Evaluating Instrument Quality: Rasch Model – Analyses of Post Test of Curriculum 2013 Training

Komalasari
Lembaga Penjaminan Mutu Pendidikan, Kalimantan Tengah, Indonesia

Abstract: The main purpose of this study was to evaluate the quality of post test utilized by LPMP Central Kalimantan Indonesia in curriculum 2013 training for X grade teachers. It uses Rasch analysis to explore the item fit, the reliability (item and person), item difficulty, and the Wrigh map of post test. This study also applies Classical Test Teory (CTT) to determine item discrimination and distracters. Following a series of iterative Rasch analyses that adopted the "data should fit the model" approach, 30 items post test of curriculum 2013 training was analyzed using Acer Conquest 4 software, software based on Rasch measurement model. All items of post test of curriculum 2013 training are sufficient fit to the Rasch model. The difficulty levels (i.e. item measures) for the 30 items range from –1.746 logits to +1.861 logits. The item separation reliability is acceptable at 0.990 and person separation reliability is low at 0.485. The wright map indicates that the test is difficult for the teachers or the teachers have low ability in knowledge of curriculum 2013. The post test items cannot cover all the ranges of the teachers' ability levels. Items discrimination of post test of curriculum 2013 training grouped into fair discrimination (item 2, 4, 5, 8, 11, 18) and poor discrimination (1, 3, 6, 7, 9, 10,12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30). Some distracters from item 1, 2, 6, 7, 8, 9, 11, 13, 16, 17, 18, 19, 20, 22, 24, 25, 27, 28, 29 and 30 are problematic. These distracters require further investigation or revision.

Key words: Rasch analysis, training, curriculum 2013, post test

Introduction

A new national curriculum for primary and secondary school was implemented by Indonesian Government in 2013. The new curriculum is an improvement of the previous curriculum which is known as curriculum 2006. The new curriculum is named curriculum 2013. In addition, there are some reasons for the changing of the former curriculum such as internal challenges, external challenges, mindset completion, strengthening curriculum management and reinforcement of material.

The curriculum 2013 is being implemented gradually started from academic year 2013/2014 until for academic year 2018/2019. Furthermore, for academic year 2017/2018, the implementation of curriculum 2013 is extended to the whole of Indonesia's regency/city in total number of the school is 4510 of senior high schools or about 60%, including Central Kalimantan Province. Moreover, in implementing the curriculum 2013, the teachers are trained an activity that is called "curriculum 2013 training". In every province of Indonesia, the training session of curriculum 2013 is conducted by LPMP (Educational Quality Assurance Institution). The funds for curriculum 2013 training is come from the budget of state expenditure and revenues (APBN).

The curriculum 2013 training is done in order to improve the competence of teachers and for the preparation of Curriculum 2013 implementation exactly.

The objective of this training to improve the participant's competence in: a) understanding the dynamics and policies of curriculum development, the policy of strengthening character education and the application of literacy in learning; b) analyzing the learning objectives, material, learning, and assessment covered: Document: learning objectives and guidelines subjects, the material in textbooks, Implementation learning model, assessment of learning; c) designing the learning implementation plan with 21st century skill (critical thinking, creativity, communication, and collaboration); d) doing a Higher Order Thinking Skill (HOTS) learning practice and assessment and doing a review of the practice result; and d) practicing the process and reporting the assessment of learning and the introduction to e-raport aplication (The Guidence of Curriculum 2013 Training in 2017, Education and Cultural Ministry of Indonesia).

Educational Quality Assurance Institution of Central Kalimantan conducted the curriculum 2013 training for 800 senior high school teachers on 25-30 May 2017. It was conducted for six days or 39 hours at 20 Schools as the place of training activity. The participants of the training(the trainee) came from14 regions/ cities and 82 schools around Central Kalimantan Province. The trainee were senior high school teachers of 18 subjects and counseling teacher.
The post test was held on the end of the training to assess the trainee's kowledge competency of curriculum 2013. The post test of curriculum 2013 also measures the extent of training effectiveness. It was consist of 30 items in the form of multiple choice question and prepared by Direktorat Pembinaan SMA Direktorat Jenderal Pendidikan Dasar dan Menengah Education and Cultural Ministry of Indonesia.

Assessment should be precise, technically sound, producing accurate information for decision making in all circumstances (Dubois and Rothwell 2000; Stetz and Chmielewski 2015, in Nornazira S, 2015). It is very important to check the credibility of the instrument utilized in curriculum 2013 training at LPMP Central Kalimantan. Thus, this study concerns on analyzing post test items of curriculum 2013 training to evaluate the quality of curriculum 2013 training post test. It is beneficial to ensure that the assessment of teacher in curriculum 2013 training using post test instrument could give accurate information of teacher competency.

