APPRAISAL OF COURSE LEARNING OUTCOMES USING RASCH MEASUREMENT: A CASE STUDY IN INFORMATION TECHNOLOGY EDUCATION

[1]Azrilah AbdulAziz, [2]Azlinah Mohamed, [3]NoorHabibah Arshad,
[4]Sohaimi Zakaria and [5]MohdSaidfudin Masodi

***Abstract*—** The development of Rasch measurement model in social science educational measurement has rapidly expanded to other areas of education including technical and engineering fields. This paper is an attempt of a paradigm shift in testing and evaluation in it education towards bio-based Rasch model. It is believed compatibility exist with the fundamental measurement currently used based on Kuhn's explanation on the role of measurement; in particular, Learning Performance Measurement System (LPMS). These cannot be gleaned from textbooks in engineering science or statistics. Taking the paradigm shift, faculty A, in one of Malaysian Institution of Higher Learning has embarked on the application of Rasch model to measure the achievement of it's Course Learning Outcomes (CLO) -Decision Support System, as a pre-requisite stipulated in the newly enacted Malaysia Qualification Framework, 2005 (MQF). Rasch model tabulates the students; i.e. Person and Items on a distribution map (PIDM) which gives a precise overview of the student's achievement on a linear scale for measurement. Comparative analysis against the traditional histogram tabulation and simple mean shows that Rasch measurement was found to give a better exploratory depth in understanding problems in information technology education. Despite the small sample size, the students were clearly categorized according to the respective cognitive skills hence; CLO's which was structured based on Bloom's taxonomy. Subsequently, it is possible to extend this approach in assessment of generic skills in students or professionals. This leads to a new paradigm in assessing competency of individuals using Rasch model.

Manuscript sent May, 4th 2007. This work was supported in part by the Ministry of Science, Technology & Innovation, MALAYSIA under National Science Fund, Grant No.: UiTM/ 100-PPSis(16/1/2). Revised received Dec.6,07

A. Azrilah, a PhD (IT) candidate is the main recipient of the Scholarship Grant. Her research on latent trait measurement is extended Rasch Model; a paradigm in technical education which has lead to new approach in performance evaluation. She can be reached at Tel/SMS : +6013 2937963 Fax : +603 41072262 or e-mail: azrilah@gmail.com

A. Mohamed, Ph.D; an Associate Professor in Intelligent System at the Faculty of Information Technology and Quantitative Science Universiti Teknologi MARA, 40800 S.Alam, MALAYSIA and reachable at azlinah@tmsk.uitm.edu.my,

N. Arshad, an Associate Professor in Intelligent System at the Faculty of Information Technology and Quantitative Science Universiti Teknologi MARA, 40800 S.Alam, MALAYSIA and reachable at habibah@tmsk.uitm.edu.my

Z. Sohaimi, Ph.D; an Associate Professor, is the Director of Division of Academic Affairs, Universiti Teknologi MARA 40450 Shah Alam, Selangor, MALAYSIA can be reached at e-mail: sohaimiz@salam.uitm.edu.my

M. Saidfudin, is a Research Consultant/ ISO9000 Lead Assessor at Centre for Teaching and Learning, Research Management Centre, University Teknologi Malaysia, Skudai, Johor, MALAYSIA reachable at e-mail:saidfudin@gmail.com

***Keywords*—** Rasch Model, Course Learning Outcomes, Kuhn paradigm, continuous improvement, quality education.

I. INTRODUCTION

"Science students accept theories on the authority of teacher and text, not because of evidence. Hence the standards of critical rationality are not present." Thomas Kuhn statement in his famous book *The Structure of Scientific Revolutions* (1970) remain a *paradigm* unchanged until complexity becomes apparent and more mind boggling, full of contradictions rather than providing solutions. The phenomenon is global and Malaysian education is of no exception. The quality of recent graduates in Malaysia has equally become more increasingly complex. This calls for a new model of education to generate scientist and engineers whose revolutionary ideas rather than a typical *scientists who are entrenched within a certain way of seeing things, and this clouds their vision* [1]. Malaysia Qualification Framework, 2005 (MQF) was developed and has promoted Outcome Based Education (OBE) learning process as an option. OBE calls for the evaluation of the Course Learning Outcomes (CLO) as specified in the program specification. This good practice has been implemented in Faculty A, in one of Malaysian Institution of Higher Learning (IHL), where teaching and learning processes was duly certified to ISO9001:2000. Assessment methods include giving students' tasks such as tests, quizzes or assignments at intervals during the 14 weeks study period per semester. CLO's were evaluated based on the students' performance which gives an indication of their learning achievements. Despite the marks obtained is orderly in manner, it is on a continuum scale. Hence, further evaluation using the raw score in Traditional Test Method is rather superficial and intractable to be carried out for this purpose.

This paper describes an alternative approach using bio-based Rasch Unidimensional Measurement Model as a more accurate performance assessment tool to measure the CLO. Use of similar nature based model is not new in computer engineering. Gauss Law deterministic empirical approach is found to be rigid and has been re-looked at using bio-inspired approach probabilistic model known as 'bootstrap method' [2] in advanced signal processing. Similarly, 'jack-knifing method' in digital imaging [3].

