Outcome Based Education Performance Measurement: A Rasch-based Longitudinal Assessment Model to measure Information Management Courses LO's

[1] AZRILAH ABDUL AZIZ, [2] ABD RAHMAN AHLAN
[3] SOHAIMI ZAKARIA and [4] MOHD SAIDFUDIN MASODI
[1] PhD Candidate, [3] Associate Professor
Universiti Teknologi MARA, Faculty of Information Management
Kampus Puncak Perdana, No.1, Jalan Pulau Angsa A U10/A
Seksyen U10, 40150 Shah Alam, Selangor, MALAYSIA
Tel/SMS : +6019 210 5172
[2] Director, Centre for Information Technology Advancement
International Islamic University Malaysia, Gombak, 53100 Kuala Lumpur, MALAYSIA
[4] Researcher / ISO QMS9000 Lead Assessor
Center of Teaching and Learning, Research Management Center
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, MALAYSIA.
Tel/SMS : +6012 2402821 Fax : +603-41084292

Abstract:- Malaysia Qualification Framework, 2005 (MQF) promotes outcome based education (OBE) learning process. OBE calls for the evaluation of the subjects Learning Outcomes (LO) as specified in the Program Specification. This good practice is implemented in the Faculty of Information Management, Universiti Teknologi MARA (FIM) teaching and learning processes which was duly certified to ISO 9001:2000. Assessment methods include giving students' tasks such as tests, quizzes or assignments at intervals during the 14 weeks study period per semester. Subject LO'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 is rather complex to be carried out. This paper describes a Rasch-based measurement model as performance assessment tool to measure the subject LO. Rasch Model uses 'logit' as the measurement unit thus transforms the assessment results into a linear correlation. An overview of the measurement model and its key concepts were presented and illustrated. An assessment form using Bloom's Taxonomy as parameter was designed showing each dimension of the ability 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 Course LO's Maps for consistency and used as a guide for future improvement of the teaching method and style. This provides the lecturers a more accurate insight of the student level of competency achieved. The study shows that this model of measurement can classify students' grades into competency scale accurately using only very few primary data sets to enable corrective action to be taken effectively at the early stage of learning.

Keywords: Learning Outcomes, performance assessment, evaluation, competency, Bloom's Taxonomy, quality.

1 Introduction

A major impetus for the performance assessment movement has been the need to reconnect large-scale and classroom assessment to mapped LO's so that assessment affects learning positively thus enhancing instruction in classroom. When we are better informed of the subjects LO'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 maximize the student's learning. This knowledge enrichment process is vital to inculcate students' zest for knowledge.

In information management education we make three(3) types of decisions using assessment results[1]:

a. Instructional placement decisions: what the student knows and where he or she should be in the instructional sequence i.e., what to teach next.
b. Formative evaluation decisions: information to monitor an information management students' learning while an instructional program is underway; how quickly progress is being made, whether the instructional program is effective, and if a change in instructional program is needed to promote the information management student's learning.
c. Diagnostic decisions: which specific difficulties account for the information management student's inadequate progress so a lecturer can remediate learning progress and design a more effective instructional plans.

In the Faculty of Information Management, Universiti Teknologi MARA (FIM), a series of assessment in the form of tests, quizzes, and final examination were designed to validate such LO's which is primarily defined by Bloom's Taxonomy for skills measurement [2]. It has been argued that, a good education system should generate graduates in information management who are 'ingenious'; that are also able to think creatively, take calculated risks and adopt exploratory attitudes; collectively termed as affective or generic skills. A graduate is deemed to be of competence when they possess good interpersonal skills, oral and written communication, leadership skills, teamwork, problem solving, creativity and sound computer literacy. This behavioral measurement on students' affective conduct is sometimes not given enough emphasis. Response from the industry could provide clear indication of any gaps on graduates' generic skills since this is crucial for them to secure jobs in an already very stiff market. In view of its importance, the need for skill based measurement is already gaining wider attention within the institution of higher learning community.

FIM, for example, observed the MQF Guideline, Malaysian Institution of Higher Learning Quality Assurance (JK-IPTA, 2005) which provides the LO framework to be assessed. Thus, a thorough method of measurement; both cognitive and affective, is is provided by the Guideline to gauge the achievement of the expected Learning Outcomes (LO) of the subject effectively.

