VOT 75213

DO OUR ENGINEERING STUDENTS HAVE WHAT IT TAKES?: THE CRITICAL THINKING SKILLS.

ROSMAH RAMLAN
AKBARIAH MOHD MAHDZIR

RESEARCH VOTE NO:
75213

Unit Bahasa Moden
Kolej Sains dan Teknologi
UTM Citycampus
Kuala Lumpur

2007

TABLE OF CONTENTS

ACKNOWLEDGEMENT

ABSTRACT **BAHASA INGGERIS**

ABSTRAK **BAHASA MELAYU**

CHAPTER 1 **INTRODUCTION/LITERATURE REVIEW**

CHAPTER 2 **METHODOLOGY**

CHAPTER 3 **FINDINGS**

CHAPTER 4 **CONCLUSION AND FUTURE WORKS**

ACKNOWLEDGEMENT

We wish to express our heartfelt gratitude to the College of Science and Technology lecturers who were willing to support us throughout the study. We would also like to thank the 1st year and 2nd year students of UTM Citycampus who were kind enough to sacrifice their precious time to be a part of this research.

ABSTRACT (Bahasa Inggeris)

A lot has been said recently regarding the products of Malaysian universities. One of the most highlighted issues was the perception from many parties that our graduates were not able to think critically. But that statement was merely a statement that has been blown out of proportion. As to date, there was a great shortage in researches that thoroughly looks into university students' critical thinking ability, and produces a result that can authenticate the statement that Malaysian graduates are not critical thinkers. The aim of this paper is therefore to look at how the Rasch Model will be able to facilitate the validity process in an effort to measure critical thinking aptitude in a multiracial, multi religion and multilingual society in a developing country in South East Asia like Malaysia. The instrument used for this research was adapted and translated from William Critical Thinking Assessment (Williams CTA, 2002). The CTA was designed to measure 4 sub-constructs: analysis, evaluation, inferences and interpretation. This instrument was claimed to be unidimensional and comprise a single factor, and the most recent Cronbach Alpha Coefficient was contended to be between .80 and .83 (William, 2002). The definition guiding the development of the instrument is based on the Delphi Mode l(Facione, 1992). Data collection were analyzed based on the Rasch Model for dichotomous data using the computer application, WINSTEPS version3.57.3 (Linacre). In addition, the data were also analyzed using SPSS version 11.5. The sample consisted of 567(2nd and 3rd Year) students representing students from 12 different courses from two public universities. The result shows that the instrument has very high item separation reliability (0.99). However, person separation reliability is too low (.14). On the whole, the students' critical thinking aptitude seems to be at a low level too. Are these supposedly selected students unable to think critically then? Are we using the correct measurement tool? This paper will answer this question.

ABSTRAK (Bahasa Melayu)

Sejak akhir-akhir ini banyak yang telah diperkatakan tentang graduan yang lahir dari universiti-universiti di negara ini. Antaranya ialah persepsi pelbagai pihak yang menganggap graduan Malaysia tidak mampu berfikir secara kritis. Ini adalah satu andaian yang telah diperbesarkan tanpa sebarang bukti yang kukuh dari mana-mana hasil penyelidikan. Tujuan penyelidikan ini adalah untuk melihat samada "Rasch Model" boleh membantu dalam usaha mengukur kebolehan pelajar universiti yang terdiri daripada berbilang bangsa, agama dan budaya berfikiran secara kritis. Instrumen yang digunakan telah diadaptasi dan diterjemahkan daripada "William Critical Thinking Assessment (William CTA, 2002)". Instrumen ini telah direkabentuk untuk mengukur 4 sub-konstruk : analisis, penilaian, inferens dan interpretasi. Ia juga dikatakan bersifat satu dimensi dan mengandungi faktor unggul dan nilai Cronbach alpha Coefficient ialah di antara .80 dan .83 (William, 2002). Instrumen ini berasaskan Delphi Model 1 (Facione, 1992). Data yang telah

