Pak J Commer Soc Sci
Pakistan Journal of Commerce and Social Sciences
2015, Vol. 9 (1), 269-281

Development and Rasch Analysis of an Achievement Test at Master Level (Philosophy of Education)

Zunaira Fatima
Department of Education, University of Sargodha, Sargodha, Pakistan
Email: neerasalman@gmail.com

Shamim Haider Tirmizi
Department of Education, Bahuddin Zakariya University Multan, Pakistan
Email: farooqmashhadi@bzu.edu.pk

Muhammad Ijaz Latif
Department of International Relations, The Islamia University of Bahawalpur, Pakistan
Email: ijaz.latif@iub.edu.pk

Aqueel Raza Gardezi
Department of Education, University of Sargodha, Sargodha, Pakistan.
Email: razaaqueel@hotmail.com

Abstract

The specific purpose of the research was to construct an achievement test in the area of philosophy of education for master level students in universities of Punjab (Bahauddin Zakariya University Multan, The Islamia University of Bahawalpur and University of Sargodha, Sargodha. Test comprises 60 multiple-choice items, selected from the item bank constructed by researcher. This test was administered to 231 male and female students of M.A. Education and M.Ed. selected randomly. Data were analysed through Rasch model. As the result of Rasch calibration, three items were tossed out of the test. Figure latent continuum showing position of items and persons was made. This study suggests that to cover the whole syllabus items should be increased. An effort should be made to take a greater number of samples for the study, so that item analysis through Rasch Model can show its probability.

Keywords: rasch model, philosophy of education, achievement test, test standardisation.

1. Introduction and Literature Review

The study aims at the development and analysis of an achievement test in the subject area of “Philosophy of Education” for master level students from Pakistan. Shah (1995) says that modern teacher education is interested in the overall development of the learner. Many teachers are observed arguing about the nature of learning and evaluation. Their efforts towards the assessment of learner’s achievement made their way towards achievement testing. Achievement tests play an important role in the evaluation process

of the learner (Chatterji, 2003), and these are being used to measure the current status of the individuals in a particular area of knowledge and skill. Aiken (2000) is of the view that achievement tests have a responsive sample of the source content (possess content validity) and are designed to measure the present status of the individuals. Its construction demands clarity of objectives and a complete and careful representation of the content (Gronlund, 1998). Then the test is administered to a sampled group of students. It is suggested that a good test is administered in such a way that the students may perform their best. Gronlund (1998) is of the view that after administration and scoring the test it is desirable to evaluate the effectiveness of the test items. This is done by analysing students' responses to each item. When formalized, the whole procedure is called "Item Analysis" traditional. That provides information concerning each of the following points (Linn & Miller, 2008).

i. The difficulty of the items "p"
ii. The discriminating power of the items "D"
iii. Effectiveness of the distracters.

Item analysis traditional is most commonly used for analysing the items. Still it possesses some weaknesses that it doesn't provide information concerning about the strength and weaknesses about the persons. For this, a mathematical approach named "Rasch calibration/model" is used.

The Rasch model of measurement is based on two expectations regarding the outcomes of a person attempting an item in a test (Kline, 2004). Firstly, that a person with higher attainment will have a greater probability of success on any item from that topic than a person with lower attainment; and secondly, that any person should always be more likely to answer correctly an easier item on that topic than a hard one. Rasch calibration involves examining the student's performance on each item and calculates the item difficulty and the person attainment on the same scale. Due to this practical nature and usefulness of Rasch calibration the researcher decided to analyse the test by using Rasch model of item analysis. In past many researchers were intended to develop and analyse the test items. But this research study is prominent in two aspects. More over the results are drawn in the form of item difficulty and person attainment and are presented on one line called figure "Latent continuum." ICC (Item character curve) and PCC (person character curve) are also drawn to make the study more significant (Rudner, 2001).

1.1 Objectives of the Study

The purpose of the study was to construct and analyse an achievement test in the area of Philosophy of Education at master level.

