IDENTIFYING POTENTIAL MISFIT ITEMS IN COGNITIVE PROCESS OF LEARNING ENGINEERING MATHEMATICS BASED ON RASCH MODEL

Sh Ataei[1]
PhD Student, *Faculty of Computer and Mathematical Sciences*
Universiti Teknologi MARA Shah Alam, Selangor, Malaysia
Z Mahmud[2]
Assoc Prof Dr, *Faculty of Computer and Mathematical Sciences*
Universiti Teknologi MARA Shah Alam, Selangor, Malaysia
M N Khalid[3]
Assoc Prof Dr, *Faculty of Computer and Mathematical Sciences*
Universiti Teknologi MARA Shah Alam, Selangor, Malaysia

E-mail: sholehataei@gmail.com, zamal669@salam.uitm.edu.my,
mnasir@tmsk.uitm.edu.my

Abstract. The students learning outcomes clarify what students should know and be able to demonstrate after completing their course. So, one of the issues on the process of teaching and learning is how to assess students' learning. This paper describes an application of the dichotomous Rasch measurement model in measuring the cognitive process of engineering students' learning of mathematics. This study provides insights into the perspective of 54 engineering students' cognitive ability in learning Calculus III based on Bloom's Taxonomy on 31 items. The results denote that some of the examination questions are either too difficult or too easy for the majority of the students. This analysis yields FIT statistics which are able to identify if there is data departure from the Rasch theoretical model. The study has identified some potential misfit items based on the measurement of ZSTD where the removal misfit item was accomplished based on the MNSQ outfit of above 1.3 or less than 0.7 logit. Therefore, it is recommended that these items be reviewed or revised to better match the range of students' ability in the respective course.

1. Introduction

According to Kilpatrick & Swafford [1], for students to be successful in later mathematics endeavours and to use mathematics effectively in life, they must have a sound learning of mathematics concepts. Mathematics helps to strengthen and to accelerate students understanding in many varieties of engineering courses.

There are many ways to assess learning. Among them are the traditional and conventional methods such as currently practiced by educators in the current education system. Assessment comes in various forms and methods. In the current Outcome-Based Education (OBE) system, measurement of learning is gauged against the course learning outcome while integrating Bloom's Taxonomy concept in the measurement process. However, it is still unclear as to the degree of learning that has actually taken place with the current assessment methods. Therefore, having one measurement tool that can help to measure one domain of Bloom's Taxonomy is important.

One such tool is based on the IRT (Item Response Theory) model, namely the Rasch measurement model [2]-[3]-[4]-[5], [6]. It is a scale measurement which is estimated by responses of persons and items' properties [7]. According to some researchers [8]-[9]-[10]-[11], [12], this measurement is able to measure spread and separation of the items and task areas based on Bloom's Taxonomy.

Item Response Theory (IRT) tries to apply the mathematical model in order to anticipate the probability of a person of certain ability attempting an item of certain difficulty successfully. This would depend on the log transformation of the difference between logit difficulty of item and logit ability of a person.. In addition, Person- Item Distribution Map (PIDM) in the Rasch model shows the logit transformation of raw score of person ability and item difficulty that both are calibrated against a single logit ruler within the range. [5]- [9]-[11]-[13]-[14]-[15], [16].

2. Measurement models

Assessment plays an important role in education measurement. A suitable measurement should be able to (a) provide linear measures, (b) overcome missing data, (c) give precision estimates, (d) detect misfits and (e) item or task to be completed and person must be separated from each other [5]. Therefore, the selection of proper tools for assessing mathematics learning is through the application of the Rasch measurement model which is expected to respond to all the properties above. Rasch model is a mathematical model that is applicable for ascertaining certain properties in data. The Rasch model expresses the relation between the probability of completing a task, the ability of persons and the difficulty of the task. A Rasch model transforms raw scores (ordinal data) into a linear interval scale data.

