International Journal of Research and Review
www.gkpublication.in E-ISSN: 2349-9788; P-ISSN: 2454-2237

Research Paper

Statistical Item Analysis of Basic & Integrated Science Process and Skill Test (BISPST) Using Rasch Model and SPSS

Rommel S. de Gracia[1], Melanie G. Gurat[2]

[1]Schools Division of Nueva Vizcaya, Philippines
[2]Mathematics Department and University Research Center, Saint Mary's University, Nueva Vizcaya, Philippines

Corresponding Author: Melanie G. Gurat

Received: 26/10/2016 *Revised: 12/11/2016* *Accepted: 12/11/2016*

ABSTRACT

Item analysis is a general term that refers to the specific methods used in education to evaluate test items, typically for the purpose of test construction and revision. Construction and analysis of science examinations in high school curricula needs the integration of integrated process and skill type which help students develop their rational and logical thinking.
This exploratory study aimed to investigate the comparative results of item analyses using Rasch Model Analysis and SPSS. It examined the reliability, mean difficulty, mean ability, discrimination index, mean trend scores, *infit* and *misfit*, and *item-person* and *fit maps* using the results of the Basic & Integrated Science Process and Skill Test (BISPST) administered to secondary science high school teachers in the Division of Nueva Vizcaya.
The Rasch Model Analysis and SPSS revealed the same unsatisfactory value of internal consistency or the reliability coefficient *Cronbach's alpha* (0.66 and 0.67), mean score (19.52) and standard deviation (4.53) of the BISPST. The mean item difficulty was set at 0 by definition (SD = 2.02) and the mean examinee ability was 1.56 (SD = 0.98), which means that the examinees were able to correctly answer 61% of items on average. 33 items with difficulty lower than the least able examinee. The mean ability using Rasch Model Analysis per item showed that high performers in the test got the correct answers compared to the low performing group. This is the same and consistent with the results of the mean trend scores of the upper, middle and lower groups as generated by the SPSS. Choices in the BISPST were found to be effective distracters. Although most items needs revision as revealed by the values generated by both softwares, no items were found to be misfit in the distribution.

Keywords: *Item Analysis, integrated process and skill, mean ability, mean trend, effective distracters, infit, misfit, threshold, Rasch Model Analysis*

INTRODUCTION

The K to 12 Science and Mathematics Curricula comprise different strands of which commonly fall in developing process and skills in learning the different competencies. Student achievement in these strands can be measured using an assessment tool. The result of this assessment can be used to improved curriculum instruction. Test construction involves making decisions based on knowledge of content and analysis of items.

The construction of an assessment tool for Science and Mathematics curricula involved the following processes: (1) creation of test items aligned with the national curriculum as blueprint and existing assessment tool as model, (2) statistical item analyses, (3) examining and interpreting differences in difficulty of test items in the context of curriculum content, and (4) decision-making to finalize test items (Par, Rosary, Fua, Cagasan, Peclaro, Care and Vista, 2014).

Reliability and validity are central to the formulation of an assessment tool. Knowledge on the reliability and validity of a test can greatly help teachers in building future tests or exams. According to Par et. al (2014) determining the item difficulties paved the way to the generation of developmental skills progression, from concrete to more abstract thinking in Science and Mathematics context. This is because the result on assessment can be used as instructional emphases inside the classroom setting.

Item analysis is considered a powerful technique used by instructors for the improvement of their instruction. It also refers to the specific methods as a means to evaluate test items, usually for test construction and revision. Items must be valid and it measures what it is intended to be measured. Furthermore, items must also be diagnostic which means that it must be a guide of the misunderstanding through the selection of wrong options. It must also be a prescriptive of appropriate remediation. In addition, result of item analysis may greatly improve instructors or teachers who construct their own examinations.

Bilgin (2006) defined Science process skills test as understanding of methods and procedures of scientific investigation. Harlen (2000) include abilities relating to identifying investigable questions, designing investigations, obtaining evidence, interpreting evidence in terms of the question in the inquiry and communicating the investigation process. Process skill learning, an important component of science curricula at all levels of high school was deemed necessary with the introduction of a paper-and-pencil test to see the acquisition of skill.

According to Livermore as cite by Harlen (2000) basic science process skills deliver a foundation for learning the integrated science process skills. Process which includes observing, classifying and recording data perform as prerequisites for integrated processes.

