Diagnosis of Perceived Attitude, Importance, and Knowledge in Statistics Based on Rasch Probabilistic Model

Zamalia Mahmud

***Abstract*—** Postgraduate students who are involved in the analyiss of survey research data are expected to have a sound knowledge in statistics. However, statistics learners should firstly inculcate interest, feel the importance of statistics and instill positive attitude toward learning statistics. Based on Rasch Unidimensional Measurement Model, this study shall investigate the attitude, importance and knowledge in statistics among postgraduate students. The investigation involved a small classroom experiment measuring students' perceived attitude prior to and after attending the statistics remedial classes and then their perceived knowledge on various statistical tools and concepts. The validity of the response patterns were measured and whether their responses can be taken as measuring the latent trait or otherwise. This paper also describes how Rasch probabilistic model is used to measure the latent trait of students' attitude in response to the agent or items measuring the trait. This involves a calibration between person ability to respond towards the five-point Likert scale items and whether the items function in unison to form a single underlying pattern. Logit scores as derived from the Rasch measurement model, Person-Item Distribution map and other related Rasch statistics are executed to endorse the trait. The study shows that student demonstrate a higher cognitive positive attitude toward learning statistics after attending the remedial classes with most post-attitude items superseding the pre-attitude items during the calibration. However, there are traces of misfit responses as identified by the fit indices due to unusual or/and inappropriate responses which are highlighted in this paper.

***Keywords*—**Attitude towards learning statistics, Pre-post attitude Rasch probabilistic model, PIDM, Person-Item misfit, Perceived Importance and Knowledge

I. Introduction

OVER the past few years, there has been a concerted effort to improve the teaching and learning of statistics. The use of technology in the statistics classroom as well as new and innovative teaching strategies continue to offer teachers (and students) with many teaching (and learning) alternatives. Some of these innovative teaching and learning strategies have also been used to improve student attitudes.

Statistics is often considered as a difficult subject to learn by many students. It is also quite often associated with students' having negative feelings toward statistics which may result in poor performance in statistics. Knowledge about students' learning attitudes in statistics should help to understand why certain students show initiative in learning for statistics and attribute success on their learning behaviour. Negative student attitudes toward statistics may create a major obstacle for effective learning [1] – [2].

Zamalia Mahmud is an Associate Professor at the Center for Statistical Studies, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA.(phone: +603-55435367; fax: +603-55435301; e-mail: zamalia@tmsk.uitm.edu.my).

However, some studies relating to teaching of statistics varies from being no changes in student attitudes toward statistics [3] to some changes in certain aspects of statistics [4].

Studies on teaching and learning of statistics at school and tertiary level have found that student learning of and attitudes toward statistics are positively related to previous mathematics and statistics experience [5], gender [2], types of student, and teacher [6], learning practices and ability of students [7]. Innovative teaching and learning strategies have also been used to improve student attitudes [3].

In assessing students' attitude toward statistics, many efforts have been made through development of attitude scale to investigate factors that could affect students' attitude towards learning statistics. Among them include Attitude Toward Statistics Scale [8], Multifactorial Scale of Attitudes Toward Statistics [9], and Survey of Attitudes Toward Statistics [10]. Other studies relating to learning behaviour include characterization of anxiety in statistics such as Statistical Anxiety Rating Scale [11] and Statistics Anxiety Inventory [12].

However, in many empirical studies relating to the measurement of constructs, it is quite often that items do not accurately measure the construct and range of persons' trait. This is due to various reasons which include poor item design, or the possibility that the item is measuring a different construct. In an empirical study which involve small number of participants such as this, the possibility that the person responses to the items and the construction of the items may not calibrate well due to the small range of person ability. Hence this study attempts to perform a diagnosis to gauge students' attitude based on the calibration of responses and items on these instruments, namely, Statistics Attitude and Competency forms. These instrumetns were adminsitered on postgraduate students at a major local university in Malaysia.

II. Study Design and Method

This study used a descriptive and survey method to gather information from twenty-five students who are currently pursuing their postgraduate studies in the Science and Technology and Social and Management Science disciplines at Universiti Teknologi MARA (UiTM). The study involved various class activities which include teaching, in-class assignments, take home assignments, in-class tests and administration of pre-test and post-test questionnaires. The main activity is that respondents are required to attend a series of three classes covering five basic statistical topics where each lesson lasted for about three hours. At the beginning of the first lesson, students were assessed on their attitudes toward statistics where they need to respond to 27 items based on five-point Likert scale. After completing three lessons, the students were again assessed on their attititude towards learning statistics using the post-attitude towards statistics questionnaire. They were given between 10 and 15 minutes to fill up the questionnaire which was developed and adapted based on several other attitude toward statistics instruments [13]. The improvised instrument was subjected to a reliability and validity analysis based on Rasch model analysis using Winsteps 3.68.2 software.

III. RASCH MEASUREMENT MODEL AS A DIAGNOSTIC TOOL

Rasch analysis is based on a stochastic or probabilistic model where rasch measurement takes into account two parameters – test item difficulty and person ability. These parameters are assumed interdependent.However, separation between the two parameters is also assumed. For example, the items (questions) within a test are hierchically ordered in terms of their difficulty and concurrently, persons are hierchically ordered in terms of their ability. The separation is achieved by using a probabilistic approach in which a person's raw score on a test is converted into a success-to-failure ratio and then into the logarithmic odds that the person will correctly answer the items [14]-[18]. This is represented in a logit scale. When this is estimated for all persons, the logits can be plotted on one scale.

The items within the test can be treated in a similar manner by examining the proportion of items answered incorrectly and then converting this ratio into the logarithmic odds of the item being incorrectly answered. The logits can also be plotted on one scale. A person's logit score can then be used as an estimate of that person'sability and the item logit score can then be used as an estimate of that item's difficulty. Since person ability was estimated from the proportion of correct answers and items difficulty from the proportion of persons with incorrect answers, both these estimates are related and the relationship between them can be expressed as a mathematical equation, i.e., Rasch Unidimensional Measurement Model as follows,

$$P_{ni}(x_{ni} = 1 / B_n, D_i) = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}} \quad (1)$$

The model expresses the probability of obtaining a correct answer (1 rather than 0) as a function of the size of the difference between the ability (*B*) of the person (*n*) and the difficulty (*D*) of the item (*i*).This Rasch model is used to calculate person abilities, to calculate item difficulties, and then to plot the person abilities and item difficulties on the same scale. According to the model, the probability of a personbeing successful on a given item is an exponential function of the difference between that person's abilityand the difficulty of the item.

