Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 3 No. 1 2023, 19 - 27
DOI: *https://doi.org/10.22487/27765660.2023.v3.i1.16412*
eISSN: 2776-5660

APPLICATION OF THE RASCH MODEL TO TEST TOOLS IN THE ANALYSIS OF SURVEY DESIGN

Nini Anggraini[1*], Chantika Nabillah[1], Herdi Dermawan Lonan[1], Hartayuni Sain[1]
[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, Tadulako University
Soekarna-Hatta Street, Palu 94118, Central Sulawesi, Indonesia.

***e-mail**: *anggraininినia@gmail.com*

ABSTRACT

The purpose of this study was to examine the instruments used to assess students' abilities in the designer analysis course of the statistics study program at Tadulako University. There were 90 students enrolled in this course, and the questions included 25 multiple-choice items related to survey design content. The test instrument for understanding the survey designer course was the subject of this study. The Rasch method, which is used to get fit items, is used. Winsteps software was used to carry out this analysis. In accordance with the Rasch model, the Winsteps program produced 23 items with an average value of 1.11 and -0.08 for MNSQ Outfit for person and item, respectively. In spite of the fact that the person's and the item's Outfit ZSTD values are 1.11 and 0.26, respectively, and despite the fact that the instrument's reliability, as measured by Cronbach's alpha, is 0.86, 23 of the 25 question items fit and 2 do not.

Keywords: *Fit, the Rasch Model, Reliability, and Winsteps*

Cite: Anggraini, N., Nabillah, C., Lonan, H. D., & Sain, H. (2023). *Application of the Rasch Model to Test Tools in The Analysis of Survey Design*. *Parameter: Journal of Statistics, 3*(1), 19-27, https://doi.org/10.22487/27765660.2023.v3.i1.16412.

*Received: Jun 2023; Reviewed: Jun 2023; Published: Jun 2023

INTRODUCTION

Student comprehension and mastery of the material are measured through the evaluation of learning outcomes. According to Marjiastuti and Wahyuni (2014), this indicates that evaluating learning outcomes provides an overview of the information provided by each student regarding their comprehension and accomplishment of the competencies derived from the process. Pisca (2014) states that assessment plays a vital role in advancing education by conveying teacher objectives, increasing motivation, fostering good learning habits, and providing feedback through the identification of strengths and weaknesses in individual learning. This statement explains that good learning quality is the result of effective evaluation. Consequently, educators must conduct appropriate evaluation activities to assess students' academic abilities. One way to measure learner abilities is by using evaluation instruments, whether in the form of tests or non-tests. Test instruments are commonly used to measure students' abilities. In the form of descriptive questions, test instruments are believed to be able to accommodate any learning material taught to students by lecturers. The questions used in evaluation activities need to be of high quality because high-quality questions can precisely indicate which students have mastered the material taught by the lecturer. A test instrument is considered of good quality if it has high validity and reliability. According to Hayati (2016), the accuracy of a study's data increases with the validity and reliability of the instrument. This assertion is supported by Wahyuningsih (2015), who explains that aspects such as difficulty level and distinguishing power also contribute to the quality of the question set. Validity and reliability are important factors in determining whether a test is of good quality or not (Wahyuningsih, 2015).

In general, the development of various instruments in the educational context produces products that can be used to measure the understanding of educational outcomes in schools and universities (Sumarni et al., 2016). The development of good instruments is an important part of assessment and evaluation as a means to improve the standard of education (Sinaga, 2016).