Several science curriculum guides and textbooks have cited significantresultson the acquisition of integrated science and process skills. These processes are rooted in the simple processes and deemed necessary to the purpose of acquiring a scientific approach to knowledge. This is for the reason that skills represent the rational and logical thinking used in science.

Interestingly, with the pending upshots from the K to 12 initiatives because of its ongoing implementation, curiosity leads to the investigation of analyzing test items of Basic and Integrated Science Process and Skill Test using Rasch Model and using SPSS software.

Objective of the Study

This study aimed to investigate the comparative results of item analyses using Rasch Model Analysis and using the SPSS software. The study explored the reliability, mean difficulty, mean ability, discrimination index, mean trend scores, *infit* and*misfit*, threshold and *item-person* and *fit maps* using the results of the BISPST.

METHODOLOGY

***Research Design*:** This study is a descriptive comparative type and exploratory in nature.

***Participants*:** One hundred fifty one secondary science teachers in the Department of Education Division of Nueva Vizcaya were involved in the study.

***Measures*:** The main tool of the study was the Basic and Integrated Science Process and Skill Test (BISPST). Primarily, it consisted of publicized sample items of the BISPST.

***Procedures*:** The BISPST was administered during the Science Teachers Congress held at Nueva Vizcaya General Comprehensive High School last Decembe13, 2014. Results of the test were used in item analysis.

Data Analysis: Descriptive Statistics was used like mean scores and standard deviations, frequency counts, mean ability, discrimination index, and mean trend scores.

Data were run through two different softwares; Quest and SPSS. Rasch model analysis was used to describe infit, misfit, thesold and item person map.

The *Index of Discrimination* is the difference between the proportion of the upper group who got an item right and the proportion of the lower group who got the item right.

Reliability Coefficient is measured in terms of the proportion of true score variability that is captured across subjects or respondents, relative to the total observed variability.

In test construction, *item difficulty* is determined by the number of people who answer a particular test item correctly.

Threshold is also known as the difficulty parameter or the threshold parameter. This value tells us how easy or how difficult an item is.

The *infit* mean-squared is the Chi-squared/degrees of freedom with weighting, in which a constant is put into the algorithms to indicate how much certain observations are taken into account. A *misfit* is an observation that cannot fit into the overall structure of the exam.

Item-Person and *Fit Maps* display the item difficulty parameter expressed in *logit* scale.

RESULTS AND DISCUSSION

Table 1 shows the reliability coefficients, mean score and standard deviation of scores obtained in BISPST as generated by the QUEST and SPSS softwares.

Table 1. Comparison of Internal Consistency, Mean Score and Standard Deviation Generated by Quest and SPSS

Statistics	Rasch Model Analysis	SPSS
Internal Consistency	0.670	0.664
Mean Score	19.52	19.52
Standard Deviation	4.39	4.39

Table 2. Item Analysis on Discrimination, Mean Ability, Difficulty and Decision in Basic and Integrated Science Process Skills Test

Item	Difficulty (f% of correct answer)	Discrimination (Disc)	Mean Ability	Decision
Item 1	87.9(VE)	0.12 (VL)	0.57	Discard
Item 1	87.9 (VE)	0.12 (VL)	0.65	Discard
Item 2	74(E)	0.27(L)	0.64	Discard
Item 3	64.4(E)	0.14(VL)	0.61	Discard
Item 4	64(E)	0.45(MH)	0.77	Revise
Item 5	29.7(D)	0.21(L)	0.78	May need revision
Item 6	60.3(E)	0.43(MH)	0.78	Revise
Item 7	71.3(E)	0.26(L)	0.66	Discard
Item 8	87.4(VE)	0.31(L)	0.62	Discard
Item 9	85.8(VE)	0.27(L)	0.63	Discard
Item 10	82.4(VE)	0.32(L)	0.64	Discard
Item 11	83.3(VE)	0.19(VL)	0.62	Discard
Item 12	43.9(MD)	0.36(L)	0.82	Revise
Item 13	20.8(D)	0.22(L)	0.86	May need revision
Item 14	76(E)	0.16(VL)	0.61	Discard
Item 15	61.3(E)	0.35(L)	0.74	Discard
Item 16	45.3(MD)	0.18(VL)	0.69	Revise
Item 17	61(E)	0.42(MH)	0.78	Revise
Item 18	21.5(D)	0.11(VL)	0.72	May need revision
Item 19	42.6(MD)	0.08(VL)	0.62	Revise
Item 20	40.5(MD)	0.24(L)	0.76	Revise
Item 21	67.3(E)	0.37(L)	0.72	Discard
Item 22	84.6(VE)	0.24(L)	0.61	Discard
Item 23	43.6(MD)	0.15(VL)	0.66	Revise
Item 24	48.7(MD)	0.52(MH)	0.92	May need revision
Item 25	50(MD)	0.17(VL)	0.67	Revise
Item 26	79.5(E)	0.3(L)	0.69	Discard
Item 27	24(D)	0.16(VL)	0.75	May need revision
Item 28	54.5(MD)	0.34(L)	0.76	Revise
Item 29	78.5(E)	0.51(MH)	0.73	Revise
Item 30	59(MD)	0.38(L)	0.76	Revise
Item 31	75.5(E)	0.38(L)	0.7	Discard
Item 32	64.3(E)	0.51(MH)	0.81	Revise
Item 33	66.9(E)	0.33(L)	0.7	Discard