1.2 Procedure of the Study

A test comprising 60 multiple-choice items was constructed. For the administration of the test ten per cent randomised subjects were selected out of the population (231) i.e. the Master level students of Education at three universities of Punjab (Bahauddin Zakariya University Multan, The Islamia University of Bahawalpur and University of Sargodha, Sargodha. Item analysis was done through Rasch model. By applying Rasch model, items were arranged in order of their difficulty. The analysis was dealt with up to the extent of item person interaction and probability curves i.e. ICC and PCC (item character curve and person character curve)

2. Test Calibration and the Rasch Model

Traditional item analysis is dependent of students taking a test. Item difficulty and item discrimination is based on the sample. The Rasch approach of item analysis is independent of the sample and the item. The item calibration is sample free and the person measurement is item-free. This quality gives the Rasch Model a specific objectivity as shown in the figure 1.

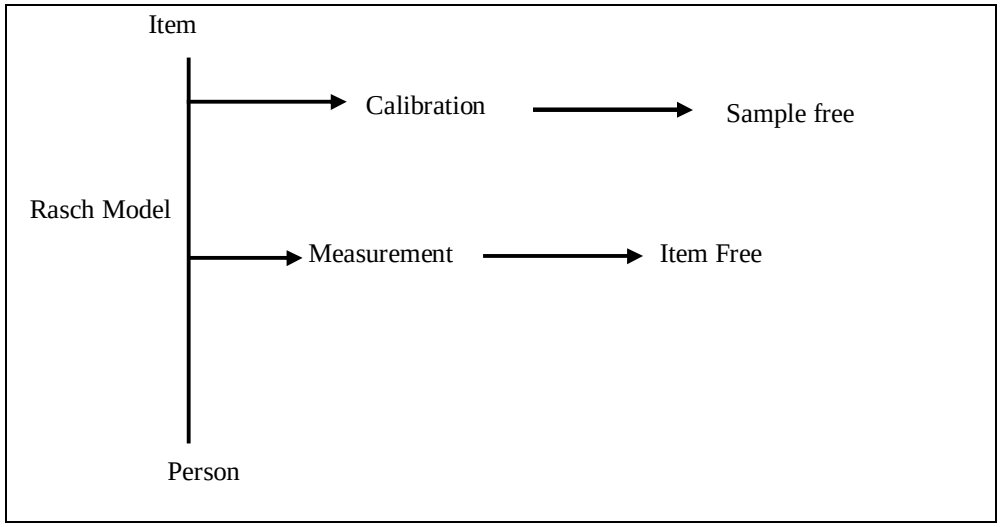

Figure 1: Test Calibration and the Rasch Model

2.1 The Model in Action

Item calibration and person measurement were done with the help of the procedure, named PROX that could be performed by hand. Sample item data and the results were derived using the procedures that have just been presented in the Table (1) and Table (2).