dikumpul, dianalisis dengan menggunakan Rasch Model untuk “dichotomous” data dibantu oleh WINSTEPS version 3.57.3 (Linacre). Selain itu, data juga dianalisis menggunakan SPSS 11.5. Sampel yang terlibat di dalam penyelidikan ini terdiri daripada 567 pelajar tahun 2 dan 3 yang mewakili 12 bidang pengajian daripada dua buah universiti awam. Hasil kajian menunjukkan , instrumen yang digunakan mempunyai “high item separation reliability (0.99)”. Walaubagaimanapun, “person separation reliability” adalah rendah (1.14). Kesimpulannya, secara keseluruhan keupayaan pelajar universiti berfikir secara kritis agak rendah. Persoalan yang timbul, adakah pelajar yang terlibat benar-benar tidak mampu berfikiran secara kritis atau adakah alat pengukur yang digunakan tidak sesuai?

INTRODUCTION / LITERATURE REVIEW

In the early 80’s Dr. Mahathir Mohamad, the prime minister at that time presented a great idea in a working paper that was later known as WAWASAN 2020. He incorporated critical thinking (CT) among his recommendations for national education goal, a recommendation adopted by all educational levels today. As of today, most colleges and universities in Malaysia had included CT skills as an important educational objective in their goal statements, and many accrediting agencies included measurable gains in CT skills into their criteria. However, the effort in the integration of the skills in the education system is indistinct.

Quite a number of universities abroad had mandated the requirement of formal instruction in critical thinking throughout their universities e.g. California State University and all its 19 campuses. In Malaysia, a drastic movement as such is a long way to come. The call for critical thinking pedagogy and assessment has become ubiquitous in U.S., British and Canadian education (Morgan, 1995; Tucker, 1996). The same situation is felt locally.

In the local scenario, a lot has been said recently regarding the products of Malaysian universities. One of the most highlighted issues was the perception from many parties that our graduates were not able to think critically. But that statement was merely a statement that has been blown out of proportion. As to date, there was a great shortage in researches that thoroughly looks into university students’ critical thinking ability, and produces a result that can authenticate the statement that Malaysian graduates are not critical thinkers.

There are a number of reasons for this situation. One of the major problems is the lack of a shared understanding and agreement on what constitute critical thinking in the local context. The second reason, which is the focus of this paper, is due to the fact that most research conducted locally were utilizing instruments that were developed for a different population. In investigating a particular group or society, it is of utmost important for a researcher to be sensitive to the culture and norms of the target group. In doing a research on CT in the local context, it is important to understand the true meaning of critical thinking from Malaysian perspective, and therefore conceptualize a local, Malaysian definition of Critical Thinking.

METHODOLOGY

Respondents

The respondents were Second and Third Year students from two Malaysian Public universities whose undergoing 12 different courses ranging from the hard sciences (i.e.Mechatronics) to the arts (i.e TESL).

The subject consisted of 568 participants, approximately half of them were female (n=229) while the rest were male (n=339). Figure 1 shows the percentages of ethnic status. Approximately 88.9 % of the participants were Malay, and the remaining participants relatively evenly distributed across Chinese (4%), Indian (5%) and others(2%). Table 1 reports the frequency associated with the respondents' current cumulative grade point average(CGPA). 155 respondents received CGPA between 3.33 to 4.0, 150 respondents received CGPA between 2.7 to 3.29, 79 respondents with CGPA of 2.0 to 2.69 and 2 respondents with CGPA of 1.99 and below.

Figure 1: Ethnic

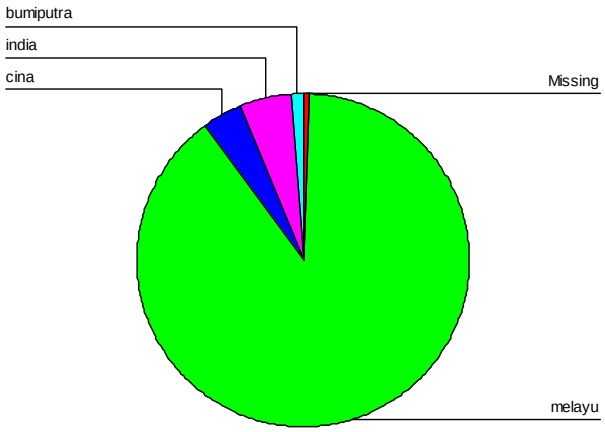


Table 1:
Frequencies of Respondents' Current CGPA

CGPA	Frequency
3.3 - 4.0	155
2.7 – 3.29	150
2.0 – 2.69	79
0 - 1.99	2

From Table 3, it is noticed that 56.5 % managed to get at least B6 in their SPM English, the remaining participants got C7 and below. For their SPM Additional Mathematics, the results were relatively evenly distributed across grades as shown in Table 3 except for grade 7, 8 and 9.

