International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:8, No:1, 2014

Validity and Reliability of Competency Assessment Implementation (CAI) Instrument Using Rasch Model

Nurfirdawati Muhamad Hanafi, Azmanirah Ab Rahman, Marina Ibrahim Mukhtar, Jamil Ahmad, Sarebah Warman

***Abstract*—**This study was conducted to generate empirical evidence on validity and reliability of the item of Competency Assessment Implementation (CAI) Instrument using Rasch Model for polythomous data aided by Winstep software version 3.68. The construct validity was examined by analyzing the point-measure correlation index (PTMEA), infit and outfit MNSQ values; meanwhile the reliability was examined by analyzing item reliability index. A survey technique was used as the major method with the CAI instrument on 156 teachers from vocational schools. The results have shown that the reliability of CAI Instrument items were between 0.80 and 0.98. PTMEA Correlation is in positive values, in which the item is able to distinguish between the ability of the respondent. Statistical data obtained show that out of 154 items, 12 items from the instrument suggested to be omitted. This study is hoped could bring a new direction to the process of data analysis in educational research.

***Keywords*—**Competency Assessment, Reliability, Validity, Item Analysis.

I. Introduction

MALAYSIA'S Ministry of Education has introduced a method of assessment Competency Based Education (CBE) by using a Modular System. Since the introduction of vocational education in Malaysia, the linear assessment is used to evaluate students' achievement. Starting 2006, CBE will be used to evaluate students who choose the vocational education in secondary school and the curriculum was designed by using modular system [1]. Table I shows differences between conventional assessment and CBE.

Modular approach in this sense requires a competency-based. Competency assessment comes up as an accepted evaluation method which suits present learning theories [4]. CBE in Malaysia using a portfolio as an alternative method in valuing students in progressing their work. As a general definition portfolio includes sample works that the student has done. According to few educators [5], [8], [11] portfolios give opportunity to see the improvement of students produced work regarding time.

Nurfirdawati Muhamad Hanafi is with the Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia (phone: 607-456 4183; fax:+607-4536585; e-mail: nurfirda@uthm.edu.my).

Azmanirah Ab Rahman, Marina Ibrahim Mukhtar, and Sarebah Warman are with the Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia (e-mail: marina@uthm.edu.my, azmanira@ uthm.edu.my, sarebah@uthm.edu.my).

Jamil Ahmad is with the Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia (e-mail: jamil3191@yahoo.co.uk).

TABLE I
Differences between Conventional Assessment and Competency-Based Assessment [12]

Conventional Assessment (1986-2005)	Competency-Based Assessment (2006 until now)
Comprehensive/overall assessment	Assessment based on each unit of the module of the subject - Using Portfolio
Norm-referenced test	Criteria on-referenced test
System Centred	Student Centred
Non-flexible	Flexible
Assessment based on learning (to assess the student at the end of the lesson)	Learning assessment (to assess student continuously to enhance the learning)
SPMV	SPM and Modular Certificate SPM - Graded
SPMV - Graded	Modular Certificate - Competent / Not Yet Competent

Due to the shift in the assessment paradigm in vocational education, the research to identify the implementation of CBE among the educators in vocational teachers is needed. Therefore, the aim of this study is to explore the empirical evidence in measuring validity and reliability in CAI instrument that have been developed. Key indicators of the quality of a measuring instrument are the reliability and validity of the measures. The process of developing and validating an instrument is in large part focused on reducing error in the measurement process. Reliability estimates evaluate the stability of measures, internal consistency of measurement instruments, and inter rater reliability of scoring student performances. Validity is the extent to which the interpretations of the results of a test are warranted, which depends on the particular use the test is intended to serve.

In recent years, Rasch models and item-response theory (IRT) or latent-trait models have provided an alternative framework for understanding measurement and alternative strategies for judging the quality of a measuring instrument. The Rasch measurement model is an application used in the study of the reliability and validity of CAI Instrument. Results from the item analysis will be used for deleting or modifying the overall test items. Using the Rasch measurement model, the quality of item is based on realiability and the separation of item and person or individuals [9], [10]. Rasch measurement model had performed item analysis and to determine the reliability and validity in the instrument in social science research and health care [6].