The analysis will be done based on items respond theory approach by Rasch's model. The Rasch measurement model was opted for this study because of its sophisticated approach to evaluate patterns of items responses and scale, and item performance (Linacre 2002; Bond and Fox, 2015 in Nornazira S, 2015). Analysis using Rasch measurement model is a more sophisticated approach to evaluate patterns of items responses and scale, and item performance (Chen et al. 2014 in Nornazira S, 2015). This study also uses Classical Test Teory (CTT) to analyse item discrimination and distracters.

Methods

In order to maintain the accuracy of the curriculum 2013 training post test instrument, it is very important to analyzed it. The study emphasized on four aspects of Rasch Analysis diagnoses which are (i) item and person reliability, (ii) item fit, (iii) item difficulty, and (iv) Wright Map, This study also used CTT for two aspects: (i) item discrimination, and (ii) distracter

The post test instrument of curriculum 2013 training was tested on 711 teachers that follow the curriculum 2013 training in 2017. Data for post test item analysis of curriculum 2013 training were secondary from Educational Quality Assurance Institution Central Kalimantan Indonesia.

This stduy was population study. The summarized of data obtained can be seen in this following table.

Table 3. Data for Item and Test Analysis

Training Names	Participant of Training	Subject	Total Items (Multiple Choice)	Number of Examinees
Curriculum 2013	Teachers (X grade)	Curriculum 2013	30	711

The post test of curriculum 2013 training was analyzed using Acer Conquest 4 software, software based on Rasch measurement model. The analysis covered:

Item Fit

Bond anf Fox (2015) describe the concept of fit is a quality-control mechanism'(akin to the use of fit in industrial statistics). Fit statistic provide one indication as to whether the researcher has completed a task of sufficient quality to allow that values for person and items can be represented with interval-level measures. Fit indices help the investigator to ascertain whether the Rasch requirement for unidimensionality holds up empirically. Fit statistics help to determine whether the item estimations may held as meaningful quantitative summaries of the observation.

Item fit statistics (i.e., infit/outfit and ZSTD) show which items fit the estimated model (Neumanna, Neumannb and Nehmc, 2011). Infit and outfit statistics adopt slightly different techniques for assesing an item's fit to the Rasch Model. Table below describe Reasonable range for item MNSQ infit and outfit as suggested by Wright and Linacre (1994).

Table 4. Reasonable range for item MNSQ infit and outfit as suggested by Wright and Linacre (1994).

Reasonable Item Mean-square Ranges for INFIT and OUTFIT	
Types of test	Range
MCQ (Hight Stakes	0.8 – 1.2
MCQ (Run of the mill)	0.7 – 1.3
Rating scale (survey)	0.6 – 1.4
Clinical observation	0.5 – 1.7
Judged (agreement encoraged)	0.2 -1.2

Reliability

There are two kinds of reliability in Rasch Model. First the person reliability index and second the item reliability index. The person reliability index indicates the replicability of person ordering we could expect if this sample of person were given another a paralel set of items measuring the same construct (Wright and Masters, 1982 in Bond and Fox, 2015). The item reliability indicates the replicability of item placements along the pathway if these same item were given to another same-sized sample of persons who behaved in the same way (Bond and Fox, 2015).

Wright Map

Rasch Analysis software produce some form of the map is often called a Wright map. Wright map report the relations between only two key aspect of the variable: the item difficulty estimates and person ability estimates. One delightful aspect of this Rasch representation of data analysis is that many of the person and item relations are shown in meaningful pictorial, or 'map' form (Bond and Fox, 2015). The distribution of persons (on the left) and items of instrument (on the right) are displayed on the same so-called logit scale.

Item Discrimination Index

A good question is a question that can be distinguishing the groups of students who have a high ability and low ability. The index which can be measured that difference is item discrimination. Also, to determine the item discrimination also can be used discrimination index, biserial correlation index, point biserial correlation index, and alignment index. Thus, the item discrimination of question is exactly same with the question validity.

According to Rahmah Zulaiha (2012) the question item discrimination the proportion difference of right answer on the students group with the high ability (high group) and students group with the low ability (low group). The question item discrimination is between -1 till +1. The negative sign means the group of students with low ability who answer correctly certain items of question more than the student's group with high ability. The item-discrimination index for item i, di, is calculated by the formula (Allen & Yen, 1979):

$$d_i = \frac{U_i}{N_{iu}} - \frac{L_B}{N_{iL}}$$

Where:

d_i = Item Discrimination index

Ui = the number of examinees who have total test scores in the upper range of total test scores and who also have item i correct,

L$_i$ = the number of examinees who have total test scores in the lower range of total test scores and who also have item i correct,

n_{iu} = the number of examinees who have total test scores in the upper range of total test scores, and

n_{iL} = the number of examinees who have total test scores in the lower range of total test scores.

If $n_{iu} = n_{iL} = n_i$, this formula reduces to

Distractor Effectiveness

According to Saifuddin Azwar (1996) the item difficulty and item discrimination are not sufficient to assess the item test quality, how are the examinees' information answers distributed on the applicable optional answer that needs to note in order the function of a test item can be fulfiled maximumly.