2.1. Dichotomous Rasch model

Dichotomous data is acceptable to be fitted into the Rasch model where the difference between two parameters namely, person ability (δ_i) and item difficulty (θ_n) are governed by a single logit scale as a result of the log transformation of the probability of completing a task given the ability of a person and the difficulty of the task or item given. The Rasch model is the simple logit model of IRT such as the Dichotomous Rasch model which has two options for answering either a yes (1) or no (0), correct or incorrect, agree or disagree. The Rasch measurement model function based on Rasch is illustrated as follows:

θ_n is a parameter: Person ability for responding to questions (items) and δ_i is a parameter: Difficulty of questions(items) for answering.

$$P(X=1)=\frac{\exp(\theta_n-\delta_i)}{1+\exp(\theta_n-\delta_i)} \tag{1}$$

$$P(X=0)=1-P(X=1)=\frac{1}{1+\exp^{(\theta_n-\delta_i)}} \tag{2}$$

$$\ln\left[\frac{\dfrac{e^{\theta_n-\delta_i}}{1+e^{\theta_n-\delta_i}}}{1-\left[\dfrac{e^{\theta_n-\delta_i}}{1+e^{\theta_n-\delta_i}}\right]}\right]=\theta_n-\delta_i \tag{3}$$

Correct response probability, incorrect response probability and parameter separation are achieved in Equations (1), (2) and (3). [3]- [4], [5]- [6]-[17], [18].

2.2. Simple Rasch Logistic model

The simple Rasch model (SRM; Rasch, 1960/1980) for the analysis of test data provides a way to place persons and items on a scale with a clear probabilistic interpretation of distance on the scale. Let the attitude of a person be characterized by the variable and the affective value of item. Item difficulty is 0.5 when person ability and item difficulty are equal.

The success probability is a function of the difference of person ability and item difficulty where $\log \frac{p}{1-p} = \theta_n - \delta_i$, $\theta_n - \delta_i$ is the difference between person ability and item difficulty.

If two persons are of different abilities but are attempting the same item $\delta_1 = \delta_2 = \delta_i$, (a) first person will have $\log \frac{p}{1-p} = \theta_1 - \delta_i$, and (b) the second person will have $\log \frac{p}{1-p} = \theta_2 - \delta_i$ and so on. This Rasch property is very important because we can achieve item difficulty without considering specific person or achieve person difficulty without considering specific items.

For example, in determining differences between two persons (item free), it will be

$$\log \frac{p_1}{1-p_1} - \log \frac{p_2}{1-p_2} = \theta_1 - \delta_i - (\theta_2 - \delta_i) = \theta_1 - \theta_2 \tag{3}$$

These equation represents probability of success where probability of success $= f(\theta - \delta)$ (4) [4]-[5]-[6]-[19]-[20]-[21] and [22].

2.3Misfit

In Rasch measurement, fit statistics is able to help us expose the discrepancies between observe real data we have gathered in practice and expected data from Rasch model.

On the other hand misfit data is defined as items which fall outside certain range of acceptable fit. Green and Frantom [2] and [3] defined item fit as an index of items function which can provide a continuum of answers. In this study, we use outfit that is based on the sum of squared standardized residuals. Consider X is an observation; E is its expected value underpinning Rasch parameter estimates while σ^2 is its modelled variance on its expectation. Therefore, the standardized unweighted mean square outfit is below:

$$t_i = (u_i^{\frac{1}{3}} - 1)(\frac{3}{q_i}) + (\frac{q_i}{3}) \tag{5}$$

when u_i is fit mean square and q_i is fit standard error[3]-[4]-[20],[21].

Green and Frantom [2]believed that outfit happens if mean square lies between 0.7 and 1.3.The standard deviation of the standardized outfit is an index of overall misfit for persons and items when it be bigger than +2 and less than -2 [3]- [4]-[20]-[21],[22].

3. Data Analysis and Results

The results of engineering students in a Calculus III course from the Faculty of Mechanical and Electrical Engineering, Universiti Teknologi MARA are discussed in this study. In order to examine students' understanding of Calculus III, the examination questions were prepared according to the six Bloom's taxonomy cognitive levels. The questionnaire consists of 31 Calculus III multiple choice questions and 72 attitude items which were administered to 54 students.

3.1. Summary statistics

Table 1 shows the summary of 54 measured persons and 31 measured items. It shows the mean measured person at -0.24 logit and mean measured item at 0.0 logit. This indicates the test is slightly difficult for the majority of the students. Reliability index for a person is measured at 0.78 logit which indicates a moderately acceptable fit with a separation index of about 2.0. This indicates a moderately acceptable fit with a person's ability ranging from high to low and vice versa. The reliability index for the item is measured at 0.88 logit with a separation index of about 3.0. This suggests a slightly broader continuum for items than persons.