Table 1 shows that the internal consistency or the reliability coefficient *Cronbach's alpha*, mean score and standard deviation generated using Quest and SPSS were the same. The difference in internal consistency was only 0.006. The test reliability was an unsatisfactory 0.66-0.67, which means that the test results were not so reliable (Yang, 2010). To improve the test reliability, increasing the item numbers should be considered.

Consistency of reliability coefficients using the Rasch model analysis and SPSS indicates that both may be utilized in in analyzing reliability, mean score and standard deviation.

The overall difficulty index of the test is about 61%. Based from the discrimination and difficulty indices, table 2 shows that the decision for items 1 (f % = 87.9, Disc = 0.12), 2(,3,7,8 ,9,10,11,14,15 ,21,22,26,31 and 33 was to discard. While items 4,6,12,16,17,19,20,23,25,28,29,30 and 32 were recommended to revise and items 5,13,18,24 and 27 may need revision.

This also reveals that almost all items in the basic and integrated science process skills test did not meet the standard of a good test. However, not all the items will be discarded.

Arranged from highest to lowest mean ability of takers who got a correct items in the specified item, item 24 (MA=0.92) was the highest, followed by item 13 (MA=0.86) item 12 (MA=0.82), item 32(MA=0.81), items 5 ,6 and 17 (MA=0.78), item 4(MA=0.77), items 20, 28 and 30 (MA=0.76), item 27 (MA=0.75), item 15 (MA=0.74), item 29 (MA=0.73), items 18 and 21 (MA=072), items 31 and 33 (MA=0.70), items 16 and 26 (MA=0.69), item 25 (MA=0.67), items 7 and 23 (MA=0.66), items 2 and 10 (MA=0.64), item 9 (MA=0.63), items 8,11 and 19 (MA=0.62), item 3, 14 and 22 (MA=0.61) and item 1 (MA=0.57).

Table 3. Item Analysis on Mean Trend Scores in Basic and Integrated Science Process Skills Test

Group	Upper Group Mean (n=54)	Middle Group Mean (n=60)	Lower Group Mean(n=37)	Total Mean(n=151)	Trend
Item1	.8889	.8833	.8108	.8675	Decreasing
Item2	.8333	.7667	.5676	.7417	Decreasing
Item3	.7407	.5833	.6216	.6490	Erratic
Item4	.8333	.6667	.2973	.6358	Decreasing
Item5	.3704	.3000	.1892	.2980	Decreasing
Item7	.8519	.6667	.5676	.7086	Decreasing
Item8	.9630	.8333	.7568	.8609	Decreasing
Item9	.9259	.7500	.6486	.7881	Decreasing
Item10	.9259	.8000	.6486	.8079	Decreasing
Item11	.9630	.7667	.7297	.8278	Decreasing
Item12	.6481	.3333	.2703	.4305	Decreasing
Item13	.2963	.1333	.1622	.1987	Erratic
Item14	.8519	.7500	.6216	.7550	Decreasing
Item15	.8333	.5000	.4865	.6159	Decreasing
Item16	.5556	.4167	.3243	.4437	Decreasing
Item17	.7963	.5333	.3784	.5894	Decreasing
Item18	.2778	.1667	.2162	.2185	Erratic
Item19	.4259	.5000	.2703	.4172	Erratic
Item20	.5741	.2833	.3243	.3974	Erratic
Item21	.7963	.7667	.3243	.6689	Decreasing
Item22	.8704	.8667	.7027	.8278	Decreasing
Item23	.4444	.5000	.3243	.4371	Erratic
Item24	.8148	.3500	.1892	.4768	Decreasing
Item25	.6111	.4000	.4324	.4834	Erratic
Item26	.9444	.6667	.6486	.7616	Decreasing
Item27	.2593	.2333	.1892	.2318	Decreasing
Item28	.7222	.4667	.2432	.5033	Decreasing
Item29	.9630	.8167	.3243	.7483	Decreasing
Item30	.7778	.5667	.2432	.5629	Decreasing
Item31	.9444	.6833	.4324	.7152	Decreasing
Item32	.9259	.5333	.2703	.6093	Decreasing
Item33	.7963	.6833	.3243	.6358	Decreasing