Table 3: English

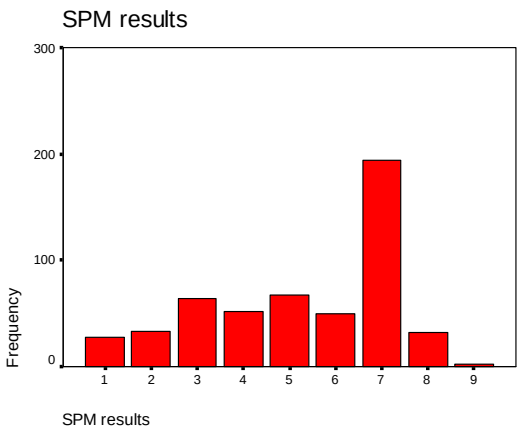
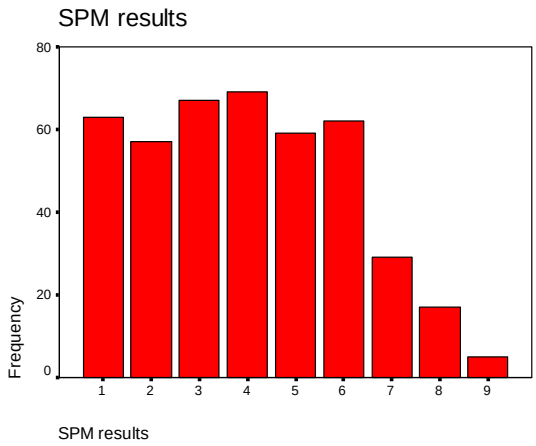


Table 4: Additional Mathematics



Instrumentation

The instrument used for this research was adapted and translated from William Critical Thinking Assessment (Williams CTA, 2002). Three language experts did the back-to-back translation. The CTA was designed to measure 4 areas: analysis, evaluation, inferences and interpretation. This instrument was claimed to be unidimensional and comprise a single factor, and the most recent Cronbach Alpha Coefficient was contended to be between .80 and .83 (William, 2002). The definition guiding the development of the instrument is based on the Delphi Model (Facione, 1992).

Data collection were analyzed based on the Rasch Model for dichotomous data using the computer application, WINSTEPS version3.57.3.(Linacre,). In addition, the data were also analyzed using SPSS version 11.5.

FINDINGS

We use WINSTEPS Version 3.57.3 (Linacre, 2005) to estimate ability, B for person, item difficulties (D), and fit statistics.

In 1960, George Rasch developed a probabilistic model in which each item difficulties (D) and person difficulty (B) was concurrently estimated. The Rasch analysis will result in a scale on which both persons and items are mapped onto the same construct in the same scale-free units, together

with the standard error. This mean the ability parameter for person and difficulty parameter for item represent the position of person and items on a latent trait they share.

The equation for Rasch model is:

$$P_{ni}\{x_{ni}=1 \mid B_n, D_i\} = \exp (B_n - D_i) / [1 + (B_n - D_i)].$$

Where $P_{ni}\{x_{ni} = 1 \mid B_n, D_i\}$ is the probability of person *n* on item *i* scoring a correct ($x = 1$) response rather than an incorrect ($x = 0$) response, given person ability (B_n) and item difficulty (D_i). This probability is equal to the constant *e*, or natural log function (2.7183) rose to the difference between a person's ability and an item's difficulty (B_n - D_i), and then divided by 1 plus this same value (Bond & Fox, 2001, p. 201).