WinSteps uses 'logit' as the measurement unit thus transforms the assessment results into a linear correlation. An overview of the measurement model and its key concept is presented and illustrated. An assessment in Decision Support System (DSS) course using Bloom's Taxonomy, a behavioral science learning characteristics, as parameter was designed showing each dimension of the abilities to be measured. The results there from, were evaluated on how well it relates to the attributes being assessed and scrutinized. It is further checked against the CLO's maps for consistency and used as a guide for future improvement of the teaching method and style. This provides lecturers a more accurate far reaching insight of the student level of learning competency achieved.

WinSteps is a probabilistic model, where conceptually the probability of success is $\bar{x}$, the mean of the events; and in this case to be taken as the student's learning capability index; CL_i . This is represented by the Person's mean denoted as $Mean_{\text{person}}$ on the Person-Item Distribution Map (PIDM). This approach is backed up by vast literature in mathematical statistics, demonstrating it's theoretical validity. $Mean_{\text{person}}$ can be used to decide on the necessary course of action to achieve the desired level of learning competency through improved and more effective instructional plans. The assessment were evaluated on how well they relate to the content domain being assessed as stated in the CLO and results were analyzed to determine whether a gap exist in the IT student's capabilities or psychological construct that is supposedly to be developed.

The model employed a simple framework where learning performance assessment utilizes the Table of Test Specification designed on Bloom's Taxonomy parameters as the dimension of the student's cognitive abilities to be assessed. The results obtained were found to give a better in-depth understanding of the student's learning progression. The study also shows that Rasch Model of measurement can classify students' grades into unidimensional competency scale accurately using only very few primary data sets to enable corrective action to be taken effectively even at the early stage of learning IT.

This information is vital where it can be of meaningful use by reconnecting large-scale and classroom assessment onto mapped CLO's effectively for enhancement of classroom instruction. When lecturers are better informed of the CLO's hence; progress and difficulties of the students, it will serve as a good guide for us to make quality decisions about what a student needs to learn next and how to teach that material in a manner that will optimize the student's learning.

In this study, Rasch Model PIDM provides meaningful information on the student's learning effectiveness where decisions on what the student knows and where he or she should be in the instructional sequence. It's ability to generate information using only a very small sample allows effective monitoring of an IT students' learning while an instructional program is underway. Most important PIDM shows the tier of learning hurdles indicating vividly which specific difficulties account for the IT student's inadequate progress. Early intervention and remediation plan to improve learning progress can be designed henceforth.

II. BACKGROUND

Learning Performance Measurement System (LPMS) in Malaysian IHL is relatively new, undeveloped and yet to be studied systematically. Although some of the functional-based performance measurement on ranked ordered assessment of student's 'achievement' exists, little is understood on the purpose of such assessment. Raters varies across the board without realizing the importance of such assessment. As such, meaningful correlation for evaluation is difficult to establish. Little was written about the alignment of functional performance achievement to the overall planned CLO. Evaluations made thus far in Traditional Histogram Report remain superficial without much regard on how either student's learning outcomes or program objectives is affected. There is a need of a LPMS to examine the construct validity of the assessment instrument used to measure the quality of teaching services rendered. We need a performance measurement system in entirety which shows the relationships between learning performance and the method in which they operate.

The use of statistical-based measurement to monitor and control process and product quality was pioneered by Shewhart (1931), Juran (1951) and Deming (1975). In addition, Kane (1986) explored the use of capability indices as a measure of process quality. Papers by the authors attempting to establish the Learning Capability Index and measure the instrument construct validity by Item Response WinSteps was presented in Sharjah, UAE (2006)], Taipei (2007), and Greece (2007) [4-6] respectively. Responses received have provided some insight for the development and identification of salient dimensions and attributes for a more comprehensive assessment and effective evaluation system.

A method of defining the required dimensional metrics in Faculty A, learning ability measurement is setforth modelled on Shewhart's (*1939*) P-D-S-A Cycle which was subsequently developed into the infamous Deming's (*1957*) P-D-C-A Cycle by the Japanese industrial community. In year 2000, this fundamental concept was adopted by the international community in Geneva for ISO9001 Standard –Quality Management System. The Standard deals in great depth in Section 6 with regards to education, training, skills and competency of the human resources.

Though closely related to Item Response Theory (IRT), Rasch Model was derived from a distinct set of fundamental postulates, and the most important concept is being specific objectivity. WinSteps is the consequence of fundamental principles deemed important and indispensable. In the Rasch philosophy, the data have to comply with these principles, or in other words the data have to fit the model; hence validity. In Rasch point of view, there is no need to describe the data. What is required is to test whether the data allow for measurement on a linear interval scale specifically in a cumulative response process i.e. a positive response to an item stochastically implies a positive response to all items being easy or otherwise. Whether the data will fit depends on many factors e.g. How good is your substantive theory?, Does the latent variable actually exist?, Are the items uni-dimensional?, etc.

If the data fail to fit the model, there is something wrong with the data. Resorting to a more general model and still claiming to measure something is obviously a self deception. Rather, in Rasch we explore the possibilities in what way can the data be wrong; possibly the items may not be good enough, the setting of the data collection may be inappropriate, or there may be more dimensions, etc. We are looking into reasoned arguments.