Distractors effectiveness from the test item are analyzed from the answer distribution toward the test item related to any provided alternative. Distractor effectiveness is reviewed to know whether all of the distractor or optional answer which is not as the key answer has functioned as it is. It means those distractors have chosen by more or all examinees from the low group, while the examinees from the high group only a few or none choose it.

A distractor can be said it is functioned well if the distractor has chosen at least 2.5% (>0.025). (Rahmah Zulaiha, 2012). The distribution of optional answer is gained through calculation by using this following formula:

$$P_{PJ} = \frac{J_{PJ}}{n}$$

Where:

P_{PJ} = the answer distribution for certain optional answer

J_{PJ} = amount of the students who chose the certain optional answer

n = total of the students

Findings and Discussions

1.   Fit Statistics

The Rasch model was conducted to analyse the quality of the items and test. Analyses were conducted to evaluate the goodness of fit of the data to the Rasch model for post test items of curriculum 2013 training. The first analysis (N=711) indicated that 46 teachers (examinees) were identified to be the misfit person in measuring the kowledge competency of teachers in curriculum 2013 training. They are person ID63, ID100, ID124, ID140, ID148, ID154, ID165, ID166, ID167, ID272, ID275, ID300, ID 302, ID 312, ID315, ID324, ID325, ID331, ID350, ID361, ID410, ID416, ID424, ID430, ID434, ID468, ID503, ID544, ID564, ID568, ID571, ID573, ID576, ID577, ID621, ID634, ID635, ID636, ID638, ID639, ID642, ID643, ID646, ID647, ID686, and ID689. The analysis shows that these people do not meet the requirement of Rasch model in analyzing the fit characteristics. Thus, suggested these people supposed to be removed from the analysis.

Further,the second analysis was conducted without misfit person, with number of case are 665. The indicator properties of the test are presented in Table 2. The evaluation of goodness of fit to the Rasch model for post test items of curriculum 2013 was based on the weighted and unweighted mean square (i.e., Infit and Outfit mean square) statistics and the t-values for each item. As a working guide, Infit and Outfit values range from 0.70 to 1.30 are considered fit while t-values between -2.00 and +2.00 are considered acceptable, with 95% confidence interval level of significance.

Table 5. *Item Properties of Post Test of Curriculum 2013 Training for Teachers (n=711)*

Item Code	ESTIMATE	SE	Unweighted Fit		Weighted Fit	
			MNSQ	T	MNSQ	T
1	0.141	0.083	1.02	0.3	1.01	0.4
2	-1.04	0.08	0.97	-0.5	0.98	-1
3	0.668	0.091	1.02	0.3	1	0.1
4	-0.432	0.078	0.97	-0.6	0.97	-2
5	-0.594	0.078	0.97	-0.5	0.98	-1.6
6	1.861	0.134	1.04	0.7	1	0
7	0.107	0.082	1	0.1	1	0
8	0.09	0.082	0.98	-0.3	0.98	-0.7
9	0.559	0.089	1.05	0.8	1.02	0.5
10	-0.529	0.078	1	0	1	-0.1
11	0.229	0.083	0.98	-0.3	0.98	-0.6
12	-0.249	0.079	1.02	0.4	1.02	1.1
13	0.788	0.094	1	0	0.99	-0.1
14	0.299	0.084	0.98	-0.3	0.99	-0.4
15	0.198	0.083	0.98	-0.4	0.98	-0.6

Item Code	ESTIMATE	SE	Unweighted Fit		Weighted Fit	
			MNSQ	T	MNSQ	T
16	-0.266	0.079	1.02	0.3	1.01	0.7
17	0.838	0.095	1.01	0.2	1	0
18	-1.5	0.086	0.96	-0.7	0.98	-0.6
19	0.856	0.095	0.99	-0.2	0.99	-0.2
20	0.385	0.086	1.04	0.8	1.02	0.6
21	-0.467	0.078	0.99	-0.3	0.99	-0.9
22	1.373	0.112	1.04	0.7	1.01	0.1
23	-1.746	0.091	0.96	-0.7	0.98	-0.4
24	0.893	0.096	1.04	0.8	1.02	0.4
25	-1.445	0.086	1.02	0.3	1.01	0.2
26	-1.007	0.08	0.99	-0.2	0.99	-0.6
27	0.712	0.092	1.03	0.6	1.02	0.4
28	-0.242	0.079	1.02	0.4	1.02	1
29	-1.4	0.085	1.05	0.9	1.02	0.7
30	0.919*	0.097	1.11	2	1.05	0.9
Person Reliability	0.485					
Item Reliability	0.990					

Table 3 presents that the difficulty levels (i.e. item measures) for the 30 items range from –1.746 logits to +1.861 logits, associated with standard errors of 0.079 or 0.134 logits. The unweighted and weighted fit MNSQs range from 0.96 to 1.11 with t-values within -2.0 to +2.0 indicating sufficient fit to the Rasch model. Appendix 1 shows the 30 plots that were produced by the plot icc command. The ICC plot shows a comparison of the empirical item characteristic curve (the broken line, which is based directly upon the observed data) with the modelled item characteristic curve (the smooth line).