*Erratic patternoccurs if the mean score from upper to lower group is not decreasing

The mean ability per item indicates that high performers in the test got the correct answer. The item with higher mean ability also means that more students who belong to high performing group got the item. Third, the examination was relatively easy for most of the respondents. To enhance the discrimination of the test, item difficulty should be adjusted to promote usefulness of the exam

Table 3 shows the mean performance of the upper, middle and lower group in a particular item as generated by the SPSS software.

The mean item difficulty was set at 0 by definition (SD = 2.02) and the mean examinee ability was 1.56 (SD = 0.98), which means that the examinees were able to correctly answer 61% of items on average. This is consistent with the findings advanced in Table 2.

As gleaned from the table, majority of the test items on Basic and Integrated Science Process Skills Test has decreasing mean trend from the "upper group" to the "lower group". This indicates that "upper group" has highest mean score on the test items 1, 2, 4, 5,6-12, 14-17, 21, 22, 24, 26-33, followed by the "middle group" and lastly the "lower group". Furthermore, items on this test are not questionable. However, items 3, 13, 18, 19, 20, 23 and 25 seem to be questionable because either the "middle group" or the "lower group" scored higher than the "upper group".

Items with erratic pattern of mean scores from the upper to lower groups were found not to have significant difference as shown in Table 4.

Table 4. ANOVA of Items with Erratic Mean Trend Pattern

Sources of Variation		Sum of Squares	df	Mean Square	F	Sig.
Item3	Between Groups	.741	2	.370	1.629	.200
	Within Groups	33.656	148	.227		
	Total	34.397	150			
Item13	Between Groups	.820	2	.410	2.614	.077
	Within Groups	23.220	148	.157		
	Total	24.040	150			
Item18	Between Groups	.351	2	.176	1.022	.363
	Within Groups	25.437	148	.172		
	Total	25.788	150			
Item19	Between Groups	1.214	2	.607	2.531	.083
	Within Groups	35.501	148	.240		
	Total	36.715	150			
Item23	Between Groups	.711	2	.355	1.444	.239
	Within Groups	36.441	148	.246		
	Total	37.152	150			
Item25	Between Groups	1.394	2	.697	2.841	.062
	Within Groups	36.314	148	.245		
	Total	37.709	150			

*significant at 5% level

Table 4 shows the Analysis of Variance for items in the BISPST with erratic mean trend pattern. Items 3, 13, 18, 19, 23, and 25 have insignificant mean scores as evidently shown by the $p>0.05$ (df = 2, 148). This implies that the mean scores of the upper, middle and lower groups do not significantly differ and erratic mean trend pattern for this matter is not doubtful. Thus, Post-Hoc analysis was no longer administered.

Table 5 shows the effectiveness of a certain choice as distracter in the particular test items.

Table 5 reveals if the options per item in the basic and integrated science process skill test were effective distracters, confusing or ineffective distracters. The result shows that all options in items 4,5,11,13,14,15,16,18,19,20,24,27 and 28 were effective distracters. The table also shows that there were at least one confusing distracters in items 1-3,6-10,12,17,22,23,25,26 and 29-33 while one ineffective distracter were included in item 21.