One should try to find out what causes the misfits by reasoned arguments rather than accounting for part of the misfit by extending the model; and doing so by incorporating discrimination parameters can lead to grossly skewed statistical outcomes. Hence, the fundamental principles will no longer apply.

Rasch Model offers an excellent and comprehensive LPMS for CLO assessment which can enhance the understanding of education alignment, and assist educators in developing and maintaining quality education in Malaysian IHL duly aligned to the national interest; in this case IT education. Psychometrically Rasch Model in dichotomous case is expressed as follows:

$$\Pr\{ x_i = 1 \} = \frac{e^{\beta v - \delta i}}{1 + e^{\beta v - \delta i}} \qquad (1)$$

where $\Pr\{ x_i = 0,1 \}$, is the probability of turn of event upon interaction between the relevant person and assessment item;

e = Euler's number, 2.71828

β_v = the ability of person *v*

δ_i = the difficulty of assessment item *i*

This yields a Sigmoidal-curve; hence Rasch Model, the locii indicating the person's ability for a given task. When responses of a person are listed according to item difficulty, from lowest to highest, it generates a likely pattern known as Guttman pattern or vector; i.e. {1,1,...,1,0,0,0,...,0}.

In this case, the probability of success can be simplified and re-written in *logit*, which is a Logistic Regression Linear Hierachical Model [7]. It is readily shown that the log-odds, or *logit* of correct response by a person to an item, based on the model, is simplified as:

$$\textbf{Logit}\ (P/1\text{-}P) = \beta v - \delta i \qquad (2)$$

Hence, it can be construed the probability of a CLO is achieved is as shown in Figure 1.

$$\boxed{\text{Pr (a CLO a success)}} = \boxed{\text{Student's Ability}} - \boxed{\text{Difficulty of a given task}}$$

Fig.1 CLO success Model

III. MEASUREMENT METHODOLOGY

This study shall address the measurement of each relevant CLO as attributes for the course DSS teaching achievement and evaluate the students' learning progress. For this purpose, Rasch Unidimensional Measurement Method (using WinSteps Software) is applied where accurate findings can be yield even by using a small data-set. The item difficulty encountered in building up the students' required learning ability and cognitive skills development can be duly analysed. Table 1 shows the linkages of the learning performance measurement for each identified student's cognitive ability based on Bloom's Taxonomy.

Table. 1 Summary of Course Learning Outcomes based

Series	Course Learning Outcomes
CLO 1	**Acquire** knowledge components required of a decision-making process, Decision Support System (DSS), group DSS and expert system.
CLO 2	**Determine** the differences between individual, group and enterprise decision-making processes.
CLO 3	**Understand** the required principles of DSS and other techniques in real-world projects.
CLO 4	Creative applications to **solve problems** in complex, unstructured ambiguous situations under conditions of uncertainty.
CLO 5	Apply high power of **critical analysis** shown by comprehensiveness of approach to decision making.
CLO 6	Critical **evaluation** of literatures on decision making.
CLO 7	**Synthesis** – a holistic understanding to deal with complexity and contradiction in the knowledge base through the application of multiple perspectives on IT managerial situations

Bloom's Taxonomy parameter of various cognitive learning level is applied when the CLO was established. Bloom's Taxonomy cognitive learning levels are; Level-1:Knowledge, 2:Understand, 3:Apply, 4:Analyse, 5:Evaluate, and 6:Synthesis; are all measurable. For course DDS, the students were expected to develop Level 4-6; i.e. evaluation and synthesize knowledge acquired to analyse situations requiring them to provide the appropriate solutions.

Assessment marks distribution is as pre-determined in the typical Course Outline. Questions were then prepared based on the percentage distribution as stated in the Table of Test Specification. Table 2 shows the spread of questions for each assessment; viz. Tests, Assignment and Final Examination in relation to the stipulated CLO's.

Table. 2 Table of Test Specification based on CLO

CLO	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Mark by LO	%
1		`	2	8	4	5		19	13.57
2	5							5	3.57
3		12	2				4	18	12.86
4			4			3		7	5.00
5		8	12		16		6	42	30.00
6	15						10	25	17.86
7				12		12		24	17.14
Sum	20	20	20	20	20	20	20	140	100.00

The course assessment forms were designed and developed according to the Table of Test Specification for the purpose of this study. This is a preliminary exercise towards actual implementation of the OBE assessment method. One of the exam papers was scrutinized to properly determine the CLO prescribed in that test. This instrument is of importance where their construct validity; where it should measure what is to be measured, is often being debated. This assessment form is vital as it gathers the presumed empirical data as the main component of this study.

WinSteps is then used to test the data obtained subsequent to the assessment done to ascertain whether it fits into the model. "Misfit" data are outliers that need to be dealt with. As stated earlier, Rasch is a probabilistic model and differs from many other models that are commonly deterministic. Thus, Rasch offers an alternative to a common practice that people do, which is the use of simple means over several items in a simple histogram; where the problem of this approach is famous. Test items are ordinal variables and they are not linear measures, hence lack accuracy.