The item separation reliability is 0.990 and is acceptable. The person separation reliability is low at 0.485. The low person reliability indicates that the individuals who took this test are likely to get different estimated ability scores if a parallel test is given to them.

2. Wright Map

Figure 1 displays a Wright Map which visually summarizes several aspects of the Rasch analysis. Figure 1 presents the map of the latent distributions and response model parameter estimates for the 30 items. Teachers are placed on the left side of the scale according to knowledge of curriculum 2013 ability, and the item difficulty indicators are shown on the right side. The teachers with the highest ability levels and the items with highest difficulty levels are located at the top of the map, while the teachers with the lowest ability level and the easiest items are located at the bottom.

The distribution of persons (on the left) and items of the instrument (on the right) are displayed on the same so-called logit scale. Persons at the same position (or 'height') on the scale as a particular item have a 50% chance of answering the item correctly. Questions of equivalent difficulty lie at the same point on the logit scale (e.g., Questions 4, and 21; 7 and 8;). Individuals ('persons') located above an item, however, have an even greater chance of answering the item correctly (i.e., the itemis likely to be easier for such individuals). Those persons located below an item have a lower probability of being able to answer it correctly (i.e., the item is more difficult for them).

The mean of the measure is shown as 0 on the map. It can be seen that the distribution of the teachers's ability tends to be lower than the mean. Only a small portion of the teachers are able to answer the difficult items (from item 7 to item 13) correctly. Items 24, 30, 17, 19, 24, 30, 6 and 22 are considered too difficult for the teachers and none of the teachers are able to get the correct answers. There is also a noticeable a teacher who have low ability and are not able to get any of the items correctly. Overall, the test is difficult for the teachers or the teachers have low ability in knowledge of curriculum 2013. The post test items cannot cover all the ranges of the teachers' ability levels.

Based on the facts obove, the curriculum 2013 training for X grade teachers that implemented by LPMP Central Kalimantan in 2017 could not improve the ability of teachers succesfully. It could be caused by the the lack of the training facilitators and the lack of teacher competency. It is very important to strengtheen mentoring activities of curriculum 2013 in every school.
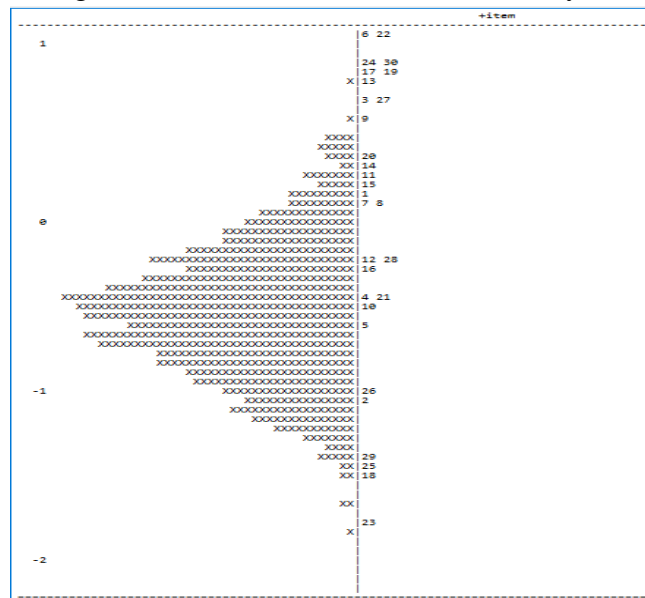
Note: Each 'X' represents 0.4 students.

Figure 1. Map of latent distributions and response model parameter estimates.

3. Item Discrimination and Distractor of Post Test of Curriculum 2013 Training

Item discrimination are indicated by the Point Biserial Index (PBI). Based on McGahee & Ball (2009), Point Biserial Index (PBI) is defined as correlation between score on an item and score on the exam; differentiates between those who have high or low test scores; Range from -1 to +1; Positive PBI indicates those who scored well on exam answered item correctly; PBI should be positive for correct answer and PBI should be negative for distractors. As a working guide, this study uses PBI's General rules: Below 0.2: Poor; revise item; 0.2-0.29: Fair; 0.3-0.39: Good; 0.4-0.7 Very good (Penn, 2009; McGahee & Ball, 2009)

The results of analysis of item discriminations and distractors using Acer Conquest 4 for post test items of curriculum 2013 training summaries in the following table. For each item, there is an item discrimination, frequency or percentage for correct answer and distractors and interpretation. They will use as information to determine the quality of the item.

Table 6. Item1

Response	Frequency	Percentage	Point Biserial
A	89	13.55	-0.07
B*	225	34.25	0.1
C	105	15.98	-0.18
D	238	36.23	0.09

The item discrimination represented by Point Biserial value (0.1) but very low (below 0.2) indicating poor discrimination. The positive PBI (0.09) on distracter D indicates that teachers who performed well on the post test selected it. This is problematic.