A Rasch analysis is a procedure for assessing the quality of raw score data using the Rasch model criteria such as fit statistics, z-standard residuals, and biserial correlations [19]. A thorough Rasch analysis involves checking the degree to which the data match a unidimensional measurement model, identifying and diagnosing sources of discrepancy, removing items or persons if they are degrading the overall quality of measurement, and finally, constructing measures which, to the degree that the data approximate the Rasch model, are both interval level and sample independent. Infit and outfit mean square fit statistics provide summaries of the Rasch residuals, responses that differ from what is predicted by the Rasch model, for each item and person. High mean square fit statistics indicate a large number of unexpected responses. High person mean square values indicate test takers who filled in responses randomly, have unusual gaps in their knowledge. Item infit mean square values between 1.5 and 2.0 are considered to be unproductive for measurement, and values higher than 2.0 are actually degrading [20].

The Rasch model is not intended to fit data or to be evaluated by how well it fits any particular data set. The Rasch model is a definition of measurement derived from the universally accepted measurement requirements that: (1) The measures of objects be free of the particulars of the agents used to estimate these measures and the calibrations of agents be free of the particulars of the objects used to estimate these calibrations, (2) The measures of objects and calibrations of agents function according to the rules of arithmetic on a common scale so that they can be analyzed statistically, and (3) Linear combinations of measures and calibrations correspond to plausible concatenations of objects and agents.

Rasch model provides two types of indices in order to help the researcher to determine whether there are enough items spread and enough spread ability among persons. Two types of reliability indices are person reliability index and item reliability index [20]. The person reliability index indicates the expectation of replicability of person ordering if the sample persons were given another parallel set of items measuring the same construct [20]-[21]. Instead of ability estimates well targeted by a suitable pool of items, person reliability also requries a large enough spread of ability across the sample. Thus when a line of inquiry in which some person score higher while some person score lower, it indicates that high person reliability with expectated consistency of person responses to the items. A well accepted measure of characteristic of persons' reliability is based on the Kuder-Richardson 20 formula.

Item reliability index, on the other hand, is based on the logit-normal model, known as the Rasch model. The reliability of measurement composed of binary data obeying the mixed effect model is illustrated in Equ. (2),

$$R_m = \frac{\text{var}[E(Y_i \mid A_i]}{\text{var}(Y_i)} \quad (2)$$

Similarly to the classical definition, there is the total observed variability in the denominator, and there is the part of the var (Y_i) due to variability of A_i in the numerator. The Separation Index, G, is the ratio of the unbiased estimate of the sample standard deviation to the root mean square measurement error of the sample. The separation is defined mathematically as [22]. The expected reliability, R and Rasch separation index, G can be obtained from Equ. (3)

$$R = \frac{G^2}{1 + G^2}$$

and

$$G = \sqrt{\frac{R}{1 - R}} \qquad (3)$$

The aspect of validity concerns the principle of invariance where item calibration invariance can also be investigated by testing for differential item functioning (DIF).

IV. ANALYSIS AND RESULTS

A. Pre- and Post Attitude towards Statistics

Table 1: Pre-Attitude toward Statistics

SUMMARY OF 19 MEASURED Persons

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	81.4	27.0	.01	.26	.99	-.2	.99	-.2
S.D.	7.0	.0	.48	.00	.44	1.8	.43	1.7
MAX.	90.0	27.0	.60	.27	1.73	2.4	1.69	2.3
MIN.	64.0	27.0	-1.19	.26	.25	-4.2	.25	-4.2

REAL RMSE .28 ADJ.SD .39 SEPARATION 1.37 Person RELIABILITY .65
MODEL RMSE .26 ADJ.SD .40 SEPARATION 1.54 Person RELIABILITY .70
S.E. OF Person MEAN = .11

Person reliability index moderately acceptable

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = .70

SUMMARY OF 27 MEASURED Items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	57.3	19.0	.00	.31	.99	-.1	.99	-.1
S.D.	12.3	.0	1.17	.01	.34	1.1	.33	1.1
MAX.	74.0	19.0	2.56	.36	1.59	1.7	1.58	1.7
MIN.	32.0	19.0	-1.59	.30	.50	-1.9	.50	-1.9

REAL RMSE .34 ADJ.SD 1.12 SEPARATION 3.33 Item RELIABILITY .92
MODEL RMSE .31 ADJ.SD 1.12 SEPARATION 3.58 Item RELIABILITY .93
S.E. OF Item MEAN = .23

Good item reliability index

UMEAN=.000 USCALE=1.000
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
513 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 1124.34 with 465 d.f. p=.0000

In the analysis of students' prior attitude toward statistics, Table 1 presents an overall information about whether the data showed acceptable fit to the model. The mean infit and outfit for person and item mean squares are expected to be 1.0 and for these data, they are all .99. The mean standardized infit and outfit are expected to be 0.0 and here they are -.2 for persons and -.1 for items. This indicates the items are overfit. This suggests that the data fit the model somewhat better than we would expect which may signal some redundancy - possibly redundant items. Redundancy gives us an indication of how we may trim items to reduce the length of the instrument.The standard deviation of the standardized infit is an index of overall mistfit for persons and items. Using 2.0 as a cut-off criterion, both persons (standardized infit standard deviation = .44) and items (standardized infit standard deviation = .34) show an overall acceptable fit.

Separation is the index of spread of the person positions or item positions. If separation is 1.0 or below, the items may not have sufficient breadth in position. For persons, separation is 1.37 for the data at hand (*real*), and is 1.54 when the data have no misfit to the model (*model*). Separation is still low representing small spread of items and persons along a continuum. Low separation index here indicate some redundancy in the items and less variability of persons on the trait. The item separation index is 3.33 which is larger, an indication of a broader continuum of items than for persons. This also indicates item difficulty can be separated into three levels. Person and item separation and reliability of separation assess instrument spread across the trait continuum. "Separation" measures the spread of both items and persons in standard error units. It can be thought of as the number of levels into which the sample of items and persons can be separated. For an instrument to be useful, separation should exceed 1.0, with higher values of separation representing greater spread of items and persons along a continuum. Lower values of separationindicate redundancy in the items and less variability of persons on the trait. Separation determines reliability. Higher separation in concert with variance in person or item position yields higher reliability. The person separation reliability estimate for this data is 0.65. This estimate can be further improved after removing some misfit response from the data. The item separation reliability estimate is 0.92 which indicate items are replicable for measuring similar traits.