Table 1. Summary of 54 measured person.

	SCORE	TOTAL COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	12.4	27.5	-.24	.46	.98	.0	.95	.0
S.D.	3.9	4.1	1.07	.17	.20	.9	.34	.7
MAX.	19.0	31.0	4.13	1.30	1.36	1.9	2.79	3.6
MIN.	5.0	14.0	-1.87	.39	.25	-2.0	.06	-1.2

REAL RMSE .51 TRUE SD .95 SEPARATION 1.87 Person RELIABILITY .78
MODEL RMSE .49 TRUE SD .95 SEPARATION 1.92 Person RELIABILITY .79
S.E. OF Person MEAN = .15

Table 2. Summary of 31 measured Item.

	SCORE	TOTAL COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	21.6	47.9	.00	.35	1.01	.1	.97	.1
S.D.	8.5	3.4	1.19	.15	.15	.9	.16	.8
MAX.	38.0	53.0	4.69	1.13	1.58	1.9	1.34	2.3
MIN.	1.0	38.0	-1.90	.29	.83	-1.6	.69	-1.0

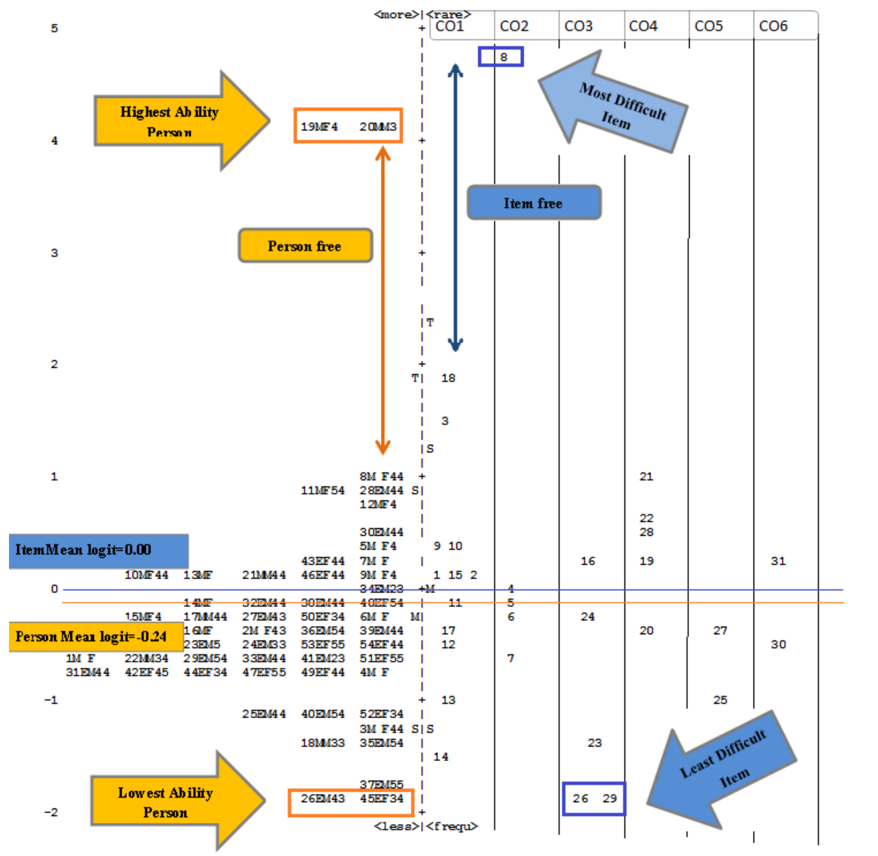
REAL RMSE .42 TRUE SD 1.12 SEPARATION 2.67 Item RELIABILITY .88
MODEL RMSE .38 TRUE SD 1.13 SEPARATION 2.95 Item RELIABILITY .90
S.E. OF Item MEAN = .22

3.2 Model Fit

3.2.1. Person-Item- Distribution Map

Figure 1 indicates the person-item distribution map which is the heart of Rasch analysis. The map shows that persons are distributed on the left side of the logit ruler and items on the right side of the logit ruler. In Figure 1, it can be seen that a person's mean logit falls slightly below item mean logit and also there are gaps on the left and right side in upper mean logit line that are known as person and item free respectively. These indicate the test is slightly difficult for students with limited ability.