Table 7. Item 2

Response	Frequency	Percentage	Point Biserial
A	68	10.24	-0.08
B*	411	61.9	0.2
C	113	17.02	-0.14
D	72	10.84	-0.07

The significant Point Biserial value (0.24) for the correct answer indicating fair discrimination for this item. Options A, C and D are acceptable as they have negative Point Biserial values.

Table 8. Item 3

Response	Frequency	Percentage	Point Biserial
A	231	35.05	0.03
B*	157	23.82	0.1
C	135	20.49	-0.11
D	136	20.64	-0.04

The item discrimination represented by Point Biserial value (0.1) but very low (below 0.2) indicating poor discrimination. The positive PBI (0.03) on distracter A indicates that teachers who performed well on the post test selected it. This is problematic. This item should be examined and revised.

Table 9. Item 4

Response	Frequency	Percentage	Point Biserial
A	54	8.14	-0.13
B	50	7.54	-0.09
C*	315	47.51	0.24
D	244	36.8	-0.12

The significant Point Biserial value (0.24) for the correct answer indicating fair discrimination for this item. Options A, B and D are acceptable as they have negative Point Biserial values.

Table 10. Item 5

Response	Frequency	Percentage	Point Biserial
A	171	25.83	-0.13
B	39	5.89	-0.14
C*	340	51.36	0.23
D	112	16.92	-0.07

The significant Point Biserial value (0.23) for the correct answer indicating fair discrimination for this item. Options A, B and D are acceptable as they have negative Point Biserial values.

Table 11. Item 6

Response	Frequency	Percentage	Point Biserial
A*	59	8.89	0.11
B	222	33.43	0.14
C	319	48.04	-0.14
D	64	9.64	-0.08

The item discrimination represented by Point Biserial value (0.11) but very low (below 0.2) indicating poor discrimination. The positive PBI (0.14) on distracter B indicates that teachers who performed well on the post test selected it. This is problematic. This item should be examined and revised.

Table 12. Item 7

Response	Frequency	Percentage	Point Biserial
A	98	14.78	-0.09
B*	232	34.99	0.15
C	218	32.88	0.07
D	115	17.35	-0.18

The item discrimination represented by Point Biserial value (0.15) but very low (below 0.2) indicating poor discrimination.The positive PBI (0.07) on

distracter C indicates that teachers who performed well on the post test selected it. This is problematic. This item should be examined and revised.

Table 13. Item 8

Response	Frequency	Percentage	Point Biserial
A	116	17.52	-0.03
B	66	9.97	-0.1
C*	234	35.35	0.21
D	246	37.16	-0.12

The significant Point Biserial value (0.21) for the correct answer indicating fair discrimination for this item. Distractor D is problematic since the majority of the teachers choose the incorrect answer (37.16%) although it has significant negative Point Biserial value. Options A and B are acceptable as they have negative Point Biserial values.

Table 14. Item 9

Response	Frequency	Percentage	Point Biserial
A	144	21.75	-0.08
B	70	10.57	-0.08
C*	171	25.83	0.05
D	277	41.84	0.08

The item discrimination represented by Point Biserial value (0.15) but very low (below 0.2) indicating poor discrimination. The positive PBI (0.08) on distracter B indicates that teachers who performed well on the post test selected it. Distractor D is problematic. This item should be examined and revised.

Table 15. Item 10

Response	Frequency	Percentage	Point Biserial
A	117	17.67	-0.07
B*	330	49.85	0.12
C	40	6.04	-0.08
D	175	26.44	-0.03

The significant Point Biserial value (0.12) for the correct answer indicating poor discrimination. Distractors A, C, and D are considered good distractors since the PBI values are negative.

Table 16. Item 11

Response	Frequency	Percentage	Point Biserial
A*	215	32.38	0.2
B	122	18.37	-0.1
C	41	6.17	-0.1
D	286	43.07	-0.07

The significant Point Biserial value (0.2) for the correct answer indicating fair discrimination for this item. Distractors B and C are considered good

distractors since the PBI values are negative. Distractor D is problematic since the majority of the students choose the incorrect answer (43.07%) although it has negative Point Biserial value.

Table 17. Item 12

Response	Frequency	Percentage	Point Biserial
A	269	40.57	-0.01
B	77	11.61	-0.05
C*	286	43.14	0.08
D	31	4.68	-0.09

The Point Biserial value (0.08) for the correct answer indicating poor discrimination for this item. Distractors A, B and D are considered good distractors since the PBI values are negative.

Table 18. Item 13

Response	Frequency	Percentage	Point Biserial
A	290	43.87	0.05
B*	144	21.79	0.14
C	167	25.26	-0.13
D	60	9.08	-0.1

The Point Biserial value (0.14) for the correct answer indicating poor discrimination for this item. Distractors C and D are considered good distractors since the PBI values are negative. Distractor A is problematic since the majority of the students choose the incorrect answer (43.87%) and it has positive Point Biserial value.