II. Methodology

This study is conducted by using a survey technique with a set of questionnaire. A set of CAI Instrument was developed with adaption from the literature review and documents from examination board. In the early stage, the questionnaire consists of 168 items which divided into three categories of CAI which are input, process and product. To ensure content validity, experts in survey research and in the subject matter were consulted and their input was used to revise and refine the items. After the item has been verified by experts, only 154 items were remained and used for the purpose of this study. Five-point Likert scales were using to evaluate the level of agreement of each item with the lowest number showed disagreement and the higher number for the agreement. A total of 156 teachers from selected vocational schools were participated as a sample. The samples are based on the random selection of teachers who were directly involved in the competency assessment. Data were conducted using SPSS for Windows Version 20.0. The data then was analyzed based on polythomous data Rasch model by using computer application, WINSTEP version 3.68 [3], [7].

III. Findings

The total number of respondents is 156. There were 84 respondents were male (53.8%) and 72 of them were female (46.2%) The total number of Malay vocational teachers was 147 (94.2%), Chinese 3 (1.9%) and others 6 (3.8%). Their ages are between 26 and 55 years old. This finding will be discussed in 3 parts which are reliability and separation index, items validity and variable map.

A. Reliability and Separation Index

Table II shows the value of item reliability and separation index obtained by analyzing the program through Winsteps. From the table it can be seen that all constructs of CAI Instrument showed number that greater than 0.8. These values indicated that every construct is strongly acceptable reliability because the values are all between 0.80 and 0.98 [3]. The value of separation index indicates the separation of item's difficulty level. All CAI Instruments constructs are accepted because the separation index is higher than 2, which consider as acceptable value [2].

B. Items Validity

There are few ways to measure item validity; however for the purpose of this study Point Measure Correlation (PTMEA Correlation), INFIT and OUTFIT mean square (MNSQ) were used. PTMEA correlation value is to test whether all items are moving in one direction with the construct. Tables III to XV display a positive index for PTMEA correlation. Tables III to VII also show Infit MNSQ and Outfit MNSQ value of CAI input items and respondents. Item C6005, C6001, C6002, C7002, C7001 has to be removed because has exceeded the range suggested by [3] the infit MNSQ and Outfit MNSQ of each item and the respondent should be in the range of 0.60 to 1.40 for likert scale. If an individual item is found not to be in the range, then it will be removed.

TABLE II
Items Reliability and Separation Index

Construct for CAI	Item ID	Item Reliability	Separation Index
INPUT	C3001-C7018	0.96	4.62
Teacher knowledge	C3001-C3011	0.85	2.35
Skills of teachers	C4001-C4011	0.94	4.11
Teacher attitudes	C5001-C5006	0.89	2.87
Teacher training	C6001-C6008	0.98	6.81
Infrastructure	C7001-C7018	0.91	3.25
PROCESS	D8001-D10013	0.94	3.88
Pre assessment	D8001-D8012	0.91	3.24
Assessment process	D9001-D9027	0.93	3.65
Scoring and keeping evidence	D10001-D10013	0.84	2.32
PRODUCT	E11001-E15011	0.94	3.94
Validity	E11001-E11011	0.95	4.36
Reliability	E12001-E12007	0.85	2.40
Feasibility	E13001-E13011	0.88	2.66
Outcome for students	E14001-E14009	0.86	2.49
Outcome for goal	E15001-E15011	0.80	2.02

TABLE III
CAI Input (Teacher Knowledge)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
4	0.92	0.93	0.54	C3004
2	0.94	0.94	0.52	C3002
3	1.06	1.07	0.43	C3003
6	0.90	0.89	0.57	C3006
5	0.92	0.92	0.54	C3005
8	0.82	0.82	0.62	C3008
9	0.86	0.86	0.58	C3009
11	0.95	0.90	0.55	C3011
1	0.95	0.93	0.52	C3001
10	0.93	0.92	0.55	C3010
7	0.91	0.90	0.55	C3007

TABLE IV
CAI Input (Skill of Teachers)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
12	0.89	0.89	0.56	C4001
20	0.90	0.89	0.56	C4009
21	0.88	0.87	0.58	C4010
17	0.90	0.88	0.56	C4006
19	0.93	0.91	0.55	C4008
15	0.98	1.00	0.49	C4004
16	0.90	0.88	0.57	C4005
13	0.82	0.82	0.61	C4002
14	0.91	0.93	0.55	C4003
18	0.91	0.89	0.56	C4007