Note that the mean for items is 0.0. The mean of the item logit position is always arbitrarily set at 0.0, similar to standard (z) score. The person mean is 0.01 which suggest that most items were well matched and easily endorsed or easy to agree with.

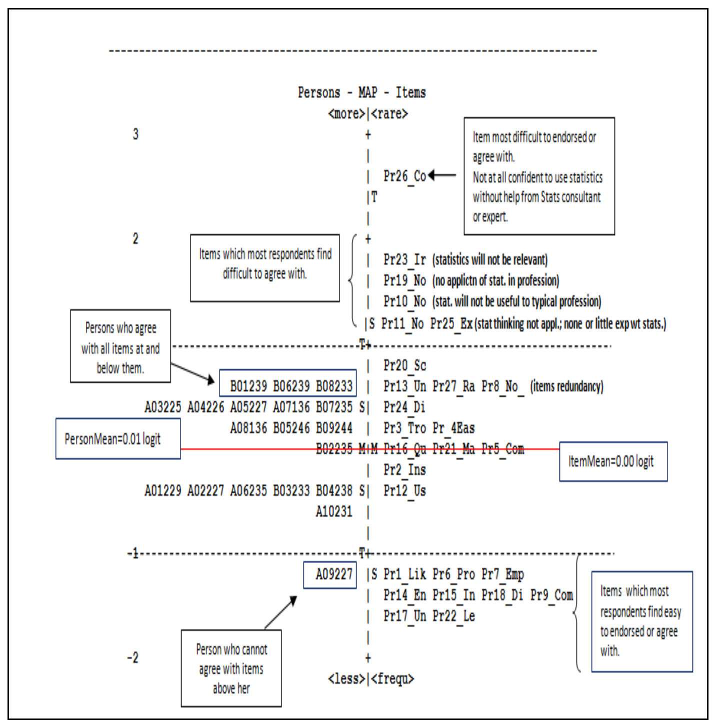

Fig 1. Person-Item Distribution Map of Pre-Attitude toward Statistics

Figure 1 illustrate the Person-Item distribution map which is the heart of the Rasch analysis. Persons are distributed on the left side of the logit ruler (center vertical line) and items are distributed on the right side."M' marks the person and item mean, "S" is one standard deviation away from the mean, and "T" is two standard deviation away from the mean. Those at the upper end of the scale agreed with more items and agreed more strongly.In the map, we can see that 99% of the respondents fall within the range of traits of 12 items. All persons are position below 5 items (Pr26, Pr23, Pr19, Pr10, Pr11). These are items most difficult to agreed upon by the respondents. It also indicates a low probability in the response to these items.There is also an indication of redundancy among the items as shown on the map. One or two redundant items can be dropped by looking at the fit statistics and retain items which fits the model. In this case, items that match these persons' levels of trait very well are items that range between -1 and +1.5 s.d. The items cover a range of -1.5 to +1.5 logits in difficulty, broader than the range of about -2.0 to +1.5 for persons. As such we may try to write items to match the range of persons' traits. Further diagnosis of item redundancy is discussed next.

Table 2: Pre-Attitude Toward Statistics - Item Misfit Order

ENTRY NUMBER	TOTAL SCORE	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
27	50	19	.66	.30	1.59	1.7	1.58	1.7	A .06	.34	31.6	48.6	Pr27 Rate
11	44	19	1.23	.31	1.56	1.6	1.54	1.6	B .16	.33	42.1	48.6	Pr11_Not appl
23	39	19	1.74	.33	1.51	1.5	1.49	1.5	C .24	.31	36.8	53.1	Pr23 Irrelevant
4	55	19	.21	.30	1.44	1.4	1.43	1.3	D .32	.35	42.1	47.5	Pr 4Easy
7	70	19	-1.18	.32	1.42	1.3	1.43	1.3	E .36	.34	36.8	50.8	Pr7 Employable
19	41	19	1.53	.32	1.42	1.3	1.38	1.2	F .24	.32	36.8	51.9	Pr19 No appl
5	57	19	.03	.30	1.39	1.3	1.38	1.2	G .20	.35	42.1	47.4	Pr5 Complicated
10	42	19	1.42	.32	1.23	.8	1.27	.9	H .00	.32	57.9	50.9	Pr10 Not useful
21	57	19	.03	.30	1.19	.7	1.20	.7	I .60	.35	52.6	47.4	Pr21 Mass comp
20	48	19	.85	.31	1.11	.5	1.11	.4	J .27	.33	47.4	48.6	Pr20_Scared
26	32	19	2.56	.36	.90	-.2	.97	.0	K-.03	.28	52.6	52.7	Pr26 Confident
22	74	19	-1.59	.33	.94	-.1	.94	-.1	L .03	.33	52.6	54.3	Pr22 Learn
2	60	19	-.24	.30	.93	-.1	.93	-.1	M .43	.35	52.6	45.8	Pr2 Insecure
6	71	19	-1.28	.32	.89	-.2	.90	-.2	N .54	.34	63.2	52.2	Pr6 ProfTrng
9	72	19	-1.38	.32	.88	-.3	.87	-.3	m .72	.33	57.9	53.1	Pr9 Communicate
1	71	19	-1.28	.32	.87	-.3	.87	-.3	l .57	.34	73.7	52.2	Pr1 Like
3	55	19	.21	.30	.82	-.5	.82	-.5	k .28	.35	42.1	47.5	Pr3 Trouble
16	58	19	-.06	.30	.81	-.6	.81	-.6	j .20	.35	47.4	47.1	Pr16 Quick
18	73	19	-1.49	.32	.79	-.6	.77	-.7	i .36	.33	63.2	53.8	Pr18 Discpline
25	45	19	1.13	.31	.79	-.6	.78	-.7	h-.07	.33	31.6	48.6	Pr25 Experience
13	51	19	.57	.30	.78	-.7	.77	-.7	g .39	.34	68.4	48.5	Pr13 Und stress
8	51	19	.57	.30	.70	-1.0	.69	-1.0	f .28	.34	52.6	48.5	Pr8 No idea
15	72	19	-1.38	.32	.62	-1.3	.61	-1.3	e .66	.33	73.7	53.1	Pr15 Interest
12	61	19	-.33	.30	.60	-1.4	.60	-1.4	d .55	.35	63.2	45.8	Pr12 Use
17	74	19	-1.59	.33	.55	-1.6	.57	-1.5	c .64	.33	63.2	54.3	Pr17 Understand
14	72	19	-1.38	.32	.51	-1.8	.52	-1.7	b .59	.33	68.4	53.1	Pr14 Enjoy
24	52	19	.48	.30	.50	-1.9	.50	-1.9	a .54	.34	63.2	47.6	Pr24 Difficult
MEAN	57.3	19.0	.00	.31	.99	-.1	.99	-.1			52.4	50.1	
S.D.	12.3	.0	1.17	.01	.34	1.1	.33	1.1			12.4	2.7	