Table 19. Item 14

Response	Frequency	Percentage	Point Biserial
A	100	15.08	-0.05
B	158	23.83	-0.08
C	200	30.17	-0.07
D*	205	30.92	0.18

The Point Biserial value (0.18) for the correct answer indicating poor discrimination for this item. Distractors A, B, and C are considered good distractors since the PBI values are negative.

Table 20. Item 15

Response	Frequency	Percentage	Point Biserial
A	168	25.45	-0.13
B*	218	33.03	0.17
C	166	25.15	-0.07
D	108	16.36	0.01

The Point Biserial value (0.17) for the correct answer indicating poor discrimination for this item. Distractors A, C, and D are considered good distractors since the PBI values are negative.

Table 21. Item 16

Response	Frequency	Percentage	Point Biserial
A	140	21.08	0.03
B*	289	43.52	0.11
C	144	21.69	-0.07
D	91	13.7	-0.11

The Point Biserial value (0.11) for the correct answer indicating poor discrimination for this item. Distractors C and D are considered good distractors since the PBI values are negative. The positive PBI (0.03) on distracter A indicates that teachers who performed well on the post test selected it. Distractor A is problematic.

Table 22. Item 17

Response	Frequency	Percentage	Point Biserial
A	112	16.89	0.05
B	251	37.86	-0.1
C	161	24.28	-0.05
D*	139	20.97	0.13

The Point Biserial value (0.13) for the correct answer indicating poor discrimination for this item. Distractors B and D are considered good distractors since the PBI values are negative. The positive PBI (0.05) on distracter A indicates that teachers who performed well on the post test selected it. Distractor A is problematic.

Table 23. Item 18

Response	Frequency	Percentage	Point Biserial
A*	476	71.69	0.24
B	35	5.27	-0.18
C	146	21.99	-0.17
D	7	1.05	0.01

The significant Point Biserial value (0.24) for the correct answer indicating fair discrimination for this item. Distractors B and C are considered good distractors since the PBI values are negative. The positive PBI (0.01) on distracter D indicates that teachers who performed well on the post test selected it. Distractor D is problematic.

Table 24. *Item 19*

Response	Frequency	Percentage	Point Biserial
A	103	15.56	-0.14
B	91	13.75	-0.13
C	331	50	0.06
D*	137	20.69	0.17

The Point Biserial value (0.17) for the correct answer indicating poor discrimination for this item. Distractors A and B are considered good distractors since the PBI values are negative. The positive PBI (0.06) on distracter C indicates that teachers who performed well on the post test selected it. Distractor C is problematic.

Table 25. Item 20

Response	Frequency	Percentage	Point Biserial
A	225	33.99	-0.12
B*	193	29.15	0.02
C	155	23.41	-0.01
D	89	13.44	0.14

The Point Biserial value (0.02) for the correct answer indicating poor discrimination for this item. Distractors A and C are considered good distractors since the PBI values are negative. The positive PBI (0.14) on distracter D indicates that teachers who performed well on the post test selected it. Distractor D is problematic.

Table 26. Item 21

Response	Frequency	Percentage	Point Biserial
A	114	17.22	-0.12
B*	320	48.34	0.19
C	69	10.42	-0.04
D	159	24.02	-0.08

The Point Biserial value (0.19) for the correct answer indicating poor discrimination for this item. Distractors A, D and D are considered good distractors since the PBI values are negative.

Table 27. Item 22

Response	Frequency	Percentage	Point Biserial
A	102	15.41	-0.15
B	344	51.96	0.09
C*	90	13.6	0.06
D	126	19.03	-0.03

The Point Biserial value (0.06) for the correct answer indicating poor discrimination for this item. Distractors A and D are considered good distractors since the PBI values are negative. The positive PBI (0.06) on distracter B indicates that teachers who performed well on the post test selected it. Distractor D is problematic.

Table 28. Item 23

Response	Frequency	Percentage	Point Biserial
A	36	5.44	-0.09
B	45	6.8	-0.08
C*	505	76.28	0.19
D	76	11.48	-0.13

The Point Biserial value (0.19) for the correct answer indicating poor discrimination for this item. Distractors A, B and D are considered good distractors since the PBI values are negative.

Table 29. *Item 24*

Response	Frequency	Percentage	Point Biserial
A	225	33.99	0.05
B	202	30.51	-0.01
C*	133	20.09	0.03
D	102	15.41	-0.09

The Point Biserial value (0.03) for the correct answer indicating poor discrimination for this item. Distractors B and D are considered good distractors since the PBI values are negative. The positive PBI (0.05) on distracter A indicates that teachers who performed well on the post test selected it. Distractor D is problematic.

Table 30. Item 25

Response	Frequency	Percentage	Point Biserial
A	48	7.27	0.03
B	81	12.27	-0.05
C*	466	70.61	0.06
D	65	9.85	-0.06

The Point Biserial value (0.06) for the correct answer indicating poor discrimination for this item. Distractors B and D are considered good distractors since the PBI values are negative. The positive PBI (0.03) on distracter A indicates that teachers who performed well on the post test selected it. Distractor D is problematic.