This paper presents a model of performance measurement for such abilities using Rasch-based evaluation tool to establish their Learning Capability Indicator; CL_i which can be used to decide the necessary course of action to achieve the desired level of competency through improved and more effective instructional plans. The assignment were evaluated on how well they relate to the content domain being assessed as stated in the course LO and results were analyzed to determine whether a gap exist in the information management student's capabilities or psychological construct that is supposedly to be developed.

The model employed a simple framework where an assessment form utilizes the Table of Test Specification designed on Bloom's Taxonomy parameters showing each dimension of the ability to be assessed.

2 Background: An Overview on Performance Assessment

Performance assessment is relatively new, undeveloped and yet to be studied systematically. Many practitioners are experimenting with its use and contributing to its development and refinement. These assessments take a variety of forms, some of which are closer than others in approximating the conceptual and theoretical underpinnings of performance assessment. Fuchs (1995) explained in great details the criteria of a good assessment in education, but not limited to;

a. Measure important LO
b. Address the prime purpose of assessment.
c. Provide clear descriptions of student performance that can be linked to instructional actions.
d. Compatible with a variety of instructional models.
e. Easily administered, scored and interpreted by fellow lecturers.
f. Communicate the expected learning outcome to Lecturers or Tutors as well as students.
g. Generate accurate, meaningful information i.e., be reliable and valid.

Performance assessment can be viewed in the correlational ABC Model on how cognitive skills and affective state is reflected in the behaviour of students during learning. Weybrew (1992) discussed at length on the repercussion of such development but believed that affective values is of significant importance in neuro-linguistic programming otherwise popularly known as NLP [3].

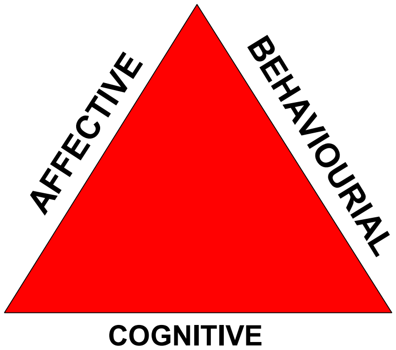

Figure.1 ABC Model
Source: Weybrew, 1992

3 Measurement Methodology

This study addresses the following questions:

a. Does this Rasch-based model result in more accurately classified examinees?
b. Is Bloom's-based Table of Test Specification effective in generating meaningful information on the student's ability? and,
c. How many examinees are needed to satisfactorily calibrate this measurement model to competency?

A method of defining the required metrics in an institution of Higher Learning Performance Assessment is setforth modelled on Aziz (*2004*) Plan – Implement – Check – Evaluate (P-I-C-E) assessment method to measure the effectiveness of a system performance [6]. This model is found very much agreeable to Shewhart's (*1939*) P-D-S-A Cycle which was subsequently developed into the infamous Deming's (*1954*) P – D – C – A Cycle by the Japanese industrial community. Then, in year 2000 in Geneva, this fundamental concept was adopted by the international community through the implementation of ISO9000:2000. McLelland (1995) developed the Knowledge-Skill-Ability-Others (KSAO) model on the same format to define competency.

The statistical approach employed is simple yet it can yield very accurate findings using data-driven approach to analyse the root causes of each learning problem encountered. It is a very disciplined approach for assessing students' ability during a learning process. Cognitive skills are clustered on and extended to include the affective skills; i.e. communication skill, teamwork, life long learning etc. These skills shall be termed as dimensions. Bloom's taxonomy is applied for this purpose.

Within these dimensions, relevant main areas or attributes, related to the learning outcome is then identified but not limited to viz; vocabulary power, technical appreciation, software development and resourcefulness. Collectively this is known as attributes which are measurable.

Table 1 shows the conceptual format of the designed course LO. It has dual purpose; first is to define the course LO and enumerate the focus of teaching and, second, serve as a guide for the preparation of an assessment.