Items overfit

Item fit is expressed as a mean square and as a standardized value. Item fit is an index of whether items function logically and provide a continuum useful for all respondents. An item may misfit because it is too complex, confusing, or because it actually measures a different construct.

Person fit to the Rasch model is an index of whether individuals are responding to items in a consistent manner or if responses are idiosyncratic or erratic. Responses may fail to be consistent when people are bored and inattentive to the task, when they are confused, or when an item evokes an unusually salient response from an individual.

Overfit is indicated by a mean square value < 1.0, and a *negative* standardized fit. Overfit is interpreted as too little variation in the response pattern, perhaps indicating the presence of redundant items. Underfit ("noise") is indicated by a mean square > 1.2 and standardized fit > 2.0 and suggests unusual and/or inappropriate response patterns.

In Table 2, 17 items were overfit as indication of too little variation in the response pattern and presence of redundant items. Dropping some items may increase the reliability index of persons. There were presence of gap between the items and this can be filled up to mark the level of trait. These evidences pointed to students' inability to precisely perceive their attitude due to redundant items thus restrict the measurement of the latent trait. The validity can be improved if similar items were removed or combined.

Table 3: Poorly Fitting Persons (Items In Entry Order)

NUMBER - NAME -- POSITION ------ MEASURE - INFIT (MNSQ) OUTFIT

17 B072352 .33 1.7 A 1.6

RESPONSE:	1:	4	3	3	1	1	4	4	3	5	2
Z-RESIDUAL:					-2	-2					
RESPONSE:	11:	2	3	5	4	4	3	5	5	1	4
Z-RESIDUAL:				2							
RESPONSE:	21:	5	4	2	2	2	2	3			
Z-RESIDUAL:		2									

Misfit items can be further diagnosed from Table 3. For person B072352, we expected a higher category selection for items 4, 5 (z-residual = -2) and a lower category selection for items 13, 21 (z-residual = 2). The person's logit position is 0.33 > PersonMean logit=0.01 meaning it tended to be easy for this person to agree with the items. Her overall infit and outfit are 1.7 and 1.6, respectively, indication of item underfit or noise. Therefore we can consider discarding this data.

Table 4 also illustrate person B072352 is at the top of the list of most misfitting response with infit and outfit value of more than 1.6 and z-std of more than 2.0. This is followed by four other persons in a similar range of fit indices.

Table 5 illustrates the person misfit order based on their responses to pre-attitude toward statistics items. Persons of high outfit meansquare (>1.47) are related to their misfit response strings. Persons who are expected to endorse or agree to the items have stated otherwise in their response towards the items. This suggests unusual or/and inappropriate and thus not truly measuring the latent trait of the person.

Table 4: Pre-Attitude toward Statistics - Person Misfit Order

ENTRY NUMBER	TOTAL SCORE	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%	Person
17	86	27	.33	.26	1.73	2.4	1.65	2.2	A .67	.66	59.3	51.0	B072352
16	90	27	.60	.26	1.72	2.4	1.68	2.3	B .80	.66	33.3	50.0	B062395
15	83	27	.12	.26	1.71	2.3	1.69	2.3	C .41	.66	33.3	51.2	B052464
6	74	27	-.49	.26	1.55	1.9	1.61	2.1	D .00	.66	29.6	48.7	A062354
5	88	27	.46	.26	1.52	1.8	1.47	1.7	E .34	.66	44.4	50.5	A052272
18	89	27	.53	.26	1.00	.1	1.04	.2	F .94	.66	48.1	50.4	B082335
3	87	27	.39	.26	1.01	.1	.99	.0	G .70	.66	44.4	50.7	A032253
4	87	27	.39	.26	.97	.0	.99	.0	H .68	.66	40.7	50.7	A042263
2	76	27	-.35	.26	.91	-.3	.98	.0	I .25	.66	40.7	49.0	A022273
1	76	27	-.35	.26	.91	-.3	.94	-.2	J .63	.66	55.6	49.0	A012293
19	83	27	.12	.26	.89	-.4	.88	-.4	i .84	.66	63.0	51.2	B092444
8	85	27	.26	.26	.88	-.4	.86	-.5	h .67	.66	55.6	51.2	A081361
10	72	27	-.63	.26	.84	-.5	.83	-.6	g .54	.66	55.6	48.1	A102313
13	76	27	-.35	.26	.71	-1.1	.72	-1.1	f .86	.66	59.3	49.0	B032335
9	64	27	-1.19	.27	.69	-1.2	.71	-1.1	e .71	.64	51.9	51.0	A092273
12	80	27	-.08	.26	.60	-1.7	.56	-1.9	d .78	.66	59.3	50.5	B022355
11	89	27	.53	.26	.55	-2.0	.52	-2.2	c .74	.66	66.7	50.4	B012394
14	75	27	-.42	.26	.42	-2.8	.44	-2.7	b .94	.66	74.1	48.8	B042385
7	87	27	.39	.26	.25	-4.2	.25	-4.2	a .83	.66	81.5	50.7	A071361
MEAN	81.4	27.0	.01	.26	.99	-.2	.99	-.2			52.4	50.1	
S.D.	7.0	.0	.48	.00	.44	1.8	.43	1.7			13.5	1.0	