2.2 Obtaining Initial Item Calibration

Table 1: Pox Item Calibration

Item Name	Item Score	Proportion		Legit Xi=ln(1-pi)/pi)	Initial Item Calibration din=xi-x'	Sample Spread =Y	Item Final Calibration (di)
		Correct (pi)	In Correct (1-pi)				
48	194	0.84	0.16	-1.66	-2.05	1.21	-2.47
7	162	0.70	0.30	-0.85	-1.25	1.21	-1.50
43	155	0.67	0.33	-0.71	-1.11	1.21	-1.33
47	145	0.63	0.37	-0.52	-0.92	1.21	-1.10
53	142	0.61	0.39	-0.47	-0.86	1.21	-1.04
58	141	0.61	0.39	-0.45	-0.84	1.21	-1.02
23	139	0.60	0.40	-0.41	-0.81	1.21	-0.97
21	134	0.58	0.42	-0.32	-0.72	1.21	-0.86
15	127	0.55	0.45	-0.20	-0.59	1.21	-0.72
56	123	0.53	0.47	-0.13	-0.52	1.21	-0.63
59	123	0.53	0.47	-0.13	-0.52	1.21	-0.63
2	119	0.52	0.48	-0.06	-0.45	1.21	-0.55
51	118	0.51	0.49	-0.04	-0.44	1.21	-0.53
1	116	0.50	0.50	-0.01	-0.40	1.21	-0.48
31	116	0.50	0.50	-0.01	-0.40	1.21	-0.48
42	114	0.49	0.51	0.03	-0.37	1.21	-0.44
55	114	0.49	0.51	0.03	-0.37	1.21	-0.44
57	113	0.49	0.51	0.04	-0.35	1.21	-0.42
60	112	0.48	0.52	0.06	-0.33	1.21	-0.40
3	110	0.48	0.52	0.10	-0.30	1.21	-0.36
46	107	0.46	0.54	0.15	-0.25	1.21	-0.30
40	107	0.46	0.54	0.15	-0.25	1.21	-0.30
44	104	0.45	0.55	0.20	-0.19	1.21	-0.23
50	104	0.45	0.55	0.20	-0.19	1.21	-0.23
54	99	0.43	0.57	0.29	-0.11	1.21	-0.13
28	99	0.43	0.57	0.29	-0.11	1.21	-0.13
25	99	0.43	0.57	0.29	-0.11	1.21	-0.13
16	95	0.41	0.59	0.36	-0.03	1.21	-0.04
34	94	0.41	0.59	0.38	-0.02	1.21	-0.02
38	94	0.41	0.59	0.38	-0.02	1.21	-0.02
45	92	0.40	0.60	0.41	0.02	1.21	0.02
19	91	0.39	0.61	0.43	0.04	1.21	0.04
49	89	0.39	0.61	0.47	0.07	1.21	0.09
52	86	0.37	0.63	0.52	0.13	1.21	0.16
32	86	0.37	0.63	0.52	0.13	1.21	0.16
37	84	0.36	0.64	0.56	0.17	1.21	0.20
27	83	0.36	0.64	0.58	0.18	1.21	0.22
18	83	0.36	0.64	0.58	0.18	1.21	0.22

9	82	0.35	0.65	0.60	0.20	1.21	0.25
10	80	0.35	0.65	0.64	0.24	1.21	0.29
22	80	0.35	0.65	0.64	0.24	1.21	0.29
20	78	0.34	0.66	0.67	0.28	1.21	0.34
35	75	0.32	0.68	0.73	0.34	1.21	0.41
36	75	0.32	0.68	0.73	0.34	1.21	0.41
39	74	0.32	0.68	0.75	0.36	1.21	0.43
30	72	0.31	0.69	0.79	0.40	1.21	0.48
24	71	0.31	0.69	0.81	0.42	1.21	0.50
33	70	0.30	0.70	0.83	0.44	1.21	0.53
26	70	0.30	0.70	0.83	0.44	1.21	0.53
12	67	0.29	0.71	0.90	0.50	1.21	0.60
8	67	0.29	0.71	0.90	0.50	1.21	0.60
13	65	0.28	0.72	0.94	0.54	1.21	0.66
17	64	0.28	0.72	0.96	0.57	1.21	0.68
6	63	0.27	0.73	0.98	0.59	1.21	0.71
41	58	0.25	0.75	1.09	0.70	1.21	0.84
11	58	0.25	0.75	1.09	0.70	1.21	0.84
29	41	0.18	0.82	1.53	1.14	1.21	1.37
14	40	0.17	0.83	1.56	1.17	1.21	1.41
5	35	0.15	0.85	1.72	1.33	1.21	1.60
4	12	0.05	0.95	2.90	2.51	1.21	3.03
				0.393	0.466		

No of items = 60 *No of students = 231 (N)*

From the edited data matrix in Table 1, there was built a distribution of the 60 different item scores (ordering from high to low 194 to 12). Their logits incorrect and computed the mean and variance of the distribution of these item logits over the test of 60 items as shown in table 1.

Explanation of Table 1 (columns 1 to 6)

Column 1 gives the name (number) of each item (i), since there are 60 items (L), the item score index (i) goes from 1 to 60.

Column 2 gives the item scores, which characterizes each item (si), i.e. the number of Persons who got a particular item correct. (Item 48 was correctly done by 194 students out of 231)

Column 3 converts the item scores into proportions correct among the sample of

N=231 persons Pi=Si/N (For item 48 Pi = 194/231 = 0.84).

Column 4 is the conversion of proportion correct (Pi) into the proportion incorrect (1-Pi) (For item 4 proportion incorrect = 1-0.84 = 0.16)

Column 5 (i) is the conversion of this proportion into logits incorrect. Each item score logits is the natural log of its proportion incorrect divided by its proportion correct.