The map shows that two students from the Faculty of Mechanical Engineering are located at 4.0 logit, seen as most able students while two from the Faculty of Electrical Engineering located at -1.8 logit are seen as least able. It is observed that item C8 of Bloom's Taxonomy comprehension level is the most difficult item while items C26 and C29 at the application level are considered easiest by the majority of the students.



BLOOMS TAXONOMY
CO1: Knowledge
CO2: Comprehension
CO3: Application
CO4: Analysis
CO5: Synthesis
CO6: Evaluation

Figure1. Person-Item Distribution Map for learning Calculus III across Bloom's Taxonomy levels.

3.2.2. Potential misfitting item

In Figure 2, the items are shown as circles and the size of the circles indicates standard errors of measurement along the latent variable. Items were plotted to items difficulty (y-axis) and outfit ZSTD statistics (x-axis). An item is considered as potential item misfit if plot range is outside of -2 to +2 or circles are big [20], [22].

In Figure 2, all items are within the -2 and +2 range (item difficulty), except item 8 (item point :(0.58, 4.69)) and the size of points is 1.13. As can be observed measure difficulty of item 8 is $4.6 > +2$ and the largest standard error with the largest circle relates (1.13) to Item 8, indication that Item 8 is potential item misfit. As a mentioned in Figure 1, Item 8 is the most difficult item where all students were not able to answer it correctly. Figure 2 shows Items 16 (item point:(2.3,0.25), size:0.33) and also ZSTD of item 2(item point:(1.92,0.1),size:0.31) is near 2 that indicates items misfit.

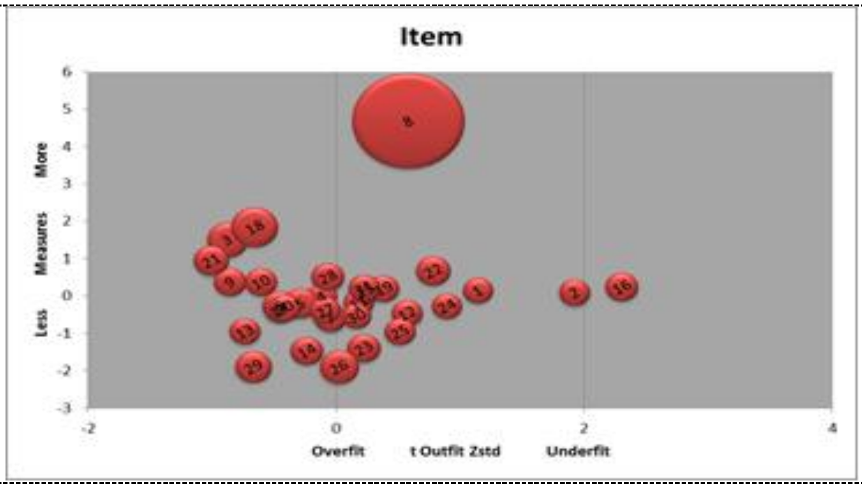


Figure 2 . Bubble chart – Students' cognitive learning In Calculus III

From Table 2, item 16 could be considered a misfit item because its standard deviation of the standardized outfit is above +2.0 and its MNSQ value is close to 1.4. Attention should be given to this item as it could allow for possible unusual responses. According to Table 2 we cannot consider item 8 as a misfit item. Similarly, Liu shows that some potential items misfit cannot be misfit items [22].