In Rasch measurement, data is multi-dimensional but measurement is unidimensional, hence putting everything on the same scale. Rasch emphasize the shifts of reliability and validity in traditional Cronbach-α and Factor Analysis to the reproducibility of measures rather than expressing the reproducibility of raw scores [8]. By focusing on the reproducibility of the latent trait rather than forcing the expected generation of the same raw score, the concept of reliability takes its rightful place in supporting validity rather than being in tension with it.

Normally decisions have often focused on reliability and measurement error rather exclusively, ignoring sampling error and sample size considerations. Rasch goes beyond Generalizability theory by providing a more comprehensive framework of measurement [9]. In Rasch, the measure variance and measurement error of the subjects is continually estimated throughout the data analysis. It also evaluates and reports data quality (fit), identifying items and persons with learning performances meriting special attention, perhaps remediation, or even omission from the current analysis.

Sample of the students assessment result is compiled, re-distributed in CLO format and tabulated as shown in Table 3. Method of CLO marks re-distribution is discussed at length by the authors in previous papers and the Guide to Program Objectives and Learning Outcomes Evaluation Model respectively. Students were coded as ST*nn*X. The study also delves into Differential Item Functioning (DIF) between genders; X coded 1 for Male and, 2 for Female.

This will enable to establish the discrimination index; the construct validity of the instrument in separating the students of different ability irrespective of gender or socio-economic background.

These raw score results are then transformed to numeric Grade Rating by cluster similar to the typical order rank A-E; in this case the following rating is used;

Table. 3 Students Assessment Result: CLO Marks Tabulation

Student	RAW SCORE (%) BY LEARNING OUTCOMES Course: DDS Session: SEM 2/2007-08							
	DIF	CLO 1	CLO 2	CLO 3	CLO 4	CLO 5	CLO 6	CLO 7
ST01	1	47.4	-	33.3	100	47.6	-	58.3
ST02	2	34.2	-	58.3	85.7	61.9	-	70.8
ST03	1	36.8	100	22.2	35.7	35.7	74.0	79.2
ST04	1	47.4	100	63.9	57.2	73.8	42.0	41.7
ST05	1	47.4	100	88.9	42.9	40.5	92.0	50.0
ST06	2	55.3	-	33.3	100	61.9	-	91.7
ST07	1	57.9	-	83.3	42.9	63.1	-	91.7
ST08	1	79.0	100	55.6	42.9	44.1	60.0	91.7

Grade Rating based on marks cluster:

>80 =**5** >70 =**4** >60 =**3** >50 =**2** >40 =**1** <40 =**0**

The grade rating is then tabulated in Excel **.prn* format as shown in Table 4. This numeric coding is necessary for further evaluation of the CLO achievement using Rasch Measurement Model.

In Rasch, the theoretical mean is deemed to be the probability of success. It is readily shown that a way to calculate the mean; $\bar{x}_i$ is simply to take the proportion of cases with each score, multiply by the value of the score, and add them up; expressed in equation form as:

$$\bar{x}_i = \sum_{x=0}^{k} p_{xi} x_i \qquad (3)$$

where, k = maximum grade rating

P_{xi} = proportion of event for each Grade Rating

x_i = ascertained Grade Rating; n=1,2,n_i… n_k

The frequency proportion of events where student obtained a certain Grade Rating is then established to compute the probability of achievement for each given CLO.

Table. 4 Students Assessment Result Tabulation **.prn*

Student	RATED SCORE BY LEARNING OUTCOMES Course: ITS754 Session: SEM 2/2007-08							
	DIF	CLO 1	CLO 2	CLO 3	CLO 4	CLO 5	CLO 6	CLO 7
ST01	1	1	0	0	5	1	0	2
ST02	2	0	0	2	5	3	0	4
ST03	1	0	5	0	0	0	4	4
ST04	1	1	5	3	2	4	1	1
ST05	1	1	5	5	1	1	5	2
ST06	2	2	0	0	5	3	0	5
ST07	1	2	0	5	1	3	0	5
ST08	1	4	5	2	1	1	3	5

Grade Rating: Marks >80=**5** >70=**4** >60=**3** >50=**2** >40=**1** <40=**0**

IV. DISCUSSION AND FINDINGS

The data is processed using WinSteps software to conduct the necessary computation. First is to establish the Person-Item Distribution Map (PIDM). Figure 3 shows the Person=ST*nn* GenderX; i.e. student's n^{th} series geographical position in relation to the test Item distribution; CLO's. Note the similarity to the Traditional Histogram tabulation; but now both Person and Item distribution are put on the same *logit* scale in line with the Latent Trait Theory. It is concerned with how likely a person *v* of an ability β on the latent trait is to response to an item *i* of difficulty δ_i.