TABLE V
CAI Input (Teacher Attitudes)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
24	1.22	1.32	0.34	C5003
22	1.20	1.26	0.33	C5001
23	1.17	1.28	0.35	C5002
25	1.15	1.26	0.30	C5004
27	1.14	1.26	0.35	C5006
26	1.25	1.30	0.24	C5005

TABLE VI
CAI INPUT (TEACHER TRAINING)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
32	1.77	1.93	0.32	C6005
28	1.41	1.41	0.39	C6001
31	1.26	1.26	0.45	C6004
29	1.4	1.55	0.28	C6002
30	1.23	1.22	0.43	C6003
33	1.26	1.32	0.37	C6006
34	1.14	1.19	0.43	C6007
35	0.96	1.01	0.49	C6008

TABLE VII
CAI INPUT (INFRASTRUCTURE)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
45	1.16	1.28	0.34	C7010
46	1.06	1.11	0.40	C7011
47	1.11	1.17	0.37	C7012
50	1.06	1.05	0.41	C7015
49	1.13	1.22	0.35	C7014
48	1.04	1.13	0.41	C7013
40	1.27	1.38	0.27	C7005
53	1.05	1.09	0.41	C7018
37	1.23	1.45	0.28	C7002
52	0.96	0.91	0.46	C7017
36	1.39	1.52	0.17	C7001
51	0.91	0.91	0.48	C7016
38	1.14	1.40	0.32	C7003
41	1.10	1.18	0.36	C7006
42	0.92	0.89	0.44	C7007
44	1.07	1.04	0.39	C7009
39	1.05	1.03	0.37	C7004

Tables VIII to X show that there are three items in the constructs that should be removed; D8001, D10004, D10002. All the three items had infit items MNSQ values and outfit items MNSQ value outside the range of 0.60 to 1.40.

TABLE VIII
CAI PROCESS (PRE ASSESSMENT)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
54	1.42	1.59	0.39	D8001
58	0.99	0.99	0.41	D8005
65	0.93	1.00	0.51	D8012
64	1.06	1.13	0.46	D8011
59	1.18	1.09	0.39	D8006
63	0.85	0.87	0.54	D8010
56	1.15	1.13	0.36	D8003
57	1.11	1.06	0.40	D8004
55	1.28	1.30	0.26	D8002
61	0.83	0.86	0.53	D8008
60	0.92	0.97	0.45	D8007
62	0.84	0.82	0.52	D8009

TABLE IX
CAI PROCESS (ASSESSMENT PROCESS)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
75	0.91	0.92	0.55	D9010
69	0.94	0.99	0.51	D9004
67	0.9	0.92	0.51	D9002
68	0.87	0.88	0.54	D9003
86	1.14	1.21	0.41	D9021
73	1.05	1.16	0.47	D9008
74	1.22	1.35	0.40	D9009
66	1.00	0.98	0.47	D9001
72	0.87	0.88	0.53	D9007
84	0.78	0.78	0.61	D9019
78	0.90	0.97	0.52	D9013
83	0.78	0.72	0.57	D9018
90	0.85	0.8	0.57	D9025
85	1.01	1.09	0.48	D9020
91	1.05	1.16	0.40	D9026
70	0.80	0.78	0.58	D9005
81	0.79	0.79	0.58	D9016
77	0.85	0.85	0.55	D9012
71	0.79	0.80	0.59	D9006
82	0.75	0.73	0.62	D9017
87	0.96	0.95	0.49	D9022
80	0.78	0.81	0.59	D9015
89	0.87	0.85	0.53	D9024
79	0.93	0.99	0.48	D9014
92	1.02	1.03	0.43	D9027
88	0.88	0.91	0.49	D9023
76	0.81	0.77	0.58	D9011