Table 2. Summary of 31 Measured Items

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
8	1	45	4.69	1.13	1.58	.9	1.17	.6	A .13	.42	97.8	97.7	8
16	16	44	.25	.33	1.23	1.9	1.34	2.3	B-.13	.27	56.8	66.1	16
1	21	52	.15	.30	1.17	1.6	1.27	1.1	C .19	.36	63.5	66.7	1
2	19	48	.10	.31	1.18	1.9	1.22	1.9	D-.06	.26	52.1	63.9	2
22	13	47	.68	.34	1.12	.8	1.15	.8	E .08	.27	72.3	73.1	22
19	20	51	.21	.31	1.11	1.0	1.07	.4	F .27	.36	60.8	67.6	19
26	37	46	-1.90	.39	1.09	.5	.98	.0	G .17	.25	80.4	80.5	26
24	23	48	-.26	.30	1.06	.7	1.09	.9	H .20	.30	62.5	61.8	24
25	30	49	-.95	.31	1.03	.3	1.06	.5	I .23	.28	63.3	65.7	25
12	24	47	-.46	.31	1.05	.6	1.06	.6	J .24	.30	53.2	62.2	12
23	35	49	-1.39	.33	1.05	.4	1.03	.2	K .20	.27	73.5	72.5	23
31	18	49	.23	.31	1.05	.5	1.02	.2	L .22	.28	63.3	66.3	31
15	20	51	.16	.31	1.04	.4	1.04	.2	M .33	.36	70.6	67.8	15
11	24	51	-.17	.30	1.02	.3	1.03	.2	N .32	.35	64.7	64.2	11
30	27	51	-.53	.29	1.00	.1	1.01	.2	O .30	.30	64.7	62.3	30
7	24	43	-.58	.33	1.01	.1	.94	.0	P .33	.32	62.8	63.9	7
4	21	51	-.01	.30	1.01	.1	.98	-.1	o .29	.30	64.7	64.3	4
27	23	47	-.38	.31	.97	-.4	.99	-.1	n .33	.29	72.3	61.6	27
28	14	46	.52	.33	.95	-.3	.98	-.1	m .32	.26	78.3	70.6	28
5	18	38	-.19	.35	.96	-.4	.87	-.3	l .41	.37	65.8	64.6	5
10	18	49	.38	.32	.93	-.5	.85	-.6	k .43	.36	67.3	69.2	10
14	38	52	-1.46	.33	.93	-.4	.82	-.2	j .34	.26	78.8	74.3	14
17	27	53	-.36	.29	.93	-.9	.86	-.5	i .40	.33	66.0	63.5	17
20	25	49	-.31	.31	.91	-1.0	.86	-.4	h .41	.33	71.4	63.5	20
6	19	39	-.26	.35	.90	-1.0	.82	-.5	g .44	.36	69.2	64.2	6
9	18	49	.38	.32	.88	-1.0	.80	-.9	f .48	.36	71.4	69.2	9
3	9	49	1.51	.41	.87	-.4	.69	-.9	e .56	.44	85.7	85.1	3
29	36	46	-1.89	.37	.86	-.7	.82	-.7	d .47	.25	78.3	78.3	29
21	13	50	.95	.35	.86	-.7	.73	-1.0	c .53	.40	78.0	77.6	21
13	31	50	-.94	.31	.84	-1.6	.79	-.7	b .46	.28	76.0	66.4	13
18	7	46	1.85	.47	.83	-.4	.69	-.7	a .60	.46	89.1	88.3	18
MEAN	21.6	47.9	.00	.35	1.01	.1	.97	.1			70.2	69.8	
S.D.	8.5	3.4	1.19	.15	.15	.9	.16	.8			9.9	8.5	

4. Conclusion

This study has demonstrated that proper measurement of students' learning of Calculus III can be measured using the Rasch measurement models**.** Rasch models clearly have an advantage to accurately measure students' learning of Calculus III across the six cognitive levels of Bloom's Taxonomy.

Rasch measurement is also able to assess the reliability of the person and item measure and identify any misfits due to unusual responses given by the students. Calibration of items and students' responses using PIDM map has demonstrated the ability level of students against the difficulty of the items. This provides an excellent diagnosis to gauge students' learning of Calculus III. The use of misfit items is able to detect not only some dimensionality problem in the items but also determine which items of test should be deleted. Also, this study shows that potential item misfit can be different from item misfit. Generally, the empirical data moderately fit the expected Rasch model. Based on the Rasch measurement, students find that some Calculus III exam questions relating to knowledge and comprehension level are too difficult but they are can answer questions relating to application level better. The analysis also showed that some of the examination questions were either too difficult or too easy for the majority of the students.

In future study, we will investigate why engineering students are able to answer questions of application level without understanding the relative concepts. Also, having discussed about the relationship between potential item misfit and item misfit via tables and figures, the strongest arguments exists with respect to their mathematical formulas.

Acknowledgment

The authors would like to thank Universiti Teknologi MARA for funding this research under the Research Intensive Faculty grant scheme No. 600-RMI/DANA 5/3/RIF (866/2012). This research would not be possible without the university's assistance.