The Rasch model is used to analyze test scores. This model is a component of item response theory used to analyze questions posed to test takers. By using a logarithmic function to produce measurements with equal intervals, the Rasch model focuses on measuring the relationship between the ability of the person answering the question (person ability) and the difficulty of the given question (item difficulty) (Olsen, 2003). When compared to other test takers, the Rasch model can identify test takers with high abilities because they are more likely to answer questions correctly. Students are more likely to answer questions with a low level of difficulty correctly compared to questions with a high level of difficulty (Azizah et al., 2020). According to Ardiyanti (2016), the analysis of questions using the Rasch model primarily focuses on response and correlation aspects. In the 1960s, Georg Rasch created an analytical model of item response theory (IRT) known as the 1PL model, which stands for "one logistic parameter." Ben Wright later popularized this mathematical model. Rasch constructed a model that connects students and items using raw data representing student abilities in the form of dichotomous data (true and false) (Purnomo, 2016).

In contrast to the analysis of items using modern theory or item response theory, the characteristics of items highly depend on the group of test takers when using classical theory in item analysis research. According to Ratnaningsih (2013), the basis of assessment shifts from the abilities of groups of test takers to those of individual test takers. Dichotomous data, also known as nominal or discrete data, represents either-or choices without fractional numbers. The simplest form of data organization is dichotomous data. An example of dichotomous data is gender in a class, where males are assigned the number one and females the number two (Pratikno et al., 2020).

Research conducted by Edi Istiyono (2013) in 10 high schools involving 1001 students in Yogyakarta produced an assessment instrument suitable for measuring high-level thinking skills of class XI high school students, but it does not cover all the material being taught. One of the materials that has not been covered is the subject matter of optics in class XI SMA/MA.

Based on the background described above, the researcher wanted to assess the quality of the test instruments used to determine students' abilities in material analysis by using the Rasch model approach. This quality is measured based on several indicators, namely item fit and item misfit. Therefore, a test instrument was designed, and subsequently, it was determined which items fit and which ones do not. Additionally, with the help of Winstep software, the Cronbach's alpha value will be determined to assess the reliability of the items.

MATERIALS AND METHODS

Data Sources

This was based on data study are primary data gathered using a questionnaire-based survey. The subjects in this study consisted of 90 students of the Statistics Study Program at Tadulako University.

Research Variables

Variables are either the subject of a study or a variety of research objects (Arikunto, 2006). In this review, the factors utilized were understudies taking overview configuration courses at the Tadulako College Measurements Review Program comprising of 90 understudies (Y) and 25 things on accomplishment test instruments tried on understudies (X).

Methods

The Rasch method was utilized in this study. The Rasch model is able to identify test takers with high ability because they are more likely to answer questions correctly than other test takers. Students are more likely to correctly respond to questions with a low level of difficulty than to correctly respond to questions with a high level of difficulty. (Sumintono, et al 2014). Rasch modeling incorporates an algorithm that mathematically expresses the probabilistic expectation of item "i" and respondent "n" for dichotomous data as:

$$P_{ni}\left(X_{ni}=\frac{1}{\beta_n},\delta_i\right)=\frac{e^{(\beta n-\delta_i)}}{1+e^{(\beta n-\delta_i)}} \qquad (1)$$

Where:

$P_{ni}\left(X_{ni}=\frac{1}{\beta_n},\delta_i\right)=\frac{e^{(\beta n-\delta_i)}}{1+e^{(\beta n-\delta_i)}}$: is the likelihood of success of item i and response n

δ_n : response ability

δ_i : item level of difficulty

$x = 1$: correct response

Data Analysis

Quantitative data analysis techniques obtained from questionnaires by respondents. The data obtained is categorical data so that it can be examined with the help of the Rasch model. In classical test theory, the Rasch model favors the total score of the results of the exam or questionnaire over each of the available options. The probabilistic expectation results for item "i" and respondent "n" in Rasch modeling are stated mathematically in the following manner: (Sumintono & Widhiarso, 2014).

1. Data Input.
2. Analyze the data by looking at the Summary Statistic.
3. Checking the item fit order using the INFIT MNSQ value of each item, the mean value and standard deviation which will indicate the item is infit and needs to be revised (discarded). According to Sumintono and Widhiarso (2015), the parameters used as a basis for revising items that do not fit refer to the following conditions: Nilai Outfit's Mean Square (MNSQ) score was 1.5, or 0.5.
   a. Accepted Value of Z-Standard Outfit (ZSTD) -2.0 ZSTD 2.0.
   b. Accepted Pt-measure Corr 0.8 Point Measure Corelation value of 0.32. inspect Test Information Function dan person measure.
4. Checking Scologram.
5. Drawing conclusions.