In short, the two parameters are used in the model to determine the probability of person *n* succeeding on item *i* (Rasch, 1960). Bond and Fox (2001) further explained that the model presumes that the probability of a certain respondents to give a right answer to particular items is a logistic function of the relative distance between the item difficulty parameter and the respondents' ability parameter. They further states that the Rasch Model is based on a simple idea that all respondents are more likely to endorse easy items than difficult items. Hence, in Rasch analysis, items that receive lower ratings are more difficult to endorse than item that receive higher ratings.

Parameter estimates for each item are expressed in logits(log-odd probability units, Wright and Stone, 1979).

i. Reliability of the questionnaire

Item and person reliability is assessed based on Rasch Item Reliability Coefficient and Person Reliability Coefficient, which are equivalent to KR_20 or Chronbach Alpha. The reliability of 41 items in the questionnaires (Refer to Table 3.1) was found to be 0.99. This shows that the item ordering has a very high likelihood of being replicated if the same items are given to a group of students of comparable ability. From the analysis it was found that Person Reliability, which is equivalent to KR_20 or Chronbach Alpha, was low at .12 suggesting that a similar ordering of person placement cannot be expected if this sample of person were given another set of item measuring the same construct (i.e. critical thinking test). This statistics does not correspond with the findings of the instrument developer (William, 2002) that states the most recent Cronbach Alpha Coefficient for the instrument was claimed to be between .80 and .83. The item separation index of 11.15 indicates that the items in this scale could be statistically differentiated into about eleven different levels of difficulties.

```
TABLE 3.1 utmukmcta1                              ZOU174ws.txt Apr 27 18:11 2006
INPUT: 567 persons, 41 items  MEASURED: 567 persons, 41 items, 2 CATS
3.57.3
-------------------------------------------------------------------------------

     SUMMARY OF 567 MEASURED persons
+-------------------------------------------------------------------------------+
|           RAW                           MODEL          INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-------------------------------------------------------------------------------|
| MEAN      17.7      41.0       46.77    3.57      1.00     .0   1.00     .0 |
| S.D.       3.1        .1        3.94     .11       .16    1.2    .29    1.1 |
| MAX.      29.0      41.0       61.15    4.23      1.49    3.2   2.35    4.1 |
| MIN.       9.0      40.0       34.42    3.48       .59   -3.5    .48   -2.4 |
|-------------------------------------------------------------------------------|
| REAL RMSE   3.69  ADJ.SD    1.37  SEPARATION   .37  person RELIABILITY  .12 |
|MODEL RMSE   3.57  ADJ.SD    1.65  SEPARATION   .46  person RELIABILITY  .18 |
| S.E. OF person MEAN = .17                                                     |
+-------------------------------------------------------------------------------+
        VALID RESPONSES:  99.9%
person RAW SCORE-TO-MEASURE CORRELATION = 1.00 (approximate due to missing data)
CRONBACH ALPHA (KR-20) person RAW SCORE RELIABILITY = .15 (approximate due to
missing data)

     SUMMARY OF 41 MEASURED items
+-------------------------------------------------------------------------------+
|           RAW                           MODEL          INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-------------------------------------------------------------------------------|
| MEAN     244.4     567.0       50.00    1.02      1.00     .0   1.00     .1 |
| S.D.     123.4        .2       11.81     .27       .02     .9    .03    1.0 |
| MAX.     537.0     567.0       75.61    1.91      1.06    2.3   1.07    2.5 |
| MIN.      32.0     566.0       16.90     .86       .96   -2.1    .93   -2.0 |
|-------------------------------------------------------------------------------|
| REAL RMSE   1.06  ADJ.SD   11.77  SEPARATION 11.15  item   RELIABILITY  .99 |
|MODEL RMSE   1.05  ADJ.SD   11.77  SEPARATION 11.18  item   RELIABILITY  .99 |
| S.E. OF item MEAN = 1.87                                                      |
+-------------------------------------------------------------------------------+
UMEAN=50.000 USCALE=10.000
item RAW SCORE-TO-MEASURE CORRELATION = -.99 (approximate due to missing data)
```

ii. Item Fit

In determining item fit, two types of statistics were utilized: the Infit Mean-square and Outfit Mean-square statistics. Using mean-square values of 0.8 to 1.2 in diagnosing misfitting items (Bond & Fox, 2001), it was found that all of the items were productive for measurement. Since the analysis was done based Rasch Model, the analysis will results in a scale on which both persons and items are mapped onto the same construct in the same scale-free units. This means that the respondents' ability is independent of the item difficulty, and item difficulty is independent of the respondents' ability. This finding is quite disturbing for it proves that there is a great probability that our students level of critical thinking is slightly lower than average. But again, the question of "whose average" arises.