Table 5. Item Analysis on the Distracters in the Basic and Integrated Science Process Skills Test

Item	Mean Ability (MA)				
	A	B	C	D	Missing
Item 1	0.13 (E)	-0.31(C)	0.57*	0.43(E)	0.74
Item 2	0.64*	-0.07 (C)	0.24(E)	0.35(E)	2.28
Item 3	0.61*	0.29(E)	-0.14 (C)	0.54(E)	1.06
Item 4	0.36(E)	0.22(E)	0.01(E)	0.77*	1.28
Item 5	0.44(E)	0.78*	0.36(E)	1.17(E)	0.54
Item 6	0.78*	-0.1 (C)	0.39(E)	-0.19(C)	NA
Item 7	0.45(E)	0.66*	0.06(E)	-0.69 (C)	-0.31
Item 8	0.08(E)	-0.54 (C)	0.06(E)	0.62*	NA
Item 9	0.63*	0.01(E)	-0.2 (C)	0.33(E)	0.4
Item 10	-0.07 (C)	0.12(E)	0.64*	0.01(E)	0.76
Item 11	0.04(E)	0.62*	0.05(E)	0.56(E)	-1.69
Item 12	0.19(E)	0.46(E)	0.82*	-0.21(C)	0.68
Item 13	0.54(E)	0.33(E)	0.11(E)	0.86*	0.65
Item 14	0.32(E)	0.51(E)	0.3(E)	0.61*	0.69
Item 15	0.74*	0.26(E)	0.13(E)	0.25(E)	0.69
Item 16	0.69*	0.45(E)	0.27(E)	0.4(E)	0.4
Item 17	0.15(E)	-0.01 (C)	0.78*	0.31(E)	0.52
Item 18	0.6(E)	0.41(E)	0.4(E)	0.72*	0.48
Item 19	0.62*	0.05(E)	0.57(E)	0.52(E)	0.42
Item 20	0.46(E)	0.76*	0.42(E)	0.24(E)	0.63
Item 21	0.72*	0.36(E)	0 (I)	0.16(E)	0.69
Item 22	-0.16 (C)	0.53(E)	0.14(E)	0.61*	0.75
Item 23	0.45(E)	-0.54 (C)	0.66*	0.77(E)	0.8
Item 24	0.92*	0.18(E)	0.14(E)	0.28(E)	0.69
Item 25	-0.39 (C)	0.61(E)	0.67*	0.26(E)	0.58
Item 26	-0.37 (C)	NA	0.66*	0.24(E)	0.32
Item 27	0.38(E)	0.75(E)	0.51(E)	0.75*	0.32
Item 28	0.38(E)	0.05(E)	0.76*	0.28(E)	0.49
Item 29	-0.48 (C)	-0.23 (C)	0.13(E)	0.73*	0.4
Item 30	-0.16 (C)	0.38 (E)	0.76*	0.47 (E)	0.53
Item 31	0.7*	0.19(E)	-0.19 (C)	0.41 (E)	0.43
Item 32	-0.07 (C)	0.81*	0.12(E)	0.17(E)	0.52
Item 33	0.47(E)	-0.26 (C)	0.26 (E)	0.7*	0.51

Legend: * (Answer), E (MA>0, effective), I (MA=0, ineffective) and C (MA<0, Confusing)