Table 5: Person Most Misfitting Response Strings

```
Person          OUTMNSQ  |Item
                         |11     2  12211122
                         |85617215437010936
                     high-----------------
  17 B072352     1.65 A|......5115.......
  16 B062395     1.68 B|......5....1.....
  15 B052464     1.69 C|...22..5.........
   6 A062354     1.61 D|22.........44...
   5 A052272     1.47 E|...........5.44.
   4 A042263      .99 H|.....5...........
   2 A022273      .98 I|..........4.....3
   1 A012293      .94 J|..2..............
  19 B092444      .88 i|.........1......
   8 A081361      .86 h|.........1......
  11 B012394      .52 c|........5........
                        |-------------low-
                        |11617225412211122
                        |85     1  37010936
```

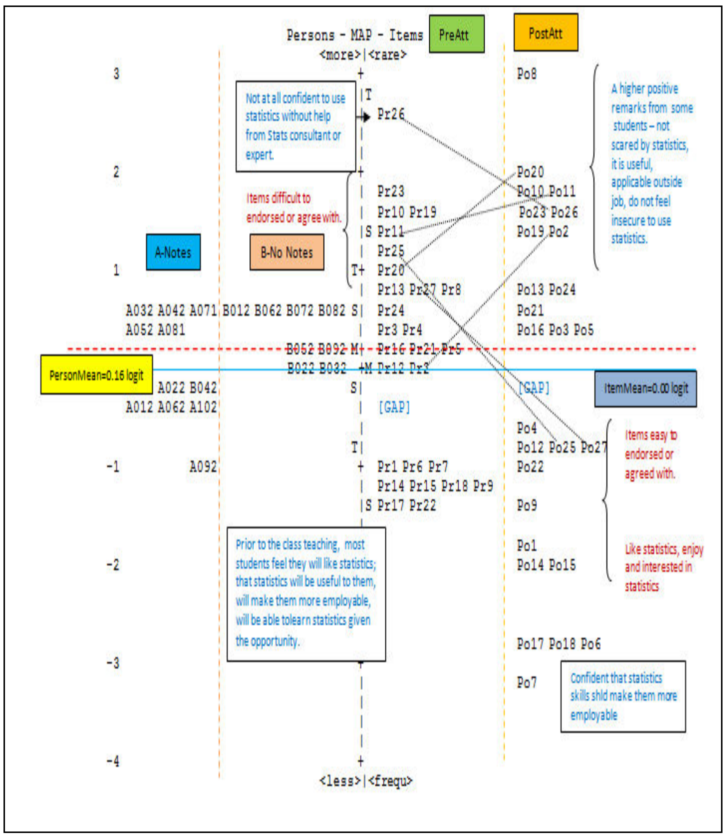

Fig 2. Person-Item Distribution Map of Pre- and Post-Attitude toward Statistics

Figure 2 shows that 48% (13/27) of the post-attitude items are located below MeanItem and 52% (14/27) are located above MeanItem. Comparatively, there are more post-attitude than pre-attitude items being endorsed which signals a more positive attitude towards learning statistics. Students gave a higher positive remarks after attending the statistics class such that they are no longer lost and scared about statistics. They also think that statistics is useful and statistical thinking is applicable in their life outside their profession. The students agreed that statistics also are relevant in their life and now more confident and secured when dealing with statistics problem.

On the other hand, there are less endorsement of the pre-attitude items (16 items) as these items are located above MeanItem while only 11 items (Pr1, Pr6, Pr7, Pr14, Pr15, Pr 18, Pr9, Pr17) are easily endorsed. Prior to the class, most students feel that will like statistics, that statistics will be useful to them and make them more employable. This is further justified after attending the class.

However, a few are still under stress during statistics class and find it a little difficult to understand the statistical concepts. They perceived that they have trouble understanding statistics because of the way they think. However, they think that statistics can be learned quickly by most people, and the formula are easy to understand. They also like statistics, interested to learn it and enjoy taking statistics courses. They are able to learn it and willing to communicate statistical information to others. They felt that they should use statistics in everyday life, and it should also be a required part of their professional training. However, they agreed that learning statistics require a great deal of discipline, and they require much more experience with statistics. Generally, there is moderate change of attitude towards learning statistics after attending the statistics class.

B. Perceived Importance in Statistics

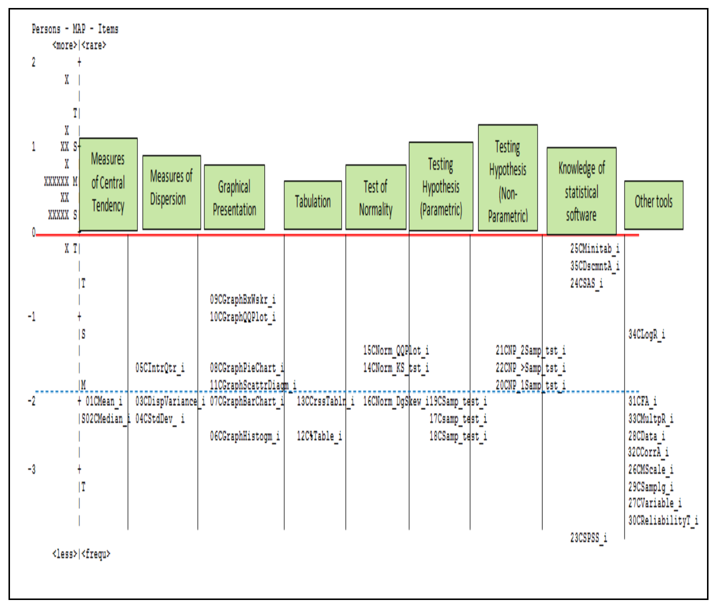

Fig 3. Person-Item Distribution Map of Perceived Importance in Statistics

Students' perceived importance in various statistical tools are compared across eight main topics as shown in Figure 3. The tools or items are calibrated against each student and displayed on a PID map based on the perceived knowledge scale ranging from (1) Not Important At All, (2) Not So Important, (3) Neutral, (4) Fairly Important, (5) Very Important.