iii. Rasch Scaling of individual item and persons

The mapping facility in Rasch analysis allows us to see the distribution of each item of the four sub-domains together with the persons along a continuum. From Table 1.2, it can be seen that item difficulties range from 2.56 to -3.31 logit whereas person ability range from 1.12 to -1.56 logit.

It can be observed that almost 95 percent of the persons were unable to answer almost 30 percent of the items in the logit continuum correctly. This is interesting for this finding indicates that these items (approximately 12 items) are too difficult and far beyond respondents' ability to answer. Why does this situation occur? Is our students' CT aptitude really too low compare to their western counterparts? Statistically speaking, the answer is yes. However, looking from another perspective, further research need to be done to thoroughly look at the suitability of the items. There is a probability that the items, especially the highlighted 12 items), have some elements of biases if used in the local context. Hence, to be used in the local context, it is recommended that these item to be improved and localized.

From the map, it can also be observed that there is a lack of items to measure some range of item difficulty. More item should be included to ensure that respondents within that range of ability are given equal chance to demonstrate their true ability. In addition, it can also be seen that there are some items that are too easy for the respondents, and these items should be revised.

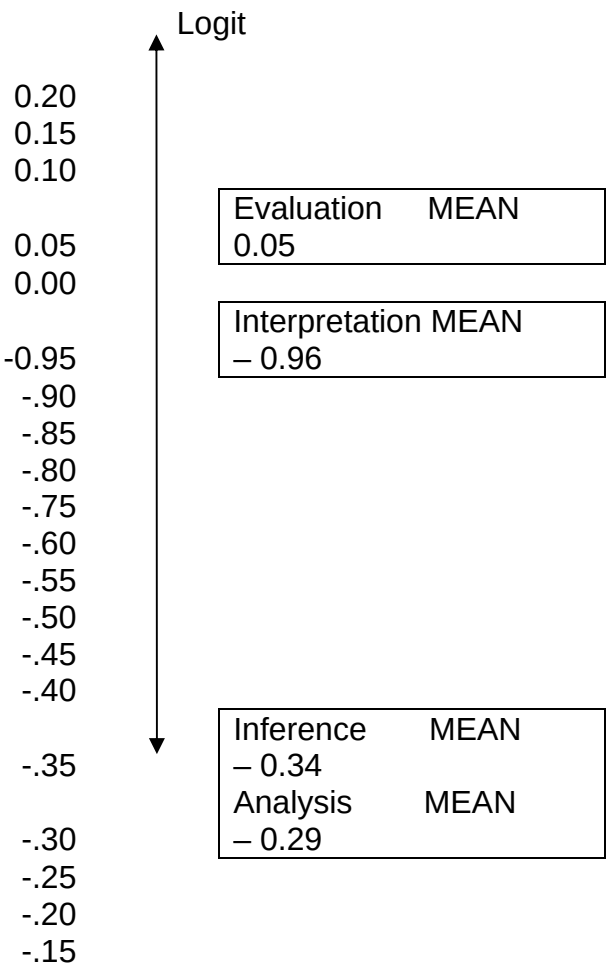
```
TABLE 1.2 utmukmcta1                              ZOU782ws.txt Apr 27 18:43 2006
INPUT: 567 persons, 41 items  MEASURED: 567 persons, 41 items, 2 CATS       3.57.3
-------------------------------------------------------------------------------

       persons MAP OF items
               <more>|<rare>
    3                +
                     |
                     |
                     |
                     |  a1     a20
                     |T
                     |
                     |
    2                +
                     |
                     |
                     |  in75
                     |
                     |  in70
                     |
                  .  |S a10    e16
    1                +  a6     in76
                     |  e19
                  .  |  in71   in72   in73
                  .  |                                          more items needed
                 .# T|  if68
                .##  |  a22    e11    if63   in74   in77
              .####  |  a4
           .####### S|  if62   if66   if67
    0   ###########  +M a7     e14    if61
       ############  |  a21
         .#########  |                                          more items needed
       ########### M|
       ############  |  a9     if65
          .########  |  e15    e18
            .###### S|  a23    a5
               .###  |  if64
   -1                +  a2     a24    a3     e12
                .## T|S                                         more items needed
                 .#  |  e17
                  .  |
                  .  |
                     |
                     |
                     |
   -2                +
                     |
                     |  e13
                     |T
                     |
                     |
                     |  in69
                     |
   -3                +
                     |
                     |  a8
                     |
                     |
                     |
                     |
   -4                +
               <less>|<frequ>
 EACH '#' IS 6.
```