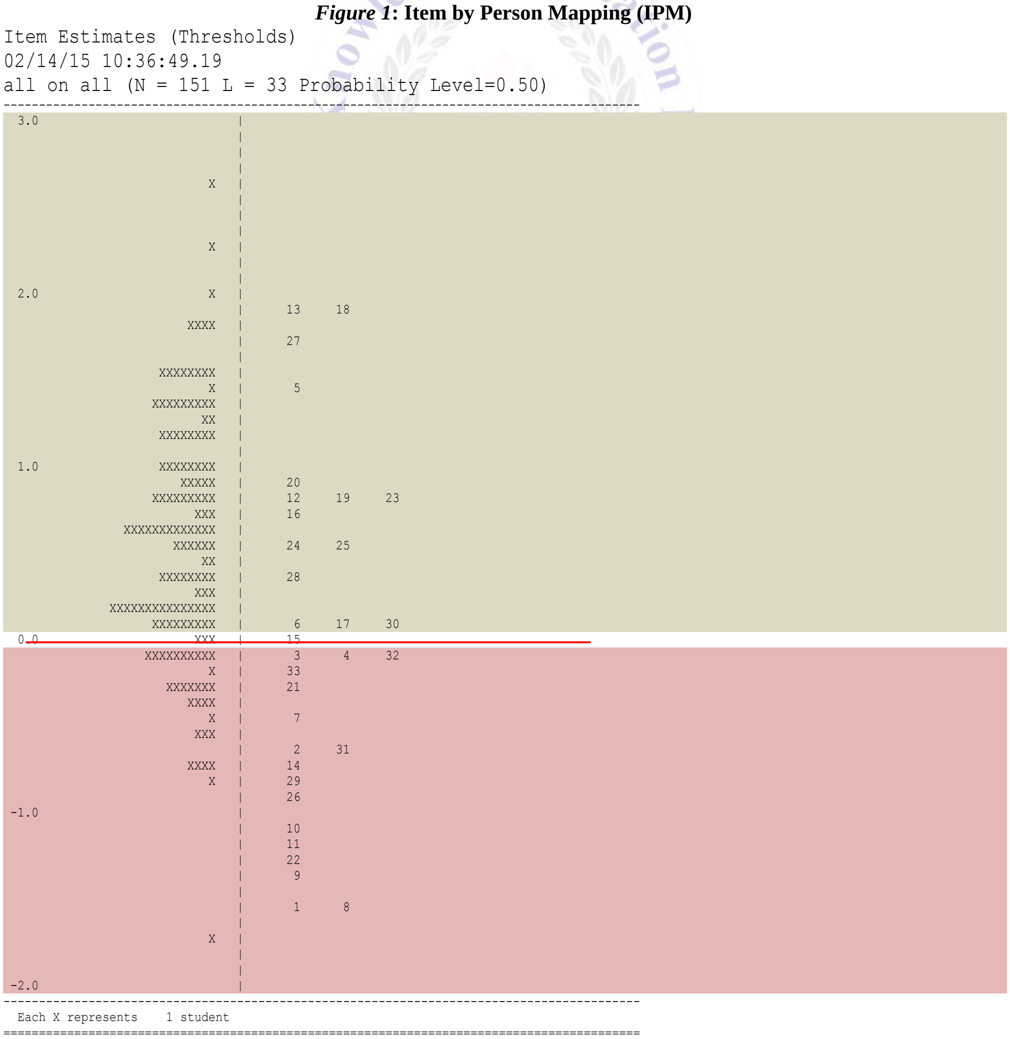
Table 6.Case Estimates of Basic and Integrated Science Process and Skill Test Using Rasch Model Analysis

ITEM	SCORE	MAXSCR	THRSH	ERROR	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
item 1	131	149	-1.57	0.26	1.02	1.3	0.2	1.1
item 2	111	150	-0.6	0.19	1.01	0.98	0.1	-0.1
item 3	96	149	-0.11	0.18	1.11	1.15	1.6	1.2
item 4	96	150	-0.09	0.18	0.9	0.88	-1.5	-1
item 5	44	148	1.45	0.19	1.05	1.06	0.7	0.5
item 6	91	151	0.08	0.17	0.92	0.92	-1.4	-0.7
item 7	107	150	-0.44	0.19	1.02	1.04	0.2	0.4
item 8	132	151	-1.52	0.25	0.94	0.9	-0.3	-0.3
item 9	121	141	-1.37	0.25	1.01	0.83	0.1	-0.6
item 10	122	148	-1.12	0.22	0.97	0.87	-0.2	-0.6
item 11	125	150	-1.16	0.23	1.03	1.03	0.2	0.2
item 12	65	148	0.79	0.18	0.97	0.96	-0.5	-0.3
item 13	30	144	1.95	0.21	1.01	1.08	0.1	0.5
item 14	114	150	-0.7	0.2	1.08	1.1	0.8	0.7
item 15	92	150	0.03	0.18	0.97	0.94	-0.5	-0.5
item 16	67	148	0.74	0.17	1.1	1.11	1.8	1
item 17	89	146	0.05	0.18	0.92	0.89	-1.3	-1
item 18	32	149	1.92	0.21	1.07	1.21	0.7	1.2
item 19	63	148	0.85	0.17	1.17	1.19	2.9	1.7
item 20	60	148	0.94	0.18	1.03	1.12	0.5	1.1
item 21	101	150	-0.25	0.18	0.95	0.94	-0.7	-0.4
item 22	126	149	-1.28	0.23	0.98	1.26	-0.1	1.1
item 23	65	149	0.8	0.17	1.13	1.13	2.2	1.2
item 24	73	150	0.58	0.17	0.85	0.82	-3	-1.8
item 25	73	146	0.53	0.18	1.1	1.12	1.9	1.1
item 26	116	146	-0.9	0.21	0.99	0.92	-0.1	-0.4
item 27	35	146	1.77	0.2	1.08	1.12	0.8	0.8
item 28	78	143	0.33	0.18	0.98	1.01	-0.3	0.1
item 29	113	144	-0.84	0.21	0.84	0.72	-1.4	-1.6
item 30	85	144	0.14	0.18	0.95	0.95	-0.8	-0.4
item 31	108	143	-0.67	0.2	0.94	0.88	-0.6	-0.7
item 32	92	143	-0.11	0.18	0.86	0.8	-2.2	-1.7
item 33	95	142	-0.23	0.19	0.98	0.98	-0.2	-0.1
Mean			**0**		**1**	**1.01**	**0**	**0**
SD			**0.98**		**0.08**	**0.14**	**1.3**	**0.9**