References

[1] Kilpatrick J, Swafford J, and Findell B 2001 *Adding it up: Helping children learn mathematics* (National Research Council (U.S.) Mathematics Learning Study Committee National Academies Press)

[2] Green K E, and Frantom C G 2002 Survey *Development and Validation with the Rasch Model. International conference on questionnaire development, Evaluation, and Testing* (Charleston, SC)

[3] Bond T G and Fox C M 2007 (2nd Ed.) *Applying the Rasch Model Fundamental Measurement in the Human Services* (Mahwah, NJ, Erlbaum)

[4] Rasch G 1960/1980 Probabilistic models for some intelligence and attainment tests Copenhagen, Danish Institute for Educational Research),ed B D Wright 1980 (The University of Chicago Press)

[5] Wu M and Adams R 2007 *Applying the Rasch Model to Psycho-Social Measurement: A Practical Approach*

[6] Alagumalai S, Curtis D D and Hungi N 2005 *Applied Rasch Measurement A Book of Exemplars: Papers in Honour of John P. Keeves* (Springer)

[7] Embretson S E and Reise S P 2000 *Item Response Theory for Psychologists* (Mahwah, NJ: Erlbaum)

[8] Othman H, Asshaari I, Ismail N A and Nopiah Z M 2011 *Evaluating the Reliability And Quality Of Final Exam Questions Using Rasch Measurement Model: A Case Study Of Engineering Mathematics Courses,* Seminar Pendidikan Kejuruteraan Kongres Pengajaran dan Pembelajaran and Alam Bina (PeKA'11)

[9] Kamsuriah A, Nazlena M A and Suhaila Z 2011 *An Improved Course Assessment Measurement for Analyzing Learning Outcomes Performance Using Rasch Model*, Procedia Social and Behavioral Sciences (Elsevier Lt) **18** 442-449

[10] Mahmud Z 2011 Diagnosis of Perceived Attitude, Importance and Knowledge in Statistics based on Rasch Probabilistic Model *Int.J. Applied Mathematics and Informatics* **5** 291-298
[11] Riccardo D N 2010 *Measure School Learning Through Rasch Analysis: The Interpretion of Results,* Procedia Social and Behavioral Sciences (Elsevier) **9** 373-377
[12] Mahmud Z, Musa N A C, Ghani N A M and Rahim A M 2012 Probabilistic Approach in Examining Quality of Survey Response Data in Statistical Education Research *International J. Mathematical Models and Methods in Applied Sciences* **6** 182-190
[13] Abdul Aziz A, Masodi M S, Mohamed Ibrahim F, Omar M Z and Zaharim A 2013 Insight into Engineering Education Learning Outcome's Assessment with Rasch Model Research *J. Applied Sciences, Engineering and Technology* **6** 3520-3526
[14] Asshaari I, Othman H, Bahaludin H, Ismail N A and Nopiah Z M 2011 *Rasch measurement in evaluation of Blooms' separation: A case study in Engineering mathematics III Course,* Seminar Pendidikan Kejuruteraan Kongres Pengajaran dan Pembelajaran and Alam bina
[15] Osman S A, Naam S I, Jaafar O, Hamidon W, Badaruzzaman W and Rahmat R A A 2012 *Application of Rasch Model in Measuring Students' Performance In Civil Engineering Design II Course* Procedia Social and Behavioral Sciences (Elsevier) **56** 59–66
[16] Osman S A, Badaruzzaman W H, Hamid R, Taib K, Khalim A R, Hamzah N and Jaafar O 2011 *Assessment on Students Performance Using Rasch Model in Reinforced Concrete Design Course Examination,* Proceedings of 10th WSEAS International Conference on Education and Educational Technology
[17] Andrich D 1988 *Rasch Model for Measurement* (Sage)
[18] Wilson M 1988 Detecting and Interpreting Local Item Dependence Using a Fannily of Rasch Models *Applied Psychological Measurement* **12** 353-364
[19] Gori E 2006 The Rasch Approach To "Objective Measurement'm The Presence Of Subjective Evaluation From"Judges" *J. Educación XXi* **9** 107-133
[20] Linacre J M 2007 *A User's Guide to WINSTEPS Rasch-model Computer Programs* (Chicago: MESA Press)
[21] Wright B D, Masters GN 1990 *Rasch Measurement Transaction (2nd ed)* (Wilmington, Delaware Wide Range, INC)
[22] Liu X 2010 *Using and developing measurement instruments in science education: A Rasch Modeling approach* (Charlotte NC: Information Age Publishing)