Xi=Ln {(1-Pi) / Pi}, (For item 4 Xi = Ln {.16/.83}), Xi= -1.66

This conversion is facilitated by the use of scientific calculator.

(ii) At the bottom of column 5, the mean of the logit incorrect is written:

$$X = \frac{\sum Xi}{L} = \frac{\text{Sum of logit incorrect}}{\text{No. of items}}$$

$$X = \frac{(-1.65) + (-0.85) + (0.71) + \ldots + (2.90)}{60}$$

Column 6 (i) gives the values of column 5 centred by subtracting their mean. These are the initial item calibrations ready to be corrected for the effected of sample spread.

di°= XI-X°, (For item 4 di° = (-1.65- 0.393) = -2.05

(ii) At the bottom of column 6, the variance u) of initial item calibration is written.

(iii) $U = \frac{\left[\sum Xi - X^{0}\right]}{L-1}$, $U = \frac{-(2.05) + (-1.24) + \ldots + (2.51)^{2}}{59} = .046$

2.3 Obtaining Initial Person Measurement

Table 2: Prox Person Measurement

		Proportions		Incorrect 1-pr	Legit Correct yr=ln(pr/1-pr)	Initial Measure brn=yr	Test Width Expansion Factor (X)	Final Measure (r)
person freq (f)	Block	possible score ®	Correct pr=r/l					
0	a1	1	0.02	0.98	-4.08	-4.08	1.11	-4.53
0	a2	2	0.03	0.97	-3.37	-3.37	1.11	-3.74
0	a3	3	0.05	0.95	-2.94	-2.94	1.11	-3.27
0	a4	4	0.07	0.93	-2.64	-2.64	1.11	-2.93
0	a5	5	0.08	0.92	-2.40	-2.40	1.11	-2.66
2	a6	6	0.10	0.90	-2.20	-2.20	1.11	-2.44
0	a7	7	0.12	0.88	-2.02	-2.02	1.11	-2.25
0	a8	8	0.13	0.87	-1.87	-1.87	1.11	-2.08
2	a9	9	0.15	0.85	-1.73	-1.73	1.11	-1.93
3	a10	10	0.17	0.83	-1.61	-1.61	1.11	-1.79
12	a11	11	0.18	0.82	-1.49	-1.49	1.11	-1.66
7	a12	12	0.20	0.80	-1.39	-1.39	1.11	-1.54
19	a13	13	0.22	0.78	-1.29	-1.29	1.11	-1.43
25	a14	14	0.23	0.77	-1.19	-1.19	1.11	-1.32
22	a15	15	0.25	0.75	-1.10	-1.10	1.11	-1.22
19	a16	16	0.27	0.73	-1.01	-1.01	1.11	-1.12
24	a17	17	0.28	0.72	-0.93	-0.93	1.11	-1.03
26	a18	18	0.30	0.70	-0.85	-0.85	1.11	-0.94
19	a19	19	0.32	0.68	-0.77	-0.77	1.11	-0.85
15	a20	20	0.33	0.67	-0.69	-0.69	1.11	-0.77
10	a21	21	0.35	0.65	-0.62	-0.62	1.11	-0.69