Table 1. Pro-Forma Learning Outcome Map

COURSE LO'S MAP									
Course: ABC1234 Course XYZ Sem:02 Yr:2007									
Topics	**M**	**LO'S**							**W%**
		1	**2**	**3**	**4**	**5**	**6**	**7**	
Topic A									
Sub-topicA1		3				2			45
Sub-topicA2			1			2			
Sub-topicA$_n$			2	1					
Topic B									
Sub-topicB1		2			2				55
Sub-topicB2			2	1					
Sub-topicB$_n$				2	1				
Σ COURSE LO%		10	25	30	20	15	-	-	100

TEACHING FOCUS: 1-Major 2-Mediocre 3-Minor
M – Marks W - Weightage

Each topic to be taught will be given a Teaching Focus rated 1 as Major, 2 for Mediocre and 3 for Minor. This provides a guide for lecturers to prepare their teaching plan focus, on how much time to be spent for each topic covered. The column on the right most; W% – show the spread of focus given; say Topic A: Introduction; less time whilst Topic B – Theory A; is given more time. The bottom row shows the spread of teaching focus; which LO is the centre of attention. Inverse diminishing scale, $\frac{1}{2^n}$ is used to calculate the Total Sum of Course LO thus giving Major; $\frac{1}{2^0}=1$, Mediocre; $\frac{1}{2^1}= 0.5$ and Minor; $\frac{1}{2^2}=0.25$ respectively.

Column marked M is used as control for mark distribution during the design of assessment questions. At this point, the course LO is turned into the Table of Test Specification based on Bloom's Taxonomy. ABET method is used to transform the LO's into Blooms Taxonomy [4]. Refer Table 2 for transformation matrix.

Table 2. Transformation of LO Into Bloom's

TABLE OF TEST SPECIFICATION Course: ABC1234 Course XYZ Sem:02 Yr:2007								
Topics	M	BLOOM'S TAXONOMY						W%
		BT1	BT2	BT3	BT4	BT5	BT6	
Topic A								
Sub-topicA1		3	2					45
Sub-topicA2			1			2		
Sub-topicA$_n$			1	2				
Topic B								
Sub-topicB1				2	2			55
Sub-topicB2			2	1				
Sub-topicB$_n$		1			2			
QUESTION SPREAD %		**15**	**35**	**25**	**15**	**10**	**-**	**100**

This is the vital pivot point on the scale of measurement academician must observe. First, we have put all the items on the same criteria of measurement. Bloom's Taxonomy is the common reference criteria used in education. Second, the lecturers have all the freedom to decide on the distribution of questions to be set but using Table 3. Table of Test Specification percentage of question spread as the framework of questions asked.

Table 3. Table of Test Specification

BLOOM'S vs LO Matrix Course: ABC1234 Course XYZ Sem:02 Yr:2007								
	LO'S							W%
	1	2	3	4	5	6	7	
BLOOM'S TAXONOMY								
Knowledge -**BT1**	3		1					**15**
Understanding -**BT2**	2	1	1	2				**35**
Application -**BT3**		2		1	2			**25**
Analysis -**BT4**			2		2			**15**
Evaluation -**BT5**		2						**10**
Synthesis -**BT6**								**-**
Σ COURSE LO %	**10**	**25**	**30**	**20**	**15**	**-**	**-**	**100**

Upon completion of the test, results were analysed for reliability using Rasch Separation Index –***G*** and Item Response Theory Discrimination Index '***a***' for construct validity. Hence, data can be duly analysed for further meaningful interpretation.

However, raw score obtained from quizzes, tests, assignments and final exam are continuum scale. It has an order but does not possess an interval scale. Hence, further evaluation using the raw score is rather complex to be carried out. Raw score has the limitation of telling exactly the extend of skill development. A student who scores 80% cannot be deduced to be twice as smart as a student who scores only 40%. A test result cannot classify examinees into the correct ability group for remedial measures. Cronbach-alpha and Factor analysis works on historical data. Rasch Model is a predictive model to give an indicator on a scale where the student is in a cohort. All the error in measurement is absorbed in Rasch Model which uses '*log odd unis' (logit)* as the measurement unit thus transforms the assessment results into a more accurate linear correlation.

In this exhibit, students were given an assignment where the assessment form is designed and developed for the attributes which is rated based on an even number scale of 1 – 4 dichotomously indicating **NO – YES** with 2 – 3 indicating their level of agreement to an attribute. This assessment form gathers empirical data as the main instrument of this study.

Dimension A, B…*n,* are the skills to be assessed based on defined Bloom's Taxonomy; i.e. ability to acquire sound knowledge in Information Management, understanding given scenarios and ability to apply appropriate knowledge; etc. The attributes are finite skills within the dimensions. In writing, it would be grammatical order, logic flow or reasoned arguments. Thus, a holistic discrete method of measurement is developed to enable the respective mean, $\bar{x}$, values for each skill can be established [5]. These values will serve as an indicator and gives a locii on the quality level of the subject LO's.

