the mean values have +2 or less than -2 it can be interpreted as less compatibility with the model expected (p >.05). Figure 4 below shows the result of acceptability of t values which conforms the data.

```
Trial1 2009_10_20
---------------------------------------------------------------------------
Item Estimates (Thresholds)                                  27/10/2009 12:10
all on mathematics (N = 285 L = 16 Probability Level=0.50)
--------------------------------------------------------------------------- Summary of item Estimates
Mean                        0.00
SD                          1.01
SD (adjusted)               1.00
Reliability of estimate     0.98
 Fit Statistics
===============
Infit Mean Square       Outfit Mean Square

   Mean     1.00            Mean   1.02
   SD       0.15            SD     0.28
   Infit t                  Outfit t
   Mean    -0.16            Mean   0.01
   SD       2.44            SD     1.87
   0 items with zero scores
```

0 items with perfect scores

Figure 4 summary of item estimates and fit statistics.

Figure 4 shows the summary of item estimates on mathematics multiple choice items. The result of fit statistic indicates that infit mean square and outfit mean square is 1.00. Figure 4 also reports that infit *t* and outfit *t* close to the expected zero values; this means that the model has good fit and it can answer the research question three.
Quest also be used to measure whether bias gender exists between the students. All group students can be tested for bias. This study only investigates on male and female students.

```
Trial1 2009_Nov15
-----------------------------------------------------------------------------------------------------
Comparison of Item estimates for groups girls and boys on the mathematics scale
L = 16     order = input                                                    15/11/2009 22:41
-----------------------------------------------------------------------------------------------------

                    Easier for girls              Easier for boys

     -3          -2          -1          0          1          2          +3
 -------+--------------+-------------+--------------+--------------+-------------+--------------
 item 1          .                  |   *            .
 item 2          .                 *|                .
 item 3          .                 *|                .
 item 4          .                  |      *         .
 item 5          .        *         |                .
 item 6          .                  |  *             .
 item 7          .                  |            *   .
 item 8          .                  |      *         .
 item 9          .                  |           *    .
 item 10          .             *   |                .
 item 11          .                 |     *          .
 item 12          .             *   |                .
 item 13          .             *   |                .
 item 14          .  *              |                .
 item 15          .               * |                .
 item 16          .       *         |                .
=========================================================================
```

Figure 5. Plot of Standardised Differences

Figure 5 depicts the plot of standardized differences between male and female students. Items that have a value greater than plus or minus two indicate significant difference between the two groups (Barret, 2001 p.129). Figure 5 identifies only bias question, however the result on the standardized difference does not indicate any gender bias in the items. Therefore it can be said that there is no evidence to claim that there is no enough evidence of gender bias or systematic guessing. These results however answer the research question 1.

The values of discrimination index obtained from this study varied from .17 to .60. 11 items (1, 2, 3, 4, 5,6, 7,8, 9, 10, 11, 12, 13, 14, 15, and 16) have good discrimination coefficient (> 0.20) while only one item (6) indicates has very low discrimination coefficient (< .20). Item 6 then should be modified or removed from the test because it has little discrimination power (0.17). Barret (2001, p. 128) argues that "a discrimination coefficient of 0.20 is considered to the threshold and question below this figure should be deleted" Example of item discrimination taken from Quest presented in Table 1 and 2 below.

Trial1 2009_Nov15

Item Analysis Results for Observed Responses 15/11/2009 22:41
all on achivement (N = 285 L = 30 Probability Level=0.50)

.....

Item 6: item 6 Disc = 0.17

Categories	1 [0]	2 [0]	3 [1]	4 [0]	9 [0]	missing
Count	126	71	70	7	10	1
Percent (%)	44.4	25.0	24.6	2.5	3.5	
Pt-Biserial	0.03	-0.15	0.17	-0.02	-0.11	

.....

Table 1.Item analysis for question 6

Conversely, Item 16 is an example of a good item that should remain in the test as it has a discrimination coefficient of 0.40.

.....

Item 16: item 16 Disc = 0.40

Categories	1 [0]	2 [1]	3 [0]	4 [0]	6 [0]	9 [0]	missing
Count	36	205	16	15	7	5	1
Percent (%)	12.7	72.2	5.6	5.3	2.5	1.8	
Pt-Biserial	-0.18	0.40	-0.17	-0.20	-0.11	-0.12	

.....

Table 2.Item analysis for question 16

It is clearly on Table 2 above that some questions discriminate well on the basis of students' ability while other reminds to be discarded from the test.

7 CONCLUSION

The goal of this study is to use Rasch model to assess the match achievement of students. It is expected that fit means square values in Rasch analysis close to one when data fit the model. Sample size is keystone to achieve means squares fit statistics. The major conclusion drawn from this research was Rasch model anlyses offers a great deal for developing and analyzing of multiple choice tests.

Measuring students' ability with multiple choice is extremely important to measure their cognitive and spatial skills. This study found multiple choice item is applicable in examine the students' performance. It can also reveal the validity and reliability of tests and the bias gender. Almost all items are able to discriminate students however there is one item needs to be revised or modified to improve its discrimination power. Unfortunately, there is no evidence finding in this study that gender is influence the students' answer. This study also suggests that in multiple choice item, students are easy to guess the answer because there are many distraction options in a test.

REFERENCE

Adam, R.J., & Khoo S-T. (1997). Conquest: The interactive test analysis, ACER Press. Canberra.

Barret, S.2001. Differential item functioning: A case study from first year economics. *International Education Journal*, 2 (3)123-132

Bond, T.G., & Fox, C.M. (2007). Applyng Rasch model: Fundamental measurement in human sciences. Lawrence Erlbaum Associates. New Jersey.

Blackman, I,. Crespigny, C.D., & Parker, S. (2006). Mappingself-confidence levels of nurses in Their povision of nursing care to others with alcohol and Ttbacco dependence using Raschs Sacling.

David, C. (2004). Person Misfit in attitude Surveys: Influences, Impact and Implications. *International Education Journal*. 5 (2),125-143

Fox, C. (1999). A Introduction to the partial credit model for developing nursing assessments. *Journal of Nursing Eeducation.* 38 (8), 34--345

Hungi, N. (2003). Measuring school effects Aaross grades. Adelaide. Shannon Research Press.

Joyce, TB. Yu., Yates, M.S. (2007). A Rasch analysis of the academic self-concept questionnaire. *International Education Journal.* 8 (2) 47-484

Keeves, J.P., & Masters,G.MS. (1999). New approach to measurement. Advance in measurement in educatioal research and assessment. Pergamon. Elsevier.Netherland.

Linacre, J.M., Wright, B.D. (1994). Chi-Squarefit statistics. *Rasch Measurement Transaction.* 8 (2),350-359.

Olson, J.F., Martin, M.O., Mullis, INA, VS. (2008). TIMMS 2007 Technical Report. TIMMS and PIRLS International Study Center. Boston. USA

Thompson, M. (2003). The Aplication of Rasch scaling to wine judging. *International Education Journal.* 4 (3), 201-223

Walstead, W.B., & Robson, D. (1997). Differential functioning and female-male differences on multiple-choice tests in economics. *The journal of Economic Education* 28 (2), 155-169.