9	a22	22	0.37	0.63	-0.55	-0.55	1.11	-0.61
8	a23	23	0.38	0.62	-0.48	-0.48	1.11	-0.53
5	a24	24	0.40	0.60	-0.41	-0.41	1.11	-0.45
3	a25	25	0.42	0.58	-0.34	-0.34	1.11	-0.37
1	a26	26	0.43	0.57	-0.27	-0.27	1.11	-0.30
0	a27	27	0.45	0.55	-0.20	-0.20	1.11	-0.22
0	a28	28	0.47	0.53	-0.13	-0.13	1.11	-0.15
0	a29	29	0.48	0.52	-0.07	-0.07	1.11	-0.07
0	a30	30	0.50	0.50	0.00	0.00	1.11	0.00
0	a31	31	0.52	0.48	0.07	0.07	1.11	0.07
0	a32	32	0.53	0.47	0.13	0.13	1.11	0.15
0	a33	33	0.55	0.45	0.20	0.20	1.11	0.22
0	a34	34	0.57	0.43	0.27	0.27	1.11	0.30
0	a35	35	0.58	0.42	0.34	0.34	1.11	0.37
0	a36	36	0.60	0.40	0.41	0.41	1.11	0.45
0	a37	37	0.62	0.38	0.48	0.48	1.11	0.53
0	a38	38	0.63	0.37	0.55	0.55	1.11	0.61
0	a39	39	0.65	0.35	0.62	0.62	1.11	0.69
0	a40	40	0.67	0.33	0.69	0.69	1.11	0.77
0	a41	41	0.68	0.32	0.77	0.77	1.11	0.85
0	a42	42	0.70	0.30	0.85	0.85	1.11	0.94
0	a43	43	0.72	0.28	0.93	0.93	1.11	1.03
0	a44	44	0.73	0.27	1.01	1.01	1.11	1.12
0	a45	45	0.75	0.25	1.10	1.10	1.11	1.22
0	a46	46	0.77	0.23	1.19	1.19	1.11	1.32
0	a47	47	0.78	0.22	1.29	1.29	1.11	1.43
0	a48	48	0.80	0.20	1.39	1.39	1.11	1.54
0	a49	49	0.82	0.18	1.49	1.49	1.11	1.66
0	a50	50	0.83	0.17	1.61	1.61	1.11	1.79
0	a51	51	0.85	0.15	1.73	1.73	1.11	1.93
0	a52	52	0.87	0.13	1.87	1.87	1.11	2.08
0	a53	53	0.88	0.12	2.02	2.02	1.11	2.25
0	a54	54	0.90	0.10	2.20	2.20	1.11	2.44
0	a55	55	0.92	0.08	2.40	2.40	1.11	2.66
0	a56	56	0.93	0.07	2.64	2.64	1.11	2.93
0	a57	57	0.95	0.05	2.94	2.94	1.11	3.27
0	a58	58	0.97	0.03	3.37	3.37	1.11	3.74
0	a59	59	0.98	0.02	4.08	4.08	1.11	4.53
						1.05		

No of items = 60 *No of students = 231*

In Table 2, there was taken identical steps with a distribution of person scores in order to obtain the distribution of person score logits and hence initial values for the abilities that go with each possible score on the test.

Explanation of Table 2 (column 1-7)

Column 1 gives the frequency of persons observed at each score. The total number of persons N= 231 equals the sum of these frequencies.
Column 2 shows the persons (against each possible score) in different blocks. As the sample for the test was very large (231), instead of writing all persons' numbers, block (A1 to A59) were allocated to each group. Score on the test was ranging from 06 to 27, so blocks were made that are shown in table 1.4 (In block a1, a2, a3 and so on there was no person).

Column 3 gives each possible score from 1 to 59 (r)

(As there was no score (0) and no person had perfect score (60) which was excluded from the calibration). So r=1 to L-1, r goes from 1 to 59)
Column 4 is the proportion of each score on a test of 60 items (Pr) Pr=r/L (In case of any person from block a1 Pr=1/60=0.02)

Column 5 is the conversion of proportion correct (Pr) into the proportion (1-Pr) (For person in blockA1N, 1-Pr=1-0.02=0.98)

Column 6 is the logit correct for that proportion using scientific calculator Yr = In {Pr/ (1-Pr)} (For person in block A1, Yr=1n {0.016/0.98}= -4.08)

Column 7

(i) Repeats the value of column 6 because, as for as this test is concerned, the score logits are already taken by the entering of the item logits. These are the initial person measure (br°) prior to correction of test width.

(ii) The variance (V) for the distribution of score logits over persons is given at the bottom of column 7.