RESULTS AND DISCUSSION

Rasch modeling in this study was carried out with the help of winstep software to analyze data in order to test the validity to reliability of the developed instruments. The Rasch model considers the respondent's ability to answer each item or question and the difficulty level of the item Based on the criteria for determining the level of validation and reliability of the instrument are presented in table 1.

The interaction between the student and the item as a whole is shown in Table 1 to be good value with an Alpha Cronbach value of 0.86 and a person reliability value of 0.83, respectively. which can be concluded. The person reliability value is then 0.85, indicating that student answers are fairly consistent and that the items used in the test instrument are of high quality. These results are presented in full in Table 1.

Table 1. Summary Statistics

	Description	Value
Logit	Person	1.86
	Item	1.91
Reliability	Person Reliability	0.83
	Item Reliability	0.87
	Alpha Cronbach	0.86
Outfit MNSQ	Person	1.11
	Item	-0.08
Outfit ZSTD	Person	1.11
	Item	0.26

To see the measures for fit and non-fit questions in light of the aftereffects of MNSQ and ZSTD outfits. According to Table 1, the Outfit Mean Squared (Outfit MNSQ) value of 1.11 falls into the fit question category, which is between 0.5 and MNSQ 1.5. This indicates that the survey questions used in the test instrument follow the Rasch model to assess students' comprehension of the survey designer analysis course. In contrast, the value of the Outfit Z-Standardized (Outfit ZSTD) is 1.11 for the individual and 0.26 for the item. The fact that the values 1.11 and 0.26 fall within the range -2.0 ZSTD 2.0 indicates that the data may have a rational value. It is possible to draw the conclusion that each item conforms to the Rasch model and is a suitable instrument for testing students' abilities throughout the survey designer analysis course. in Picture 1 below indicate the categories of appropriate and inappropriate questions. The following two characteristics must be present for an inquiry to be considered reasonable. The MNSQ outfit esteem should be between 0.5 and 1.5, so the primary limit (question) should be appropriate. The next requirement is an outfit ZSTD score between 2.0 and 2.0, as well as a point-measure connection score between 0.4 and 0.85 for the overall score. The categories of fit and unfit questions can be seen from the distribution of items (questions) in Picture 1 below. The criteria for a suitable question are if it meets the following two characteristics. The first item (question) limit is said to be fit if the MNSQ outfit value lies in the range of 0.5 to 1.5. The second requirement is to have an outfit ZSTD value in the range -2.0 to 2.0 and for a point measure correlation value with a different total score in the range 0.4 to 0.85 (Sumintono & Widhiarso, 2014).