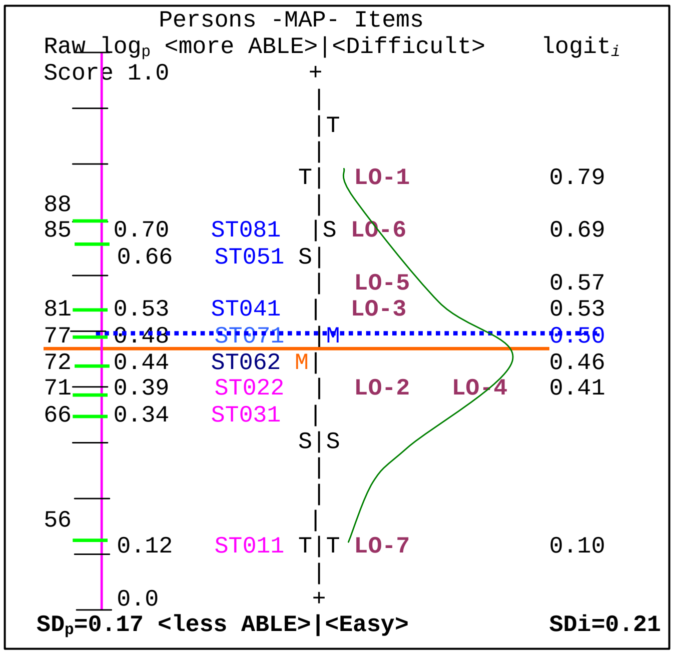


Fig.2 Person-Item Distribution Map

Thus, the parameter δ is the location of the item on the same trait: if β_n is greater than δ_i then the person is likely to be able to respond to the item correctly. The degree of a person's ability is indicated by the separation of the item against the person's location on the map: the further the separation, the more able a person likely to respond correctly to the said item. Similarly, the extent of an item difficulty is reflected by the spread of the item over the scale: akin to the high jump bar; the higher the location from the item mean, $Mean_{item}$, then the item is perceived to be more difficult as compared to an item on a lower location. Thus, the $Mean_{item}$ serves as the threshold where it is set to 0.5 on the *logit* scale.

It is observed in the PIDM that the cohort's $Mean_{person}$=0.46 which is unexpectedly lower than the threshold value, $Mean_{item}$=0.50 indicating these students have low ability. Five (5) of the students (62.5%) were found to be below $Mean_{item}$. These poor students generally have difficulty in attending all the CLO's except CLO2-*Determine*, CLO4-*Problem solving* and CLO7-*Application of multiple perspectives.* Now the Lecturer can take specific instructional measures of corrective action on the respective students. For example; a bridging tutorial can be arranged for this purpose.

The PIDM reveals that CLO1-*Acquire knowledge* is the most difficult item encountered whilst CLO7-*Application of multiple perspectives* is the easiest item understood by the students. Overall, students find it difficult to have good command on all the expected CLO. The students also shown to lack the CLO1-*Acquire knowledge*, which should has no problem to the students at this stage. On other hand, all the students fulfilled CLO7-*Application of multiple perspectives*, which supposed be achieved in the Bloom's level by this cohort.

This PIDM detail out the exact position of each student ST*nn*X in relation to the respective CLO's; where ST*nn*X is Student n^{th} series of Gender X: Male is coded as 1 and Female, 2. Take ST081-Male student; his PIDM shows that he has fulfilled all the CLO requirements except for CLO1-*Acquire knowledge* and he barely fulfilled CLO6-*Evaluation*; while ST062-Female student is having problem with CLO1-*Acquire knowledge*, CLO6-*Evaluation*, CLO5-*Critical analysis* and CLO3-*Understanding* respectively. The poorest student ST011-Male *logit* 0.12 is certainly a problematic Outfit where he did not fulfilled all the CLOs except CLO7-*Application of multiple perspectives.*

The PIDM also allows us to scrutinize the gender performance; both Female students ST062 and ST022, have low ability below the $Mean_{person}$= 0.50 while three (3); ST081, ST051, STO41, out of six (6) Male students possess high ability than the $Mean_{person}$. This is called Differential Item Function (DIF) showing the variance of learning ability between the genders.

Such detailed information is not available in a typical Histogram reporting based on Traditional Test Theory. Rasch Measurement is far superior and rich with specific information that enable the Lecturer to pinpoint the exact nature of the instruction problems and how the students progress over each course towards meeting the expected Program Objectives. Table 5 shows the detail analysis showing the correlation of each person to each item; the tendency of odds for them to be able to respond correctly expected for each LO.

It detailed out the probability of each student in achieving each LO respectively. The probability of each LO achievement is then computed;

$$P(\theta) = \beta_v - \delta_i \qquad \text{from (2)}$$

$$P(\theta) = \frac{e^{\beta v - \delta i}}{1+e^{\beta v - \delta i}}$$

Let us look at ST081 in calculating his probability of achieving LO-1 and LO-7;

$$\begin{aligned} P(\theta) &= \beta_v(ST081) - \delta_i(\text{LO-1}) \\ &= 0.70 - 0.79 \\ &= -0.09 \end{aligned}$$

Table.5 Consolidated FTMSK ITS754 EMT Learning Outcomes

Person	Logit Person Measure	Item	Logit Item measure	Probability of success of respective students on each LO							
				P(ST081)	P(ST051)	P(ST041)	P(ST071)	P(ST062)	P(ST022)	P(ST031)	P(ST011)
ST081	0.70	LO-1	0.79	0.48	0.47	0.44	0.42	0.41	0.40	0.39	0.34
ST051	0.66	LO-6	0.69	0.50	0.49	0.46	0.45	0.44	0.43	0.41	0.36
ST041	0.53	LO-5	0.57	0.53	0.52	0.49	0.48	0.47	0.46	0.44	0.39
ST071	0.48	LO-3	0.53	0.54	0.53	0.50	0.49	0.48	0.47	0.45	0.40
ST062	0.44	LO-4	0.41	0.57	0.56	0.53	0.52	0.51	0.50	0.48	0.43
ST022	0.39	LO-2	0.41	0.57	0.56	0.53	0.52	0.51	0.50	0.48	0.43
ST031	0.34	LO-7	0.10	0.65	0.64	0.61	0.59	0.58	0.57	0.56	0.50
ST011	0.12										