(iii)

$$V= \frac{\sum (br^0)^2 \times f}{N-1}$$

For this distribution of score,

$$V = \frac{(-4.77)\times(-3.38)\times\ldots\times(4.07)}{230} \quad \text{so} \quad V= 1.05$$

2.4 Calculating the Expansion Factor

The co-efficient X and Y are expansion factors, which respond in the case of X to the difficulty dispersion of items and in the case of Y to the ability dispersion of person. There was computed expansion factors for the initial estimates of item calibrations and person measures in order to correct the item calibrations for sample spread and the person measures for test width, From Table 1 (column 6) and Table 2 (column 7) there was calculated:

U = 0.46and V= 1.05

(i) The person ability expansion factor due to the test width is:

$$X = \left[\frac{1+\frac{u}{2.89}}{1-\frac{uv}{8.35}}\right]^{\frac{1}{2}} = \left[\frac{1+\frac{0.46}{2.89}}{1-\frac{(0.46)(1.05)}{8.35}}\right]^{\frac{1}{2}}$$

The value of X is written in column 7 (Table 1)

ii. The item difficulty expansion factor due to sample spread is:

$$Y = \left[\frac{1+\frac{v}{2.89}}{1-\frac{uv}{8.35}}\right]^{\frac{1}{2}} = \left[\frac{1+\frac{1.05}{2.89}}{1-\frac{(0.46)(1.05)}{8.35}}\right]^{\frac{1}{2}}$$

The value of Y is written in column 6 (Table 2)

2.5 Correcting Item Calibration for the Effect of Sample Spread

Column 8 (Table1) gives the corrected item calibration obtained by multiplying each initial value in column 6 by the expansion factor of 1.20 (column 7). It gives the final item calibration for each item (di). $Di=Y \times di°$

(For item 48, the final item calibration= -2.05x 1.20= -2.47) and for item 39, this value will be 0.35x1.20 = 0.43).

Column 9 (Table 1.3) gives the corrected person measures obtained by multiplying each initial value in column 7 by expansion factor of 1.11 (column 8). It gives the final person measure. $Br = X \times br°$

(For any person from block A1, the value of final ability measure will be (-4.08x1.11) = -4.53.

3. Latent Continuum

The latent continuum shows the position of items and persons w. r. t. sample on a vertical line.

3.1 Persons

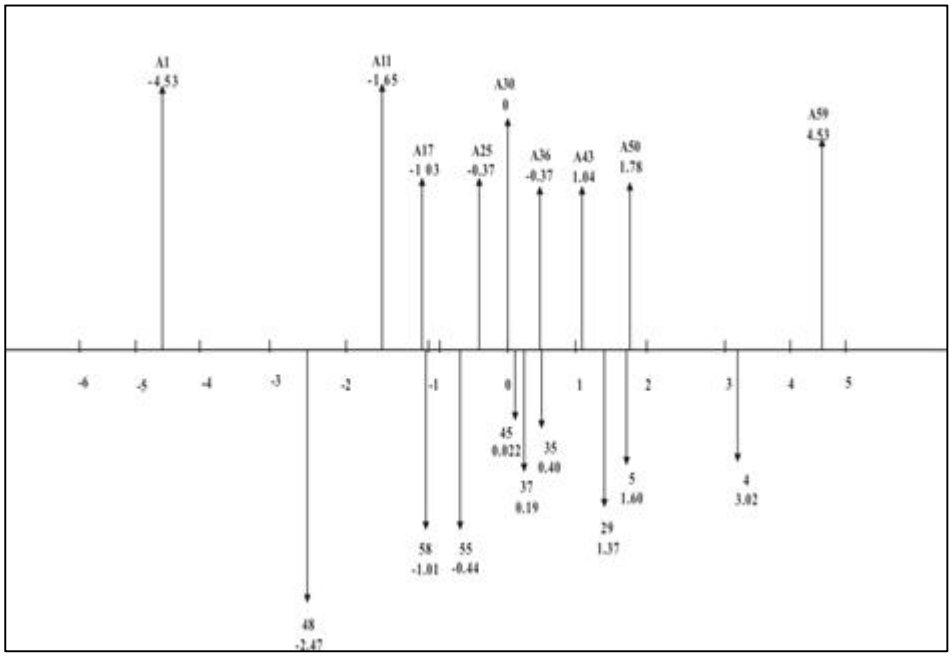

Figure 2: Latent Continuum Showing Positions of Items and Persons for Data in Table

3.2 Items

Figure 2 shows that item 4 should be discarded on account of being very hard and item 48 should also be removed, as it is very easy. Item 59 and 29 need modification. Other items are fairly good to be retained. We can also infer that persons in block a25 should have a good chance of getting items 48, 55 and 59 correct. This approach leads us to the place and importance of probability in the Rasch model, which is more precise and mathematical phase of Rasch calibration.