Table above presents the result of the scoring procedure. This was the initial result generated, the table must be regenerated after the deletion of several items for science which have poor discrimination indices or do not conform to the Rasch model. The number of an item answered correctly by an individual ranges from 32 - 132. The maximum score ranges from 142 to 151. Items not answered by some individuals were not counted in the maximum score. The third column provides information about the case estimate or the ability of an individual (estimate) and the standard error of the estimate (error). The negative threshold indicates that low performers chose the correct items while positive threshold indicates that high performers got the correct answer. This indicates that low performers got the items 1 to 4, items 7 to 11, items14, 21,22,26,29, and 31 to 33. While high performers got the correct answers in item 5,6,12,13,15-20,23-25, 27-28, and 30. The result also shows that the most difficult item was item 1 with the highest negative threshold while the easiest item was item 13 with the highest positive threshold.

The *infit* mean square is the fit statistic in the Rasch procedure and it is weighted to give greater weight to those responses near the steepest segment of the item characteristic curve. The acceptable range for the *infit* mean square, or in other words, the criterion to accept the items as conforming to the Rasch model is set by the QUEST program to lie within the range of 0.77 to 1.30, indicated by the dotted lines in the figure 2. Thus, item 1 to 33 is all considered in the analysis because it is within the range of 0.77 to 1.30.

Figure 1: **Item by Person Mapping (IPM)**



Parallel to Classical Theory analysis of item difficulty, Rasch analysis also produces the difficulty values of each of the Science Skills test items. Rasch model analysis provides how the items spread along the ability range of the group of students they are given to (see figure 1).

In *Figure 1*, observations on the left hand side are examinee proficiency values whereas those on the right hand side are item parameter values. This IPM can tell us the "big picture" of both items and students. The examinees on the upper left are said to be "better" or "smarter" than the items on the lower right, which mean that those easier items are not difficult enough to challenge those highly proficient students. On the other hand, the items on the upper right outsmart examinees on the lower left, which implies that these tough items are beyond their ability level. In this example, the examinees overall are "better" than the exam items. If we draw a red line at zero, we can see that examinees who are below average would miss a small chunk of items (the grey area) but pass a much larger chunk (the pink area). Furthermore, the figure shows that no *misfit* data (outliers) were shown.

Figure 2 maps the distribution of examinee ability and item difficulty on the same *logit* scale. The distribution of examinee ability did not obviously deviate from normality. In contrast, the distribution of item difficulty distinctly diverged from normality. No items with difficulties lower than the ability of least able respondents exists and no items with difficulties higher than the ability of most proficient respondent. Almost all of students had ability measures ranged between 0.77 to 1.30, there were only five items with extreme difficulty in this range.