Table 4 shows the simple computation of an assessment. The lecturer will give his evaluation on the student's performance using the prescribed form. He will give his own weightage, ***W*** for each dimension. This allows flexibility and freedom for each lecturer to make his own evaluation. This is vital because the lecturer is free to set his own criteria of assessment and let the student know what is expected from the assignment. Let us assume, the skills to be assessed are **BT2 -**Understanding and **BT3** -Application.

Next each attribute is given a grade. In this exhibit, the grade of attributes in Dimension A – **BT2** is totalled up:

$$\text{Attribute } A_n = 2+3+4 = 9 \quad (1)$$

Table 4. Simulated Assessment

PERFORMANCE SCORE FORM Student: *XXX YYY* Date: ddmmyy						
Ratings	**W**	**GRADE**				**W*G**
		1	**2**	**3**	**4**	
Dimension A- BT2	40					0.4x3=1.2
Attribute A_1			_/			Ave.S= $\frac{2+3+4}{3}$ = 3
Attribute A_2				_/		
Attribute A_n					_/	
Dimension B- BT3	60					0.6x3=1.8
Attribute B_1			_/			Ave.S= $\frac{2+3+4}{3}$ = 3
Attribute B_2				_/		
Attribute B_n					_/	
	Sum (W*G)					1.2+1.8 = 3.0

The sum of the grades is then averaged out; to give,

$$\textbf{BT2} \text{ average score, } 9/3 = 3 \qquad (2)$$

This raw average score; $BT2_{avS}$=3 is adjusted by multiplying ***W***, weightage to give the true score for the said dimension;

$$\mathbf{BT2}_{tS}\text{ , } \boldsymbol{W}{*}G = 0.4 \times 3 = 1.2 \qquad (3)$$

Finally, each Dimension score is summed up to give the true score the student obtained for his / her assignment;

$$\text{Sum } (\boldsymbol{W}{*}G) = (0.4 \times 3) + (0.6 \times 3) = 1.2 + 1.8 = 3.0 \qquad (4)$$

This true score is more reflective of the students' skill achievement rather than an arbitrarily assessed. Now we can pin-point exactly the finite attribute and the respective BT_N assessed.

Table 5 shows a further simulated computation to establish the learning outcome index of the subject; LO_i . Let us assume the total number of student in a cohort is 100, hence, *N* of the population surveyed; *N*=100.

Imagine, the spread of *N* for each attribute given the grade $G_{1\text{-}4}$ is;

Attribute A_1, *N*: 15, 30, 35 and 20
Attribute A_2, *N*: 25, 40, 25 and 10; etc.

Next, this value of *N* is multiplied to each respective grades;

$$15 \times \mathbf{1} = 15,\ 30 \times \mathbf{2} = 60,\ 35 \times \mathbf{3} = 105,\ 20 \times \mathbf{4} = 80 \qquad (5)$$

This gives a total sum scored of ;

$$15 + 60 + 105 + 80 = 260 \qquad (6)$$

The total sum scored for Attribute $A_{1,N=100}$ =400, is then moderated against the full score in an ideal scenario when everybody is assumed to be totally excellent.;

$$\frac{260}{400} = 0.65 \qquad (7)$$

Table 5. Computation Of LO_i

LEARNING OUTCOME ANALYSIS Student: *XXX YYY* Date: ddmmyy						
Rating	**1**	**2**	**3**	**4**		$\bar{x}$
Dimension A –BT2						
Number of N	15	30	35	20	100	
Attribute A_1 Score Obtained	15	60	105	80	$\frac{260}{100}$ =2.40	0.65
Number of N	25	40	25	10	100	
Attribute A_2 Score Obtained	25	80	75	40	$\frac{220}{100}$ =2.20	0.55
	LO Mean, $\bar{x}_i$ =				$\frac{2.40+2.20}{2}$ = 2.3	
BT2-LO$_i$: Learning Outcome Score					$\frac{2.30}{4}$x 100 = **57.50%**	

The mean value $\bar{x}$ for this particular attribute is obtained by multiplying this value of 0.65 to the expected full score of 4 to proportionately yield a value of 2.40.