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
17	3	26	3.74	.94	.94	.10	.21	-.38	.76	.70	92.0	94.0	S17
12	5	26	2.40	.71	.65	-.58	.89	.14	.77	.70	96.0	90.6	S12
16	5	26	2.40	.71	.50	-1.01	.26	-1.04	.85	.70	96.0	90.6	S16
10	6	26	1.95	.64	1.22	.62	1.35	.70	.60	.67	84.0	88.1	S10
20	7	26	1.58	.58	.91	-.14	1.18	.49	.66	.64	88.0	85.1	S20
22	7	26	1.58	.58	1.64	1.59	3.01	2.61	.32	.64	80.0	85.1	S22
1	8	26	1.26	.54	.96	-.02	1.05	.27	.62	.62	84.0	81.8	S1
7	10	26	.73	.49	1.32	1.35	1.53	1.16	.41	.57	64.0	75.2	S7
24	10	26	.73	.49	.76	-1.04	.75	-.45	.66	.57	88.0	75.2	S24
2	12	26	.28	.47	1.29	1.48	1.23	.61	.40	.52	52.0	71.5	S2
14	12	26	.28	.47	1.01	.10	.92	-.01	.52	.52	68.0	71.5	S14
18	13	26	.06	.46	1.06	.40	.92	.03	.48	.50	68.0	70.9	S18
6	15	26	-.35	.46	.79	-1.34	.69	-.35	.55	.45	80.0	70.4	S6
9	16	26	-.56	.46	.76	-1.55	.66	-.33	.54	.43	84.0	69.9	S9
13	18	26	-.99	.47	.74	-1.51	.58	-.28	.50	.38	76.0	72.0	S13
21	18	26	-.99	.47	.78	-1.23	.65	-.18	.48	.38	84.0	72.0	S21
25	18	26	-.99	.47	.93	-.33	.76	-.02	.42	.38	68.0	72.0	S25
15	19	26	-1.22	.49	1.39	1.82	3.47	1.90	.08	.35	68.0	73.5	S15
4	20	26	-1.46	.50	.86	-.56	.69	.04	.40	.33	76.0	76.0	S4
19	20	26	-1.46	.50	1.09	.46	3.63	1.85	.19	.33	76.0	76.0	S19
23	20	26	-1.46	.50	.86	-.56	.69	.04	.40	.33	76.0	76.0	S23
11	21	26	-1.73	.53	.74	-.96	.49	-.28	.43	.30	80.0	80.0	S11
3	24	26	-2.88	.76	1.00	.19	.63	-.01	.21	.19	92.0	92.0	S3
5	24	26	-2.88	.76	.92	.05	.50	-.18	.25	.19	92.0	92.0	S5
8	26	26	-4.90	1.84	MINIMUM MEASURE				.00	.00	100.0	100.0	S8
MEAN	14.3	26.0	-.20	.61	.96	-.11	1.11	.26			79.7	79.2	
P.SD	6.6	.0	1.91	.28	.26	.97	.91	.83			10.8	8.1	

Figure 1. The Rasch Model indicates that the item data distribution is inappropriate

From the results of the Winstep software analysis in Figure 1 Item 22 does not meet the criteria for the MNSQ, ZSTD, and PT Measure or is included in the invalid criteria so it must be corrected or not fit. However, if it does not meet the MNSQ and PT-Measure corr values, however, ZSTD value still meets or falls into the category of valid questions so that it only needs to be corrected, it does not need to be discarded or classified as fit, namely in question numbers 7, 15, and 19 it is still suitable for use. The number of fit items that meet the criteria is 23 items out of 25 items for the survey designer analysis course.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
17	3	26	3.74	.94	.94	.10	.21	-.38	.76	.70	92.0	94.0	S17
12	5	26	2.40	.71	.65	-.58	.89	.14	.77	.70	96.0	90.6	S12
16	5	26	2.40	.71	.50	-1.01	.26	-1.04	.85	.70	96.0	90.6	S16
10	6	26	1.95	.64	1.22	.62	1.35	.70	.60	.67	84.0	88.1	S10
20	7	26	1.58	.58	.91	-.14	1.18	.49	.66	.64	88.0	85.1	S20
22	7	26	1.58	.58	1.64	1.59	3.01	2.61	.32	.64	80.0	85.1	S22
1	8	26	1.26	.54	.96	-.02	1.05	.27	.62	.62	84.0	81.8	S1
7	10	26	.73	.49	1.32	1.35	1.53	1.16	.41	.57	64.0	75.2	S7
24	10	26	.73	.49	.76	-1.04	.75	-.45	.66	.57	88.0	75.2	S24
2	12	26	.28	.47	1.29	1.48	1.23	.61	.40	.52	52.0	71.5	S2
14	12	26	.28	.47	1.01	.10	.92	-.01	.52	.52	68.0	71.5	S14
18	13	26	.06	.46	1.06	.40	.92	.03	.48	.50	68.0	70.9	S18
6	15	26	-.35	.46	.79	-1.34	.69	-.35	.55	.45	80.0	70.4	S6
9	16	26	-.56	.46	.76	-1.55	.66	-.33	.54	.43	84.0	69.9	S9
13	18	26	-.99	.47	.74	-1.51	.58	-.28	.50	.38	76.0	72.0	S13
21	18	26	-.99	.47	.78	-1.23	.65	-.18	.48	.38	84.0	72.0	S21
25	18	26	-.99	.47	.93	-.33	.76	-.02	.42	.38	68.0	72.0	S25
15	19	26	-1.22	.49	1.39	1.82	3.47	1.90	.08	.35	68.0	73.5	S15
4	20	26	-1.46	.50	.86	-.56	.69	.04	.40	.33	76.0	76.0	S4
19	20	26	-1.46	.50	1.09	.46	3.63	1.85	.19	.33	76.0	76.0	S19
23	20	26	-1.46	.50	.86	-.56	.69	.04	.40	.33	76.0	76.0	S23
11	21	26	-1.73	.53	.74	-.96	.49	-.28	.43	.30	80.0	80.0	S11
3	24	26	-2.88	.76	1.00	.19	.63	-.01	.21	.19	92.0	92.0	S3
5	24	26	-2.88	.76	.92	.05	.50	-.18	.25	.19	92.0	92.0	S5
8	26	26	-4.90	1.84	MINIMUM MEASURE				.00	.00	100.0	100.0	S8
MEAN	14.3	26.0	-.20	.61	.96	-.11	1.11	.26			79.7	79.2	
P.SD	6.6	.0	1.91	.28	.26	.97	.91	.83			10.8	8.1	