***Figure 2*: Item Fit Map**

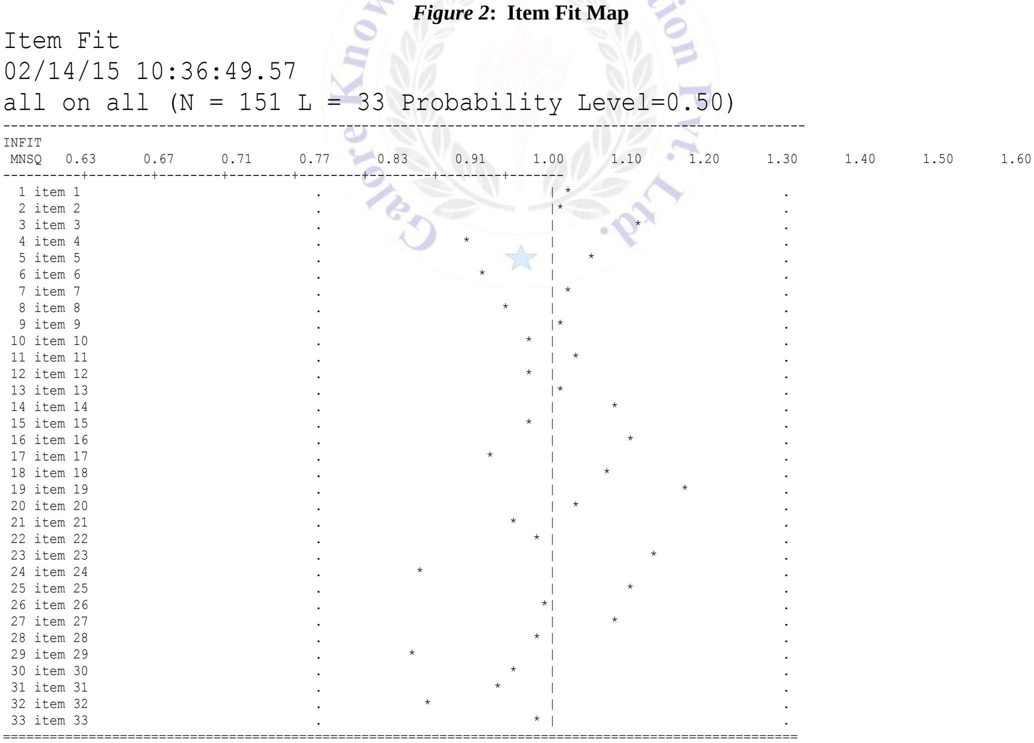


Figure 2 Shows the Weighted MSQ and ZSTD of each item in the Rasch Model Analysis. This figure illustrates the result of initial item fit *statistic* analysis. No item has a fit statistic value out of the range of pre-specified criteria for weighted ZSTD. Therefore, no item would be excluded from the second round of item fit *statistic* analysis. The processes would be continued until no misfit item was identified. After the exclusion of all misfit items, misfit examinee would also be excluded in a

similar manner (MSQ = mean square; ZSTD = standardized fit statistics). Thus, items 1 to 33 are all considered in the analysis because it is within the range of 0.77 to 1.30 as suggested by Yang (2010).

CONCLUSION

The test reliability has unsatisfactory value of 0.67. To improve the test reliability, increasing the item numbers should be considered. The examination was relatively easy for most of the respondents. To enhance the discrimination of the test, item difficulty should be adjusted to promote usefulness of the exam. However, in obtaining the value of internal consistency or the reliability coefficient *Cronbach's Alpha*, mean score and standard deviation, either Rasch Model Analysis or SPSS may be used.

The mean ability through Rasch model analysis per item showed that high performers in the test got the correct answers compared to the low performing group. This is the same and consistent with the results of the mean trend scores of the upper, middle and lower groups as generated by the SPSS. Choices in the BISPST were effective distracters. Majority of the items need revision but all the items were within the *infit* range of 0.77 - 1.30.

Acknowledgment

Sincere gratitude to the Science High School teachers in the Division of Neva Vizcaya and the Department of Education thru Dr. Samuel R. Soliven, Officer-in-Charge, Office of the Schools Division Superintendent of Nueva Vizcaya, for his invaluable research support which unquestionably led to the realization of the study.

REFERENCES

- Bilgin, F. (2006). *Science Skill Test* (3rd ed.). New York: Holt, Rinehart, & Winston.
- Harlen, R. (2000). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
- Ho Yu, C. (2013). A Simple Guide to the Item Response Theory (IRT) and Rasch Modeling.
- Kaplan, R., & Saccuzzo, D. (2004). *Psychological testing: Principles, applications, and issues* (6th ed.) Pacific Grove, CA: Brooks/Cole.
- Par, M. Anastasi, A., & Urbina, S. (2014). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Patten, M. (2001). *Questionnaire research: A practical guide* (2nd ed.). Los Angeles: Pyrczak.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.
- Yang, S. (2010). Statistical item analysis of the examination in anesthesiology for medical students using the Rasch model. Journal of the Chinese Medical Association 74 (2011) 125e129.

How to cite this article: S. de Gracia R, Gurat MG. Statistical item analysis of basic & integrated science process and skill test (BISPST) using rasch model and SPSS. Int J Res Rev. 2016; 3(11):89-97.

******