The scaling of the 41 items that represent the four skills was also presented in Table 1.2. Each item is arranged along a logit scale representing from the easiest (bottom of the scale) to the hardest (top of the logit scale). It can be seen that the distribution of items are found mostly at the upper part of the scale indicating that most of the students are unable to complete the task given due to the fact that the questions are beyond their ability.

In order to get a better picture of students' level on the four sub-constructs, the mean measures of person for each sub-construct is computed and presented in the map given in Figure 1. From the figure, it is noticed that of the four sub-constructs, the respondents seems to be able to answer analysis item most (mean -0.29 logit). The respondents also seem to find that items on inferences to be on the easy side and this are shown through the mean measure (-0.34 logit). Items on interpretation (-0.96 logit), as well as on evaluation seem to be on the difficult side. It seems that the respondents mean measure for evaluation seems to be the highest (0.05 logit). This indicates that it is very hard for the respondents to get the items on evaluation sub-construct correct.

FIGURE 1: Overall Critical Thinking Assessment (mean measure)



CONCLUSION AND RECOMMENDATIONS

There are quite a number of established and available instruments in the market that measure critical thinking abilities such as Watson-Glaser Critical Thinking Appraisal, California Critical Thinking Skill Test, and Cornell Critical Thinking Test. From previous research that has been conducted locally, these instruments have been widely used, whether in its original form, translated form or adapted. However, there has yet been a thorough research that has been done to measure higher education students' critical thinking aptitude based on the Malaysian conceptualization of critical thinking.

How are we able to accurately measure our students' level of critical thinking if we are still dependent totally to the current available definition that is based on the thinking of a totally different society and culture?

The researcher also discovers that although these instruments have been proven to be good in measuring particular aspects of critical thinking, none attempted to measure every critical thinking domain. The researcher believes that there are still wide gaps in the definition itself to be used in the local context. Therefore, it was not impossible then that the available instruments in the market are less suitable and incomplete to measure the level of critical thinking among our Higher Education institutions students.

In conducting research regarding a particular group or society, it is of utmost important for a researcher to be sensitive to the culture and norms of the target group. It is important to understand the true meaning of critical thinking from Malaysian perspective, and therefore conceptualize a local, Malaysian definition of Critical Thinking. Hence, there is a crucial need for us to construct our own interpretation of critical thinking and develop an instrument that is able to accurately measure our students' critical thinking aptitude. The result of the studies will be very beneficial to the universities as a guide for them to take initiative to improve students' critical thinking abilities hence able to produce quality graduates.

REFERENCES

Bond, T.G; Fox, C.M. 2001. *Applying The Rasch Model Fundamental Measurement In The Human Sciences.* Lawrence Erl Baum Associates Publishers , London.

Facione, P.A. (1990). Critical Thinking: A statement of expert consensus for purposes of educational assessment and instruction. ("The Delphi Report"). Commisioned by the American Philosophical Association and published by the California Academic Press:Milbrae, CA as ERIC Document Reproduction Service

Linacre, J.M. 2005. *A User's Guide To WINSTEPS®MINISTEPS: Rasch-Model Compute Programs*

Morgan,Jr., W.R. (1995). Critical Thinking – What does it mean? Journal of College Science Teaching, 336-340

Rasch, G. 1960. *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

Wright, B., & Stone, M.H. 1979. *Best Test Design.* Chicago: MESA Press.