Students' perceived importance can be clearly gauged from the position of each item based on the response scale. The majority of the students perceived the importance of all listed topics/concepts in the instrument. However, the level of perceived importance differ across the topics. Most items are located far below the MeanItem logit, an indication of students' agreement to the importance of those topics.

From Figure 3, it is observed that all students which is distributed on the left hand side of the logit ruler are located well above all items which is distributed on the right hand side of the logit ruler. All items are also located far below 0.00 logit and even below the lowest student which is at -0.28. This indicates that all students felt the importance of these statistical tools in research.

Among the descriptive tools, majority recognizes the importance of measures of central tendency, measures of dispersion, graphical presentations and tabulation. The most important among them all is pointed to Q12 (Table – Frequency/Percent Tables which is located at -2.43*logit*, while the least important item is Q25 Minitab which is at -0.23*logit*, -0.05*logit* below the lowest person *logit*. Most of the least important items are among the Knowledge on Statistical Software.

The students find the questions on Measures of Central Tendency to be quite important. Both the items are located below ItemMean=-1.71*logit* and the gap *logit* between ItemMean and Q01 is -0.36*logit* (-2.07 –(-1.71)). Within the Measures of Dispersion; Q3 DispVariance and Q4 StdDev are of importance in statistics compared to Q5 IntrQtr – Interquartile range which is deemed less important. In Graphical Presentation, 33% (2/6) of the items are perceived to be of importance by the students, while the rest 67% (4/6) are considered less important. The important items are Q6 GraphHistogram which is considered most important within Graphical Presentation, and Q7 GraphBarChart as the second most important item. On the other hand, Q11 Scatter Diagram, Q8 Pie Chart, Q10 QQ Plot, Q9 Box-and-Whisker Plot are perceived less important by the students. Q13 Cross tabulation and Q12 Percentage Table are both perceived to be of importance within the Tabulation topic.
In Test of Normality, only Q16 Degree of Skewness is perceived as important compared to Q14 Kolmogorov-Smirnov test/Shapiro Wilks and Q15 QQ Plot.

For Testing Hypothesis (Parametric), all three items [Q18 two sample test (t-test, z-test, paired t-test), Q17 one sample test (Sign), and Q19 more than two sample test (1-way Anova, 2-way Anova)] are perceived as important by the students. Testing Hypothesis (Non-Parametric), all of the items are perceived to be less important in statistical analysis. The less important appears to be Q21 of two sample test (mann-Whitney, Wilcoxon), followed by Q22 more than two sample test (Kruskal-Wallis) and last is Q20 one sample test.

Under Knowledge of Statistical Software, 31% (4/13) perceived it as less important and 69% (9/13) perceived as important. The items which are considered to be of importance are Q23 SPSS, Q30 Reliability Test, Q27 Knowledge of Variables, Q29 Knowledge of Sampling Techniques, Q26 Knowledge of Measurement Scales, Q32 Correlation Analysis, Q28 Knowledge of Types of Data, Q33 Multiple Regression, and Q31 Factor Analysis. While the less important items are Q34 Logistic Regression, Q24 SAS, Q35.

C. *Perceived Knowledge in Statistics*

Students' perceived knowledge in various statistical tools are compared across eight main topics as shown in Figure 4. The tools or items are calibrated against each student and displayed on a PID map based on the perceived knowledge scale ranging from (1) No knowledge at all, (2) Little Knowledge, (3) Neither, (4) Some Knowledge, to (5) A Great Deal of Knowledge.

Students' perceived knowledge can be clearly gauged from the position of each item based on the response scale. The majority of the students agreed that they have a great deal of knowledge in measures of central tendency and some knowledge in measures of dispersion, graphical presentation and tabulations. These items are either located close to the MeanItem logit or far below the MeanItem.

Few were comfortable with Test of Normality but majority were unsure of the concept. Majority perceived that they have some knowledge on Parametric tests as the items are position slightly below the MeanItem. However, half of them perceived that they have little knowledge on the non-paramteric tests. More than half perceived that they either have little knowledge or no knowledge at all on certain statistical softwares with majority having no knowledge at all on most multivariate topics.

On the Knowledge of Statistical Software, the students perceived that they are less knowledgeable in 54% (7/13) of the items. The least knowledgeable item is Q25 Minitab, and the rest are SAS, Logistic Regression, Discriminant Analysis, Multiple Regression, Correlation Analysis, and Factor Analysis. On the hand, the items which are found to be quite knowledgeable to the students are related to the Types of Data, Identification of Variables, Sampling Techniques, Measurement Scale, Reliability Test, and SPSS software.

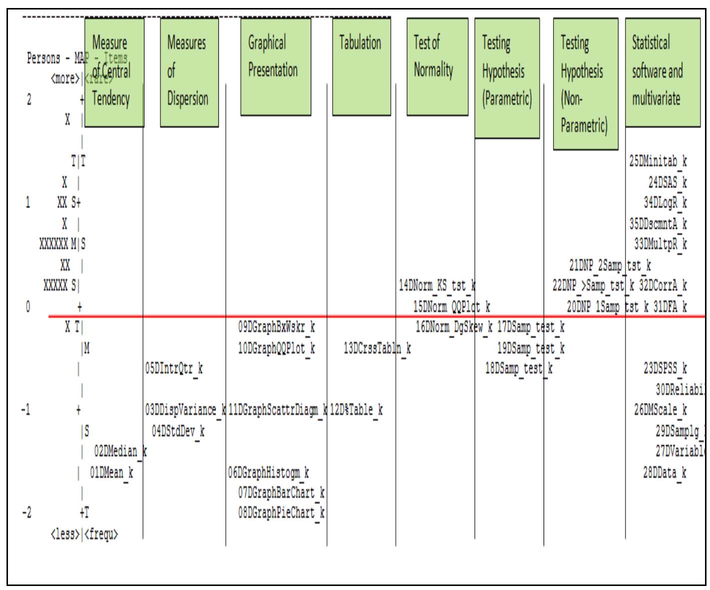

Fig. 4 Person-Item Distribution Map of Perceived Knowledge in Statistics

D. *Model and Empirical Item Characteristic Curve*

Further analysis was carried out by investigating both the model and empirical ICCs between perceived knowledge and students actual ability in variables and types of data prior to the remedial class (see Figure 5). Students perceived themselves as having good knowledge in the topic, however this was not reflected in their actual ability prior to the remedial class, as illustrated by the low empirical scores on the items relating to the topic on variables and types of data.