Subtitute into:

$$P(\theta) = \frac{e^{\beta v - \delta i}}{1+e^{\beta v - \delta i}}$$

$$= \frac{e^{-0.09}}{1+e^{-0.09}}$$

$$= 0.4775$$

ST081(LO-1) % Level of Achievement = 47.75%

$$P(\theta) = \beta_v(ST081) - \delta_i(\text{LO-7})$$

$$= 0.70 - 0.10$$

$$= 0.60$$

Subtitute into:

$$P(\theta) = \frac{e^{\beta v - \delta i}}{1+e^{\beta v - \delta i}}$$

$$= \frac{e^{0.60}}{1+e^{0.60}}$$

$$= 0.6457$$

ST081(LO-7) % Level of Achievement = 64.57%

It can be observed from Table 5 that 6 out of 8 students (75%) have problems with their CLO achievement. Further scrutiny reveals that these students have difficulties in achieving LO-1; *Acquire knowledge*. Out of the 6 students having difficulties in achieving their CLOs, 1 of them (17%) have difficulties in achieving almost all the LO's except LO-7. This is quite interesting to note, since it is considered normal if students achieved the lowest LO better as compared to the highest LO.

However in this case it is a reversed findings. This warrants a further investigation for the root cause either on the students, the test questions, or perhaps the instructional method.

The reverse findings are atypical to all the students in this cohort; all the students find it easy to achieve the high LO instead of the low LO. This indicates a different conclusion to the phenomena. It can caution us that there is nothing wrong with the students; it only reflected that there should be further scrutiny on the test validity itself.

Evaluation of Person-Item Correlation Order is done next. This is to establish the construct validity of the instrument of assessment. Table 7 shows the value of Cronbach-α = *0.00*.

The construct validity of the test is subjected to further scrutiny.

As previously shown, in Rasch Measurement the mean of events, $\bar{x}_i$, is the probability of success. The mean of the sample averages will approximate the mean of the population [10]. The computation of such mean, $\bar{x}_i$, depends on the raw score from each respondent. But pure raw score is susceptible to respondents making a guess at the answers or not being truthful, and in such a case, the answers cannot be trusted to be reflective of the population.

In Rasch probabilistic model, estimates for each items used in the test and each respondents are given separately. This is a key feature of the Rasch model; the estimates of the locations of the items can be obtained independently of the locations of the persons [11] .

The shortcomings of using pure raw scores which tends to group the students around the median and not adequately contrasting the results of the survey by the ability of students are remedied when a Rasch measurement analysis is made [12]. Rasch measurement provides the standard errors for every person and item. These standard errors can be squared and summed to produce a correct average error variance for the sample or any subset of persons and for the items or any subset of items.

Prediction from the student's answers are deemed ordinal responses. This render measurement on the LO's are almost impossible due to absence of interval in the scale. Rasch transformed these ordinal data into linear function.

The normal solution in linear regression approach is to establish a line that fits the points as best as possible –'best fit line'; which is then used to make the required predictions by inter-polation or extra-polation as necessary as shown in Figure 3.

$$y = \beta_0 + \beta_1 m \qquad (4)$$

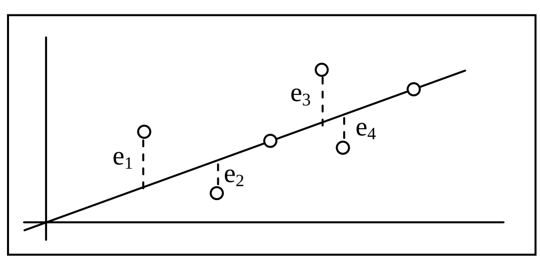


Fig. 3 Best fit Line

In obtaining this best fit line, however, there exist differences between the actual point, y_i , and the predicted point; $\acute{y}_i$, that is on the best fit line,. The difference is referred here as error; *e*

$$y_i - \acute{y}_i = e_i \qquad (5)$$

By accepting the fact that there are always errors involves in the prediction model, the deterministic model of equation (5) renders itself less reliable. This can be resolved by transforming it into a probabilistic model by including the prediction error into the equation;

$$y = \beta_0 + \beta_1 m + e \qquad (6)$$

Hence, Rasch enabled the concept of reliability from establishing "best fit line" of the data into producing a reliable repeatable measurement instrument instead. Rasch focuses on constructing the measurement instrument [8, 11]; in this case the Test paper, with accuracy rather than fitting the data to suit a measurement model with errors. By applying the appropriate probabilistic model, the error is absorbed; now Rasch measurement represents a more accurate prediction instead.

In Rasch philosophy, the data have to comply with the principles, or in other words the data have to fit the model. In Rasch's point of view, there is no need to describe the data. What is required is to test whether the data allows for measurement on a linear interval scale specifically in a cumulative response process i.e., a positive response to an item stochastically implies a positive response to all items being easy or otherwise. This dichotomous responses can take only two values, 0 and 1 , which is known as Bernoulli random variable; in our case a learned student, β_n , who produced a good learning outcomes or otherwise, due to some level of difficulty carrying out the learning task, δ_i.