TABLE X
CAI PROCESS (SCORING AND KEEPING EVIDENCE)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
96	1.4	1.58	0.36	D10004
94	1.29	1.59	0.36	D40002
100	0.93	0.94	0.52	D10008
99	0.91	0.97	0.48	D10007
104	1.04	1.12	0.43	D10012
97	0.97	0.94	0.45	D10005
105	1.01	1.01	0.42	D10013
102	0.79	0.70	0.56	D10010
98	0.85	0.83	0.53	D10006
101	0.83	0.74	0.56	D10009
95	0.83	0.76	0.52	D10003
93	0.92	0.89	0.45	D10001
103	0.93	0.84	0.47	D10011

Tables XI to XV show that items E11005, E11007, E11006, E15009 has been removed in the competency assessment implementation (CAI) because the infit items MNSQ values and outfit items MNSQ value outside the range of 0.60 to 1.40.

TABLE XI
CAI PRODUCT (VALIDITY)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
110	1.54	1.83	0.29	E11005
112	1.43	1.62	0.35	E11007
111	1.25	1.51	0.4	E11006
109	1.14	1.18	0.42	E11004
114	1.00	1.06	0.46	E11009
116	0.95	0.91	0.49	E11011
113	0.98	1.10	0.42	E11008
115	1.03	1.05	0.41	E11010
107	0.90	0.86	0.45	E11002
106	0.98	0.96	0.39	E11001
108	0.93	0.95	0.44	E11003

TABLE XII
CAI PRODUCT (RELIABILITY)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
117	0.98	0.97	0.43	E12001
120	0.86	0.85	0.54	E12004
122	1.07	1.14	0.49	E12006
119	0.96	0.96	0.43	E12003
118	1.11	1.09	0.34	E12002
123	0.77	0.73	0.62	E12007
121	0.84	0.83	0.57	E12005
115	1.03	1.05	0.41	E11010
107	0.9	0.86	0.45	E11002
106	0.98	0.96	0.39	E11001
108	0.93	0.95	0.44	E11003

TABLE XIII
CAI PRODUCT (FEASIBILITY)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
125	0.95	0.99	0.51	E13002
133	1.09	1.24	0.39	E13010
127	0.91	0.98	0.51	E13004
124	0.83	0.82	0.59	E13001
130	0.91	1.00	0.49	E13007
132	1.08	1.19	0.41	E13009
134	0.98	1.08	0.44	E13011
131	0.97	1.07	0.47	E13008
128	0.84	0.85	0.55	E13005
129	0.75	0.72	0.63	E13006
126	0.83	0.82	0.54	E13003

TABLE XIV
CAI PRODUCT (OUTCOME FOR STUDENTS)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
140	0.92	0.90	0.50	E14006
142	0.94	0.95	0.49	E14008
138	0.95	0.88	0.48	E14004
139	0.90	0.82	0.51	E14005
143	0.90	0.89	0.53	E14009
141	0.90	0.92	0.52	E14007
135	0.96	0.90	0.49	E14001
137	0.99	0.95	0.45	E14003
136	0.99	0.96	0.45	E14002

TABLE XV
CAI PRODUCT (OUTCOME FOR GOAL)

ENTRY NUMBER	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORRELATION	ITEM
144	0.98	1.08	0.51	E15001
154	0.90	0.95	0.53	E15011
153	0.89	0.93	0.51	E15010
152	1.70	2.06	0.04	E15009
151	0.85	0.85	0.56	E15008
150	0.82	0.81	0.58	E15007
148	0.82	0.83	0.61	E15005
149	0.88	0.91	0.53	E15006
145	0.82	0.89	0.58	E15002
146	0.86	0.85	0.57	E15003
147	0.74	0.71	0.66	E15004

C. Variable Map

Fig. 1 shows the hierarchy of person ability and item difficulty in a straight line for CAI input. It is found that all item are scattered and heading towards various level sample ability. Person position with high ability (very much agreeable) stays at the top scale while person with low ability (less agreeable) stays at the lower part of the scale. The most difficult item C6005 (32) stays at the scale while simplest item C7007 (42) stays at lower part of the scale.