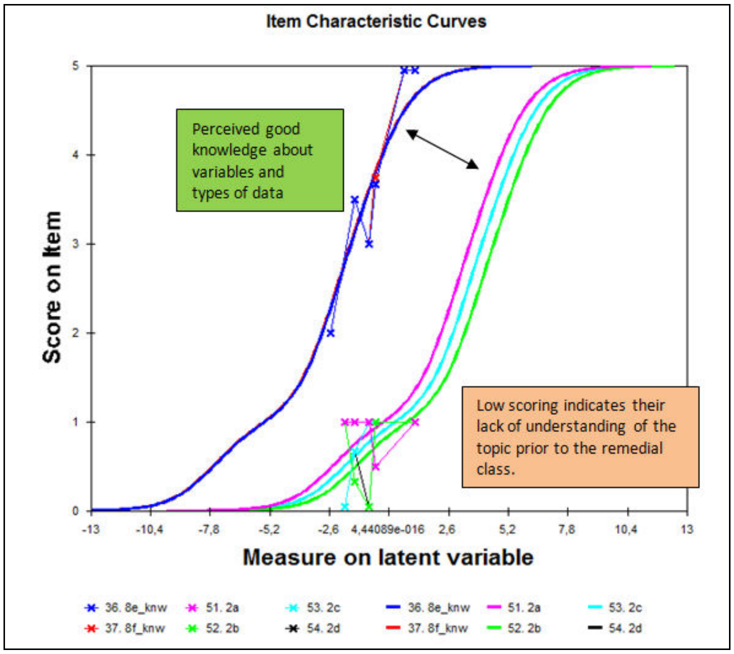

Fig. 5 Model-Empirical ICC for perceived knowledge and ability

Figure 6 illustrates a wide gap between perceived importance and knowledge of students towards the Test of Normality topic. The students realized the importance of this topic in their research, however they admitted that their level of knowledge on the subject matter is fairly low, hence the neccesity to learn more about the topic in future remedial classes.

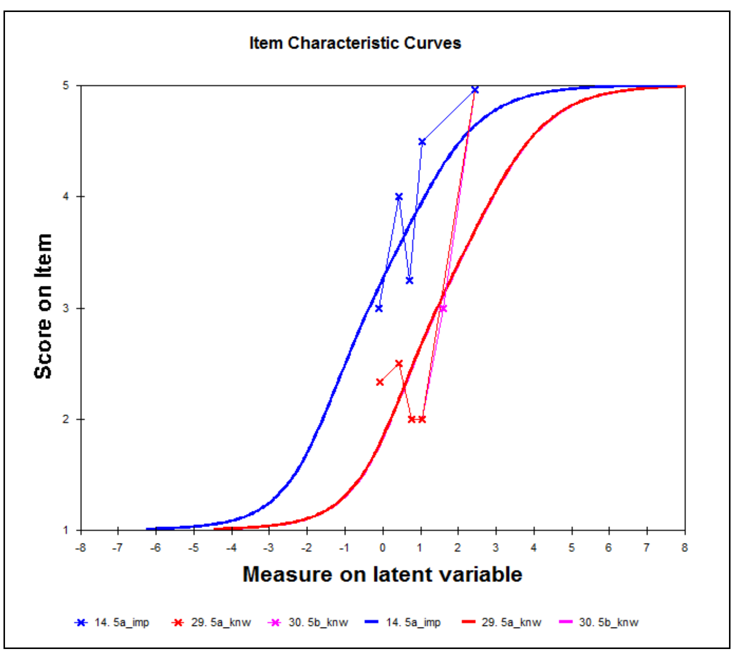

Fig. 6 Model-Empirical ICC for perceived importance and knowledge on Test of Normality

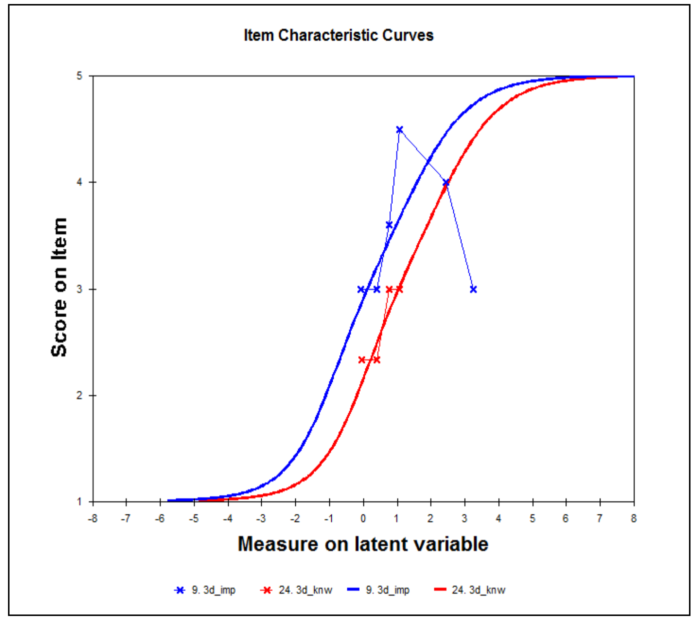

Fig. 6 Model-Empirical ICC for perceived importance and knowledge on Box-and-Whisker Plot

As in the case of of Box-and-Whisker Plot, the model and empirical ICC indicate a narrower gap between perceived importance and knowledge. Majority perceived the importance of using Box Plot in their research analysis, but again they have limited knowledge on the subject matter. As such, emphasis on indepth understanding of Box Plot is required and should be made in future remedial classes.

V. Conclusion

This study has shown that true measurement of students' attitude towards learning statistics can be precisely measured using the Rasch probabilistic model. With specific logit ruler being developed for this study, it is able to validate a question construct and calibrate students' responses with the agent of response, namely the test items based on PIDM map. This study has also shown that Rasch analysis is able to gauge and diagnose the extent of students' attitude towardes learning statistics prior to and after attending the classes. The range of attitude based on the differences in logit scores between the pre- and post attitude items has enabled the instructor to gauge students' attitude to the exact scale intensity. The person responses towards the items were found to be highly reliable as indicated by the high reliability index and therefore the responses given were truly measuring the latent trait of students' attitude based on their endorsement of the items which are clearly depicted in the PIDM map. Rasch analysis also has the ability to pick up traces of misfit answers due to unusual or inappropriate responses.