Since the Person_{mean} is the probability of success, then the level of CLO achievement can be taken as;

$$P(\theta) = \beta_v(Mean_{person}) - \delta_i(Mean_{item})$$
$$= 0.46 - 0.00$$
$$= 0.46$$

Substitute into:

$$P(\theta) = \frac{e^{\beta v - \delta i}}{1+e^{\beta v - \delta i}}$$
$$= \frac{e^{0.46}}{1+e^{0.46}}$$
$$= 0.6130$$

Mean % LO learning Achievement = 61.30%

This is a stark contrast against a mere 44.75% if measurement is attempted to be derived from pure mean raw score of 17.90 obtained from Table 7 -Mean Raw Score;

$$\text{Expected full score} = N_{students} \times \text{max. rating};$$
$$8 \times 5 = 40$$
$$\Pr(\text{Success}) = \frac{17.9}{40} \times 100 = 44.75\%$$

This is the very reason why the need to transformed the ordinal responses into *logit.* It is an interval scale which makes it measurable where in Rasch the errors are already accounted for. It can be readily shown mathematically that a series of numbers irrespective of based used, is not equally spaced but distant apart exponentially as the number gets bigger while a log series maintain their equal separation, hence interval. This equal separation is shown in Table 6 and we call it *logit.* The difference between $\log_{10}5$ and $\log_{10}2$ is constant and remains of equal distant between $\log_{10}50$ and $\log_{10}20$. Similarly for $\log_e$; hence *logit*.

Table. 6 Comparison of Original and Log intervals

Original series	**$\log_{10}$**	**$\log_e$**
1	0.000	0.000
2	0.301	0.694
5	0.699	1.609
10	1.000	2.303
20	1.302	2.997
50	1.699	3.912
100	2.000	4.606

Now the true LO measurement can be obtained by substituting the $Mean_{person}$ *logit* against the $Mean_{item}$ *logit* from the Person Item Distribution Map which serves as the measurement of the LO effectiveness.

Table. 7 Consolidated FTMSK ITS754 EMT Learning Outcomes Student's Assessment: Persons Item Statistics: Correlation Order

```
INPUT:  8 Persons  7 Items   MEASURED: 8 Persons   Items 6 CATS
Person:REAL SEP.:.00  REL:.00  Item:REAL SEP.:.00  REL.:.00
     CRONBACH ALPHA (KR-20) Person RELIABILITY .00   SD 0.17
```

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	Item
4	20	8	.41	.19	1.66	1.9	1.77	2.1	-55	LO4
5	16	8	.57	.20	.62	-1.2	.58	-1.2	11	LO5
7	28	8	.10	.21	.79	-.4	.88	-0.1	.15	LO7
6	13	8	.69	.21	1.06	.3	1.08	0.3	.51	LO6
1	11	8	.79	.23	.43	-1.3	0.39	-1.3	.55	LO1
2	20	8	.41	.19	1.44	1.4	1.41	1.3	.58	LO2
3	17	8	.53	.20	.86	-.3	.84	-0.3	.62	LO3
MEAN	17.9	8.0	.50	.20	.98	.0	.99	0.1		
S.D.	5.2	.0	.21	.01	.41	1.1	.44	1.1		

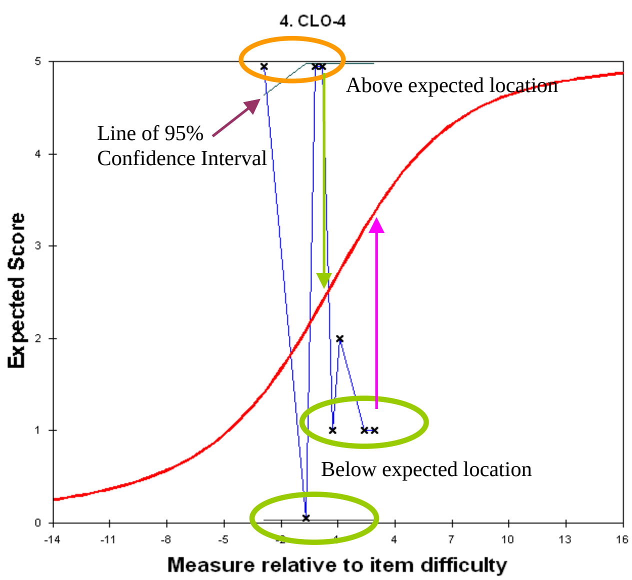
Subsequently, a check of Point Measure Correlation (PMC) gives the content validity of the items. The working parameter for an acceptable PMC value shall be between: $0.4 < x < 0.8$ [10]. CLO5- *Critical analysis* and CLO7- *Application of multiple perspectives*, both below than 0.4 which is at 0.11 and 0.15 respectively which needs further evaluation.