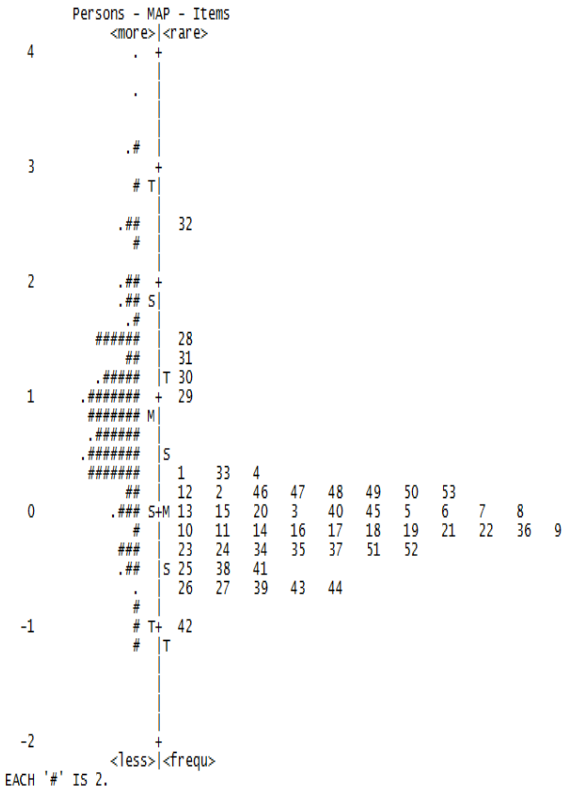

Fig. 1 Variable map for CAI input

Fig. 2 shows the hierarchy of person ability and item difficulty in a straight line for CAI process. It is found that all item are scattered and heading towards various level sample

ability. Person position with high ability (very much agreeable) stays at the top scale while person with low ability (less agreeable) stays at the lower part of the scale. The most difficult item D8001 (54) stays at the scale while simplest item D10011, D8002, D9001, D8008, D8009, D9023, D10001 and D10003 (103, 55, 60, 61, 62, 88, 93, 95) stays at lower part of the scale.

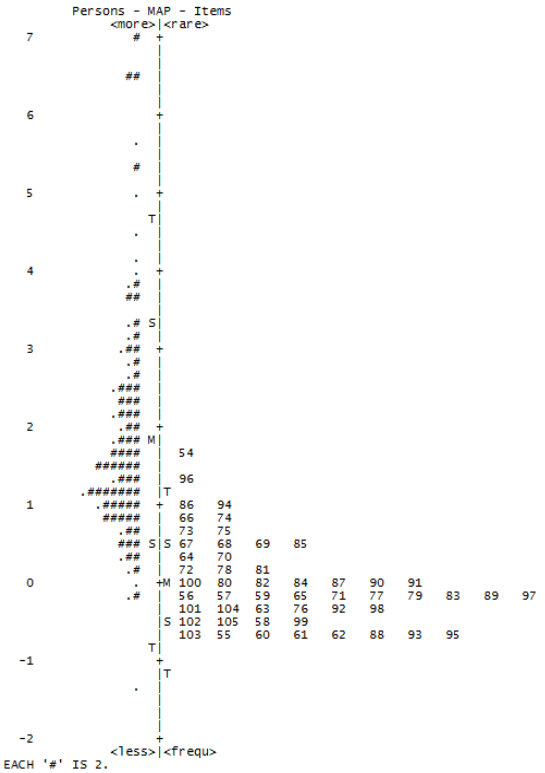

Fig. 2 Variable map for CAI process

Fig. 3 shows the hierarchy of person ability and item difficulty in a straight line for CAI product. It is found that all item are scattered and heading towards various level sample ability. Person position with high ability (very much agreeable) stays at the top scale while person with low ability (less agreeable) stays at the lower part of the scale. The most difficult item E11005 and E11007 (110, 112) stays at the scale while simplest item E11002 (107) stays at lower part of the scale.