Table 31. Item 26

Response	Frequency	Percentage	Point Biserial
A	125	18.83	-0.07
B	56	8.43	-0.1
C	77	11.6	-0.11
D*	406	61.14	0.18

The Point Biserial value (0.18) for the correct answer indicating poor discrimination for this item. Distractors A, B and C are considered good distractors since the PBI values are negative.

Table 32. Item 27

Response	Frequency	Percentage	Point Biserial
A	127	19.16	-0.18
B	111	16.74	-0.1
C*	153	23.08	0.06
D	272	41.03	0.18

The Point Biserial value (0.06) for the correct answer indicating poor discrimination for this item. Distractors A and B are considered good distractors since the PBI values are negative. The positive PBI (0.18) on distracter D indicates that teachers who performed well on the post test selected it. Distractor D is problematic.

Table 33. *Item 28*

Response	Frequency	Percentage	Point Biserial
A	175	26.48	-0.02
B	113	17.1	0.04
C	89	13.46	-0.13
D*	284	42.97	0.08

The Point Biserial value (0.08) for the correct answer indicating poor discrimination for this item. Distractors A and C are considered good distractors since the PBI values are negative. The positive PBI (0.04) on distracter B indicates that teachers who performed well on the post test selected it. Distractor B is problematic.

Table 34. Item 29

Response	Frequency	Percentage	Point Biserial
A	40	6.03	0.01
B	67	10.11	0.04
C	94	14.18	-0.08
D*	462	69.68	0.02

The Point Biserial value (0.02) for the correct answer indicating poor discrimination for this item. Distractors C are considered good distractor since the PBI values are negative. The positive PBI (0.01 and 0.04) on distracter A and B indicates that teachers who performed well on the post test selected it. Distractor A and B is problematic.

Table 35. Item 30

Response	Frequency	Percentage	Point Biserial
A	369	55.91	0.12
B*	130	19.7	-0.09
C	126	19.09	-0.02
D	35	5.3	-0.07

The Point Biserial value (-0.09) for the correct answer is problematic as poor teachers also selected this answer. This item did not discriminate well. Distractors C and D are considered good distractors since the Point Biserial values are negative. Distractor A is problematic since the majority of the teachers (55.91%) choose the incorrect answer and showing non-significant Point Biserial value of 0.12. Options A and B require further investigation or revision.

Based on table 6 – table 35, items of post test of curriculum 2013 training could be grouped in two categories: fair discrimination and poor discrimination. Table 33 shows the items for every category.

Table 36. Category Of Items Based on Item Discrimination (PBI's Values)

Category	Items
Fair Discrimination	2, 4, 5, 8, 11, 18
Poor Discrimination	1, 3, 6, 7, 9, 10,12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30

Table 37. Problematic Distracters

Item	Distracter	Explanation
1	D	The positive PBI (0.09) on distracter, the majority of the teachers choose this distractor (36.23%)
3	A	The positive PBI (0.03) on distracter A, the majority of the teachers choose this distractor (35.05%).
6	B	The positive PBI (0.14) on distracter B.
7	C	The positive PBI (0.07) on distracter C
8	D	The majority of the teachers choose this distracter (37.16%)
9	B	The positive PBI (0.08) on distracter B.
11	D	Distractor D is problematic since the majority of the students choose the incorrect answer (43.07%)
13	A	Distractor A is problematic since the majority of the students choose the incorrect answer (43.87%) and it has positive Point Biserial value.
16	A	The positive PBI (0.03) on distracter A
17	A	The positive PBI (0.05) on distracter A
18	D	The positive PBI (0.01) on distracter D
19	C	The positive PBI (0.06) on distracter C
20	D	The positive PBI (0.14) on distracter D
22	B	The positive PBI (0.06) on distracter B
24	A	The positive PBI (0.05) on distracter A
25	A	The positive PBI (0.03) on distracter A
27	D	The positive PBI (0.18) on distracter D
28	B	The positive PBI (0.04) on distracter B
29	A and B	The positive PBI (0.01 and 0.04) on distracter A and B
30	A	The positive PBI (0.12) on distracter A and majority of the teachers (55.91%) choose the incorrect answer