3.3 Item Character Curve

Item Characteristic Curve (ICC) and Person Characteristic Curve (PCC) give clear position of item difficulties and position of persons among these difficulties. Using the formula;

$$P = \exp(Bv\text{-}di)/ [1+ \exp(Bv\text{-}di)]$$

probabilities were found and curves were drawn as shown in graphs (Bond & Fox, 2001; Magno and Ouano, 2009)

Table 3: Item Character

Item No.	Items Difficulty (d_i)	Persons' Ability (B_V)		
		Series 1	Series 2	Series 3
		A14	A30	A46
		-1.32	0.00	1.32
43	-1.33	0.50	0.21	0.07
21	-0.86	0.61	0.30	0.10
15	-0.72	0.65	0.33	0.12
51	-0.53	0.69	0.37	0.14
3	-0.36	0.72	0.41	0.16
41	0.84	0.90	0.70	0.38
4	3.03	0.99	0.95	0.85

For series 1, there is 50% probability for persons of block A-14 (-1.32 ability level) to solve an item with difficulty level -1.33 (item no. 43). For series 2 there is 21% probability for persons of block A-30 (0.00 ability level) to solve an item with difficulty level -1.33 (item no. 43) and for series 3, there is 7% probability for persons of block A-46 (1.32 ability level) to solve an item with difficulty level -1.33 (item no. 43) and so on as shown in graph.

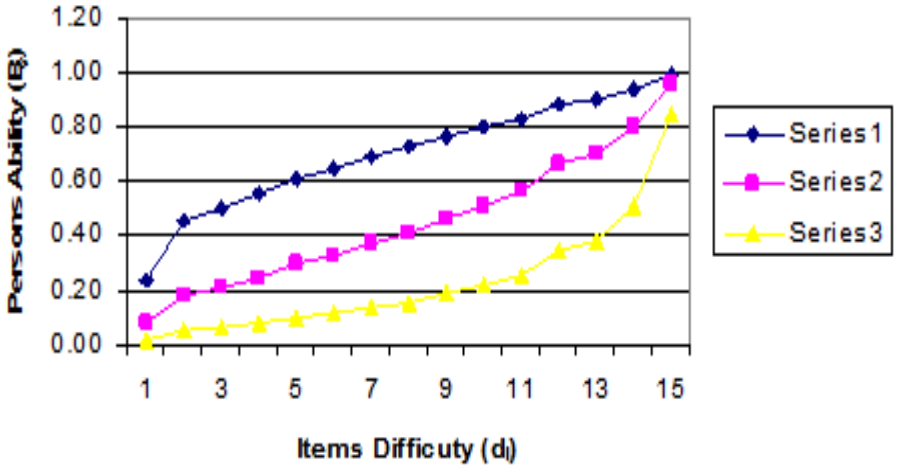

Figure 3: Item Character Curve

Table 4: Person Character Curve

Person Ability Level (Bv)	Item Difficulty Level (di)												
	-1.50	-0.86	-0.53	-0.23	0.02	0.16	0.29	0.43	0.60	0.71	0.84	1.41	3.03
-0.53	0.73	0.58	0.50	0.43	0.37	0.33	0.31	0.28	0.24	0.22	0.20	0.13	0.03
0.00	0.82	0.70	0.63	0.56	0.50	0.46	0.43	0.39	0.35	0.33	0.30	0.20	0.05
0.94	0.92	0.86	0.81	0.76	0.72	0.69	0.66	0.62	0.58	0.56	0.52	0.38	0.11

For series 1, persons with ability value -0.53 showed 73% probability to solve an item with difficulty level -1.50. For the persons of the same ability there was 31% to solve an item with difficulty level 0.29. For the persons of the same ability there was 31% to solve an item with difficulty level 3.03, similarly for series 2 and 3 as shown in graph. As the curve is sloppy it differentiates well between those with low and high ability persons (Quadir, Gillani, and Hameed, 2012).