Knowledge can be quite successfully gauged based on students' perception towards the items. Even though perception can be quite subjective to a certain extent, it can however provide some insights and understanding on the situation. This can help the researcher to develop a better instrument to gauge knowledge more effectively. Though every effort was taken to ensure all students provide their responses, there are some who were reluctant to do so and hence the occurrence of missing data. However, due to Rasch predictive property, it has enabled us to gauge students' true response irrespective of whether or not the students responded to the items. Rasch has this particular predictive properties embedded in the model to make it a very reliable validation model for both the person and item response string. This study has shown that Rasch measurement model has the capability to rigorously analyse test items and responses to the items more accurately thus making evaluation more comprehensive with better accuracy.

REFERENCES

[1] S.E.Cashin and P.B. Elmore, "The survey of attitudes toward statistics scale: A construct validity study", *Educational and Psychological Measurement,* vol. 65, no. 3, pp. 509-524, 1997.

[2] J.A. Fullerton and D.Umphrey, "An analysis of Attitudes toward statistics: gender differences among advertising majors", (ERIC Document Reproduction Service No. ED 456 479), 2001.

[3] J. Brandsma, "Data collection and analysis: Examining community college students' understanding of elementary statistics through laboratory activities, 2000 (ERIC Document Reproduction Service No. ED 461 396).

[4] R.L. Kennedy and C.J. McCallister, "Attitudes toward advanced and multivariate statistics using computers, (ERIC Document Reproduction Service No. ED 464 097), 2001.

[5] T.S.Brown and J.T. Brown, "Prerequisitie course grades and attitudes toward statistics", *College Student Journal*, vol. 29, pp. 502-507, 1995.

[6] N.H. Othman and Z. Mahmud, "Statistical Competency And Attitude Towards Learning Elementary Statistics: A Case of SMK Bandar Baru Sg Buloh", *Proceedings of the Regional Conference on Statistical Sciences (RCSS'10),* pp.335-348, 2010.

[7] N. L. Abu Kassim, N.Z. Ismail, Z. Mahmud , M.S. Zainol, " Measuring students' understanding of statistical concepts using Rasch Measurement", *International Journal of Innovation and Management Technology*, vol.1,no.1, pp. 13-19, 2010.

[8] S.L.Wise, "The development and validation of a scale measuring attitudes toward statistics", *Educational and Psychological Measurement*, vol. 45, pp. 401-405,1985.

[9] E. Auzmendi, "Factors related to Statistics: A study with a Spanish sample", *Paper presented at the meeting of the American Educational Research Association*, Chicago, 1991.

[10] C. Schau, T.L. Dauphinee and A. Del Vecchio, "Survey of attitudes toward statistics (SATS)", 1993.
[Online: http://www.unm.edu/~cschau/downloadsats.pdf]

[11] J.R. Cruise, R.W. Cash, and L.D. Bolton, "Development and Validation of an Instrument to Measure Statistical Anxiety," in *American Statistical Association Proceedings of the Section on Statistical Education*, pp. 92-98, 1985.

[12] M.Zeidner, "Statistics and mathematics anxiety in social science students: some interesting parallels", *British Journal of Educational Psychology*, vol. 61, pp.319-328, 1991.

[13] C. Schau, " Students' attitudes: The other important outcome in statistics education", *Paper presented at the Joint Statistical Meetings, San Francisc,* 2003.

[14] Z.Mahmud, M.S. Masodi, and A.A.Aziz, "Overcoming measurement hurdles in statistical education research", *WSEAS Computers and Simulations in Modern Science*, vol. 5, pp. 78-85, 2011.

[15] A. A. Aziz, A.Mohamed, N. Arshad. S.Zakaria., A. Zaharim, H.A.Ghulam, M.S.Masodi. Application of Rasch Model in Validating the Constructs of Measurement Instrument. *International Journal of Education and Information Technologies*, Vol. 2, No. 2, pp. 105-112, 2008.

[16] M.S.Masodi, A. Mohamed, A.A. Aziz, N.Arshad, H.A. Ghulman & S.Zakaria, "Application of Rasch Model in validating the construct of measurement instrument", in *International Journal of Education and Information Technologies,* Issue 2, Volume 2,. pp. 105-112, 2008.

[17] M.S.Masodi, A. Mohamed, A.A. Aziz, N.Arshad, & S.Zakaria,, "Appraisal of Course Learning Outcomes using Rasch measurement: A case study in Information Technology Education", *International Journal of Systems International Journal of Systems Applications, Engineering & Development*; Issue 4, vol.1, University Press, UK. pp.164-172, 2007.

[18] J. Sick, "Assumptions and requirements of rasch measurement, *JALT Testing & Evalution SIG Newsletter*, vol.14, no.2, 2010.

[19] B.D. Wright, and J.M. Linacre, Reasonable mean-square fit values. *Rasch Measurement Transactions,* vol. 8, no. 3, p.370, 1994.

[20] T.G. Bond and C.M.Fox, *Applying the Rasch model: Fundamental measurement in the Human Sciences*, New Jersy: Lawrence Erlbaum Associates, 2001.

[21] B.D. Wright and G.N. Masters, *Rating Scale Analysis : Rasch Measurerment,* Chicago : MESA Press, 1982.

[22] B.D. Wright and M. Stone, *Measurement Essentials*, 2nd Ed. *Test Design.* MESA Press, Chicago, 1999.

Zamalia Mahmud is currently an Associate Professor of Applied Statistics at Universiti Teknologi MARA Malaysia. She obtained her PhD in Statistical Education from the University of Strathclyde, Scotland in 1997. She is currrectly a member and reviewer of WSEAS, International Statistical Institute, Voorburg Netherland, Malaysia Institute of Statistics, and also an active reviewer of several local and international social and scientific journals. Her research interest is in the teaching and learning of statistics, measurement of statistical teaching and learning behaviour, and assessment of teaching and learning of statistics instruments using Rasch analysis.