Furthermore, the post test do not have good discrimination. Even item 30 is the most problematic item. It describe by MCC plot on figure 2 below. In particular it shows the proportion of students in each of a sequence of ten ability groupings that responded with each of the possible responses.
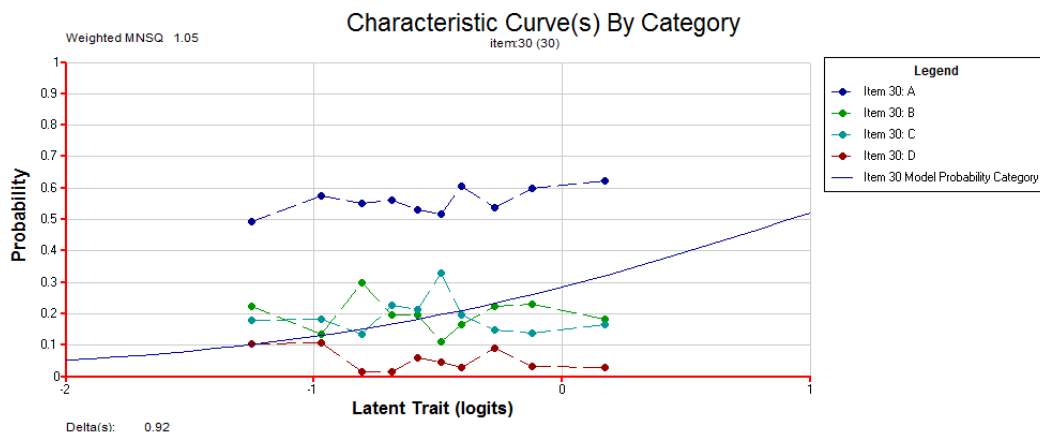
Figure 3. MCC of Item 30

Figure 3 indicates that distracter A is problematic since the majority of the teachers (55.91%) choose the incorrect answer and showing non-significant Point Biserial value of 0.12. This item did not discriminate well. The negative PBI for the correct answer B indicates that students who performed poorly on this exam answered correctly; the positive PBI (0.12) on distracter A indicates that students who performed well on the exam selected it.This item may have wrong key answer or the item is wrong. Some distracters are problematic (see table 37). These distracters require further investigation or revision.

Conclusion

As discussed earlier, Rasch measurement analysis was initially, conducted with 30 items of post test of curriculum 2013 training. The study emphasized on four aspects of Rasch Analysis diagnoses which are (i) item and person reliability, (ii) item fit, (iii) item difficulty, (iv) item discrimination, and two aspect based on CTT (i) distracter, and (ii) Wright Map.

The item separation reliability is 0.990 and is acceptable. The person separation reliability is low at 0.485. All items of post test of curriculum 2013 training are sufficient fit to the Rasch model indicated by the unweighted and weighted fit MNSQs range from 0.96 to 1.11 with t-values within -2.0 to +2.0.

The difficulty levels (i.e. item measures) for the 30 items range from –1.746 logits to +1.861 logits, associated with standard errors of 0.079 or 0.134 logits.

The wright map indicates that the distribution of the teachers's ability tends to be lower than the mean. The test is difficult for the teachers or the teachers have low ability in knowledge of curriculum 2013. The post test items cannot cover all the ranges of the teachers' ability levels.

Items discrimination of post test of curriculum 2013 training could be grouped into fair discrimination (item 2, 4, 5, 8, 11, 18) and poor discrimination (1, 3, 6, 7, 9, 10,12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30).

Some distracters from item 1, 2, 6, 7, 8, 9, 11, 13, 16, 17, 18, 19, 20, 22, 24, 25, 27, 28, 29 and 30 are problematic. These distracters require further investigation or revision. Based on the results of the analysis that has been obtained, can be implied as follows.

Post-test analysis of the 2013 curriculum training by using Rasch model is very important to know and do to acknowledge the characteristics information of test items. Based on the results of the analysis could be done improvements and development of post test instruments which is used on curriculum 2013 training which will come.

In addition, the results of the analysis show that the test is difficult for the teachers or the teachers have low ability in knowledge of curriculum 2013. The Educational Quality Assurance Central Kalimantan needs to reinforce teachers especially on difficult materials (item 6, 22,24,30,17,19,3,27) through the 2013 curriculum assistance activities in schools.

The findings of this study could provide a better knowledge basis for interpreting teachers assessment results in another training.

References

Allen, MJ., & Yen, W.M. (1979). *Introduction to measurement theory.* Monterey California: Brooks/Cole Publishing Company.

Bond, TG., & Fox, CM. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences*. New York : Routledge

McGahee, T.W. & Ball, J. (2009). How to read and really use an item analysis. Nurse Educator, 34, 166-171.

Neumanna, I., Neumannb,K., and Nehmc,R. (2011) Evaluating Instrument Quality in Science Education: Rasch-based analyses of a Nature of Science test. *International Journal of Science Education* Vol. 33, No. 10, 1 July 2011, pp. 1373–1405

Nornazira S., Aede H M., Nor Fadila M A and Adibah AL. *Reliability and Validity Evidence of Instrument Measuring Competencies for Superior Work Performance*. Rasch Report. Measurement Symposium (PROMS) 2015 Conference Proceedings, DOI 10.1007/978-981-10-1687-5_21.

Penn, B.K. (2009, August). Test item development and analysis. Presented at Creighton University School of Nursing Faculty Retreat, Omaha, NE.

Rahma Zulaiha. (2012). *Writtent test*. Jakarta:Puspendik.

Saifuddin Azwar. (1992). Reliability and validity.(3rd ed). Yogyakarta: Pustaka Pelajar.

The Technical Guidance and Assistance Guidance of Curriculum Implementation 2013 Senior High School 2017. (2017). Ministry of Education and Culture of RI