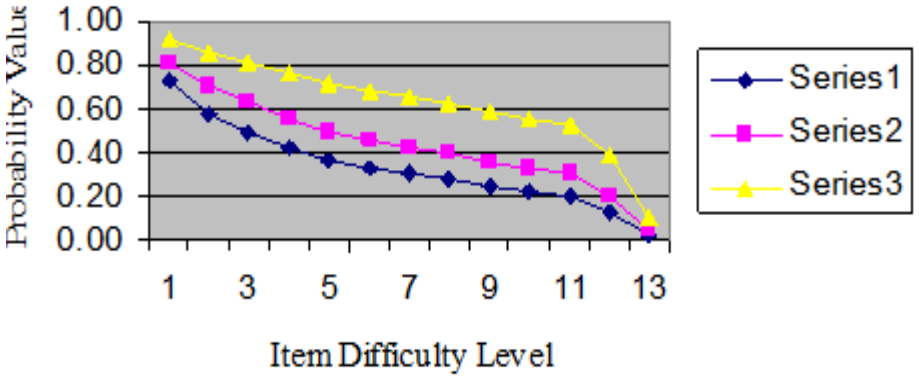

Figure 4: Person Characteristic Curve

4. Conclusions

As the major purpose of the study was to prepare and analyse an achievement test in philosophy of education. For this purpose the researcher constructed a multiple choice item test comprising 60 items and administered it in the three universities of Punjab. A randomly selected sample of 231 students was selected and the test was administered. Student's response towards the test was not really positive as the test appeared a little bit difficult for them. The range of the scores was really low. But the positive aspect of the test was it covered almost all important philosophies and philosophers of education.

After organizing and editing the data matrix, final item calibration and person measurement estimates were obtained. Position of each item and person was shown on a horizontal line. Figure of latent continuum was made to show the position of items and persons on a horizontal line. The continuum discussed the probability of attempting an item by a person and vies versa: which made the Rasch model more clear and meaningful. The results showed that, Item 4 should be discarded on account of being very hard and item 48 should also be removed, as it is very easy. Item 59 and 29 need modification. Similarly the persons from A1 have good probability to solve the items correctly, while A59 would have low probability to solve the items correctly.

Item characteristic curve and person characteristic curve were also drawn to show the probability of items and persons more clear. For the standardization of the test sample size should be increased. More universities of other provinces should be included in the sample to make it more representative. Efforts should be made to collect information about the personal bio-data of the students e.g. date of birth, parent's education, parent's profession, favorite subject, favorite teacher and the subject of favorite teacher. Teachers should make use of this data to check the effect of these factors in their achievement in their examination. (Hashami, et al, 2012).

REFERENCES

Aiken, L. R. (2000). *Psychological Testing and Assessment*, 10th ed., Boston: Allyn and Bacon.

Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in Human Sciences*, Mahwah NJ: Erlbaum

Chatterji, M. (2003). *Designing and Using Tools for Educational Assessment*, Boston: Allyn and Bacon.

Encarta Encyclopaedia, (2005). © 1993-2004 Microsoft Corporation.

Gronlund, N. E. (1998). *Assessment of Student Achievement*, 7th ed., Boston: Allyn and Bacon.

Kline, P. (2004). *Psychological Testing*, New York: Methuen & Co.

Linn, R. L. & Miller, M. D. (2008). *Measurement and Assessment in Teaching*, 9th ed., Delhi: Dorling Kindersley Pvt. Ltd.

Linn, R. L. and Gronlund, N. E. (2000). *Measurement and Evaluating In Teaching*, (8th ed.), New Jersey: Prentice Hall, Inc.

Magno, C. and Ouano, J. (2009). *Designing the Written Assessment of Student Learning*, Manila: De La Salle University.

Rudner, L. M. (2001). "Item Response Theory" [Online] Available: http://edres.org/irt/baker

Shah, R. A. (1995). *Education and Teacher Education in Pakistan*, Karachi: Pakistan Study Centre, University of Sindh.

Quadir, M., Gillani, G. and Hameed, A. (2012). Rasch Calibration of General Science Test at Grade- VIII in Pakistan. *Journal of Education and Vocational Research, 3*(3), 98-106.

Hashami, A., Zeeshan, A., Saleem, M. and Ali, R. (2012). Development and Validation of an Aptitude Test for Secondary School Mathematics Students. *Bulletin of Education and Research, 34* (1), 65-76.