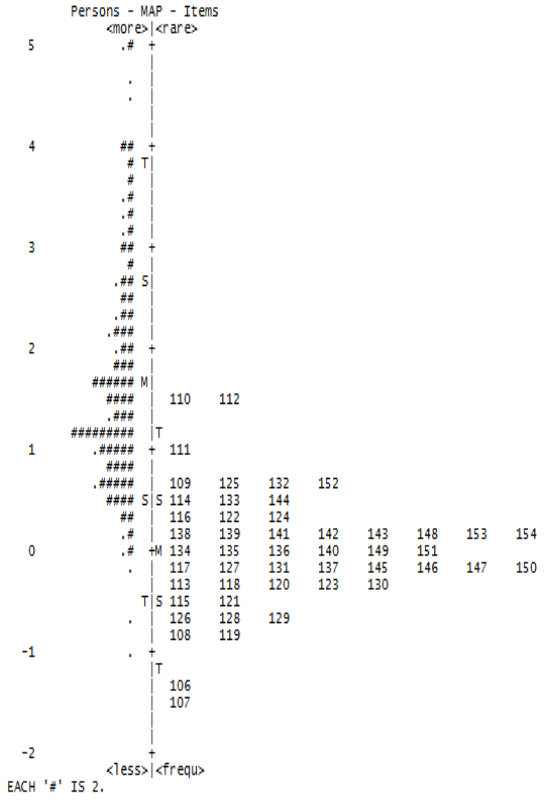

Fig. 3 Variable map for CAI product

IV. Conclusion

Item analysis is found as the best method for determining the quality of developed test items. Rash measurement model was used to analyze each item in order to measure the validity and reliability of competency assessment implementation (CAI) instrument. The Rasch model is effective as can be seen through the use of valid items to define the construct, a clear definition of the measurable constructs and consistent with theoretical expectations and the ability of the items that are consistent with the purposes of measurement. The findings have suggested all the possibilities designed to examine the suitability of the items in the CAI instrument. The item reliability is high which indicates the item is stable. Separation index for the level of difficulty exceeded the value of 2 which is strongly accepted. Twelve items with infit MNSQ and outfit MNSQ items which are out of acceptable range 0.6-1.4 for CAI input, CAI process and CAI product had been removed. This resulted to only 142 items which remained after the Rasch analysis. Further study is recommended to focus at differences based on GDIF items to remove the bias item based on gender and course.

Acknowledgment

This paper is under the sponsorship of University Tun Hussein Onn Malaysia (UTHM) and Fundamental Research Grant Scheme (FRGS).

REFERENCES

[1] Ahmad Tajudin Jab(2007), Technical and Vocational Curriculum Development. Kuala Lumpur: Department of Technical Education. Ministry of Education.
[2] Azrilah Abdul Aziz, Azlinah Mohamed, NoorHabibah Arshad, Sohaimi Zakaria, Azami Zaharin, Hamza Ahmad Ghulman & Mohd Saidfuddin Masodi. (2008). Application of Rasch Model in validating the construct of measurement instrument. *International Journal of Education and Information Technologies*, 2(2), 105-112.
[3] Bond T.G and Fox C.M, (2007), Applying The Rasch Model: Fundamental Measurement in Human Science. London: Lawrence Erlbaum Associates.
[4] Higgins, R., Hartley, P., and Skelton, A. (2002), The Conscientious Consumer: Re Considering The Role of Assessment Feedback on Student. Learning, Studies in Higher Education, 27(1), 382-400.
[5] Katz, A. M., and Gangnon, B., A. (2000), Portfolio Assessment Integrating Goals and Objectives With Learner Outcomes. Assessment 12(1).
[6] Kimberlin, C. L. and Winterstein, A. G. (2008), Validity and Reliability of Measurement Instruments Used in Research. *Am J Health-SystPharm*. 65(2276-2284).
[7] Linacre J.M. (2003), Winsteps Computer Program Version 3.48. Chicago: www.winsteps.com (10 Sept. 2009).
[8] Pitts, J. et al. (2002). Enhancing Reliability in Portfolio Assessment: Discussion Between Assessors Medical Teacher 24 (2), 197-201.
[9] Richardson R. C. (2008), The Development and validation of a personality instrument to increase collaboration, *Educational Research and Review*, 3(4), 121-127.
[10] Schumacker R.E. (2005), Item Response Theory. Applied Measurement Associates.
[11] Tigelaar et al. (2005), Quality Issues in Judging Portfolios: Implications for Organizing Teaching Portfolio Assessment Procedures. *Studies in Higher Education*. 30(5), Number: 595-610.
[12] Technical & Vocational Curriculum Development Department, 2007 & Examination Board, Ministry of Education Malaysia, 2008