This process is repeated for all the other attributes to give each mean value of $\bar{x}$;

$$\frac{\text{Sum of } \bar{x}}{\text{Total number of attributes}}$$

$$\frac{2.40+2.20}{2} = 2.30 \qquad (8)$$

Next, proportionate the result by 4, being the full score, to obtain the subject Learning Outcome score, $\mathbf{LO}_i$ is;

$$\frac{2.30}{4} \times 100 = 57.50\% \qquad (9)$$

Now the students' LO achievements can be tabulated. A hypothetical LO result is shown in Table 6. Competency achieved by each student for each LO can therefore be dissected.

This is where assessment by skilled based is more useful. All students may appear to have passed the exam. If we set 70% as the threshold of competency, then Student D is having a problem. Similarly for the lecturer, he has problem developing BT3 and BT5.

Table 6. LO Analysis

TABULATION OF MARKS BY LO

Course: ABC1234 Course XYZ Sem:02 Yr:2007

Students	LEARNING OUTCOMES SCORE						MARKS
	BT1	BT2	BT3	BT4	BT5	BT6	
1. Azrilah	86	92	88	80	78	81	**84**
2. Sohaimi	82	84	78	82	74	72	**76**
3.Saidfudin	80	82	76	80	80	70	**73**
4. Student D	77	76	72	68	66	68	**72**
5. Student E	81	72	68	72	76	74	**72**
6. Student F	76	74	68	70	66	72	**63**
7. Student G	73	72	64	72	70	70	**62**
8. Student H	78	71	63	71	64	71	**60**
LO mean	**79**	**78**	**72**	**74**	**72**	**72**	

4 Discussion

A comprehensive pro-forma evaluation for the required LO known by dimensions and attributes can be prepared to meet MQF evaluation requirement. Student learning capability is indexed as an indicator to the subject LO using Rasch Model.

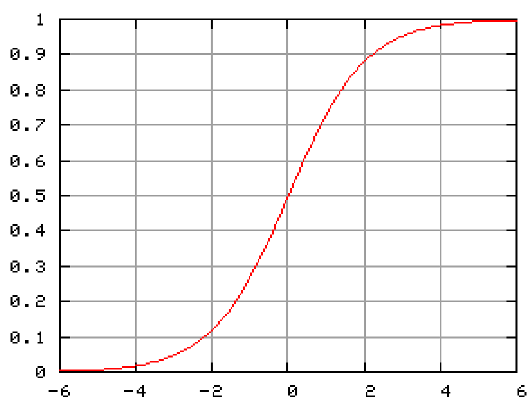

Table 7. Rasch Model Conversion Table

In this case Rasch Model can be simplified as [6];

Probability of (Possessing LO)	=	Student's Ability	-	Difficulty of given task

The Pr(PLO) value can be derived from score obtained in Table 6; hence the difficulty index. Rasch Model enable each of the Students' Ability; thus students' skill development to be clearly identified by each competency trait. Symptoms can be traced more effectively and treated specifically.

This will help guide academicians to respond with certainty on the nature of corrective actions to be taken. Teaching methods may be reviewed and new approach is developed and tested. Teaching style may need some innovation to stimulate learning ambience.

5 Conclusion And Recommendations

This simple yet prudent conceptual theoretical framework is capable of providing more comprehensive view but specific and objective evaluation. This measurement model results in more accurately classified examinees. The students' competency gap with industry expectation can now be put under better control. Though the measurement model is able to show reliably accurate result even with small number *N*, the dimensions affecting the performance of a teaching method shall be subjected to further study. The attributes in relation thereof which has material effect on the teaching and learning system must be subsequently researched thoroughly in the near future to give a more accurate account of the subject LO.

References:

[1]. Fuchs, Lynn S."*Connecting Performance Assessment to Instruction: A Comparison of Behavioral Assessment, Mastery Learning, Curriculum-Based Measurement, and Performance Assessme*nt: ERIC Digest, 1995

[2]. Hopkins; C.C. et al, "Classroom Measurement and Evaluation", 3rd edition, Peacock, 1990.

[3]. Weybrew, Benjamin B.; *The ABC's of Stress: A Submarine Psychologist's Perspective*, Praeger Publishers, 1992

[4]. Besterfield-Sacre, M., Shuman, L.J., Wolfe, H., Atman, C.J., McGourty, J., Miller, R.L, Olds, B.M. and Rogers, G.M., "*Engineering Education Assessment Methodologies and Curricula Innovation*" University of Pittsburg; Jan 2000

[5]. Aziz, A. and Saidfudin, MS.; "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.*

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