Next is to check the corresponding Outfit values. The choice is obvious because it is easier to explain an outfit as compared to infit problems. The acceptable Root Mean Square

(MNSQ) value is between; $0.5 < x < 1.5$. Value beyond has implication for measurement where it distorts or degrades the measurement system and unproductive for the construction of accurate measurement [13a]. Analysis shows CLO4- *Solving problem* has a high unexpected MNSQ=1.77 with a Z-standard score; the equivalent *t*-test, of a 2.1. CLO1- *Acquire knowledge* with MNSQ=0.39, however has Z-STD= -1.3 which is within the acceptable range of Z-STD value= ± 2 [13b].

This warrants an in-depth review of the items construct to ascertain the instrument validity in measuring what is it supposedly to measure; i.e. the Bloom's cognitive skills development as stipulated in the Table of Test Specifications. CLO4 –*Solving problem* is further evaluated by Item Characteristic Curve technique to check the misfit data. Figure 4 shows the Item Characteristic Curve, the spread of respondents against item difficulty. It was found that CLO4 –*Solving problem* is over-discriminating. The less able respondent were unexpectedly found to have scored very high beyond the 95% confidence interval whilst the more able did not score as expected, hence misfit data. Several possible reasons could have contributed to such anomaly or perhaps the assessment form itself need to be discarded.

Fig.4 Item Characteristic Curve of CLO-4:
: Creative Problem Solving



V. CONCLUSION AND RECOMMENDATIONS

Perhaps by now, it is noticeably conspicuous on the systemic inquiry of Rasch Model has provided us thus far. This is vital as the validity of the students' assessment is crucial in OBE. The CLO measurement has significant contribution towards realising the Program Outcomes; and this is a fundamental requirement in OBE accreditation program.

The binomial bell graph skewness showing tendency of inclination and the graph kurtosis dictating the object spread is easy and friendly for evaluation. Values on the map serves as an indicator; on the item easiness and gives a locii on the quality level of the respective item and person under scrutiny. Rasch modeling can be used to form valid measures on different dimensions of learning achievement. The respective students learning ability development may be tracked over their study period and instructional method and style may be improved to facilitate such latent development. Symptoms of students' weaknesses in certain generic skill trait can be traced more effectively and easily.

This simple but prudent conceptual theoretical framework of measurement is capable of examining teaching and learning effectiveness in great depth and width. Rasch measurement uses empirical data directly from the student's assessment for a given instruction method; in this case Active Learning. Using Rasch Measurement, the result more accurately classified the students according to their observed achievements. It enables each item to be evaluated discretely. Further application of Rasch Measurement in IT Education has a large potential and of significant contribution towards the development of future ingenious graduates.

ACKNOWLEDGMENT

The authors wish to express gratitude to the Experts who contributes their effort and views in the Face-validity test. Many thanks to the Ministry of Science, Technology and Innovation for their support in providing scholarship: Grant No.: UiTM/100-PPSis(16/1/2) to enable this subject matter to be pursued as a research subject towards the award of Doctor of Philosophy (IT) by University Teknologi MARA, Shah Alam, MALAYSIA.

REFERENCES

[1] Kuhn, T., *The Structure of Scientific Revolutions*, Harvard University Press. 1970

[2] Zoubir, AM and Iskander, DR.; "*Bootstrap methods and Applications in Signal Processing*"; pp. 10-19. IEEE Signal Processing Magazine, July 2007

[3] Thompson, D., DR.; "*Jacknifing Multitaper Spectrum Estimates*"; pp. 20-13. IEEE Signal Processing Magazine, July 2007

[4] "Establishing a method of measurement towards developing quality information in a Knowledge Management System"; *Proceeding, International Conference on Knowledge Management in Higher Learning Institution. Bangkok. Feb.2006.*

[5] Aziz, A., Johari, J and Saidfudin, MS.; "Outcome Based Education: A Computational Measurement On Special Librarian Intelligence Competency As Information Professional", *Proceeding, International Conference on Media and Information Warfare (Mi-WAR)* Kuala Lumpur, Malaysia; August 16 - 17th, 2007

[6] Aziz, A, Zakaria, S and Saidfudin, MS; "*A Rasch-based Longitudinal Assessment Model to measure Information Management Subjects Learning Outcomes*" Proceeding WSEAS 3rd International Symposium on Datamining and Intelligent Information Processing. Beijing, China; September 2007

[7] Uekawa; "*Explaining Rasch Model through HLM Regression*", retrieved 25 May 2007; http://www.estat.us/id115.html

[8] Mok, M. and Wright, B; "*An overview to Rasch Model and Measurement*", University of Chicago, 2004

[9] Marcoulides, G. A. (1999). *Generalizability theory: Picking up where the Rasch IRT model leaves off ?* Mahwah, NJ, USA: Lawrence Erlbaum
[10] Saunders, M. NK., Lewis, P. & Thornhill, A. (1997). *Research methods for business students.* London:Pitman.
[11] Bond, T. & Fox, C. (2007). (2nd. Ed.) *Applying the Rasch Model: Fundamental Measurement in Human Sciences.* New Jersey:Lawrence Album Associates.
[12] Stone, M. H. (2002). Quality Control in Testing. *Journal of the Institute for Objective Measurement*, 4, 15-23.
[13,a&b] Linacre, M.*"What do Infit and Outfit Mean-Square and Standardized mean?"*, Rasch Measurement Transactions 16:2 Autumn 2002.