Figure 2. The results of the distribution of the difficulty levels of the questions

From the outcomes of the Rasch Model analysis of the items as many as 25 items tested on students of the statistics study program of the survey designer course are obtained in figure 2. Based on the results of the analysis of questions with Winsteps software from picture 2, it can be read the data on the distribution of the category of difficulty level of the question, namely in the group of very difficult questions there are three questions, namely on question items number 17, 12, and 16. The difficult question group has nine questions, namely on question items number 10, 20, 22, 1, 7, 24, 2, 14, and 18. The group of easy questions contained six questions, namely on item numbers 6, 9, 13, 21, 25, and 15. The group of very easy questions contained seven questions, namely on item numbers 4, 19, 23, 11, 3, 5, and 8. With a percentage of very difficult questions 16%, difficult questions 44%, easy questions 32% and very easy questions of 34%.

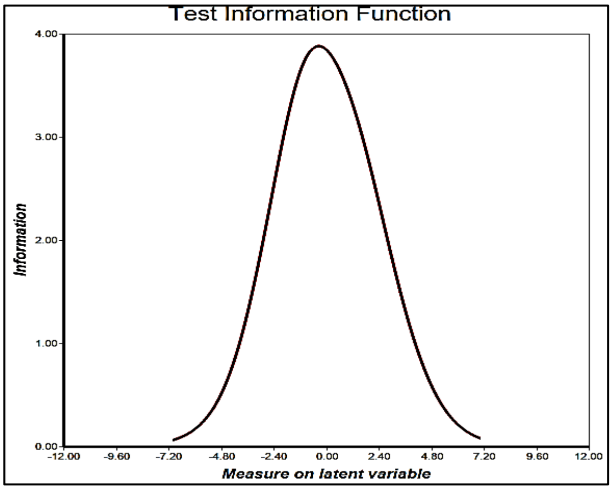

Figure 3. Test Information Function

The measurement data obtained from the survey design analysis course's student understanding test instrument is depicted in figure 3. The information function's value is represented by the y-axis, while the x-axis indicates the student's level of proficiency on the test. The measurement's information is very high on this graph at a high level of ability. so that the growth of student understanding test instruments in the survey designer analysis course is suitable or optimal if used on the ability of students with high abilities.

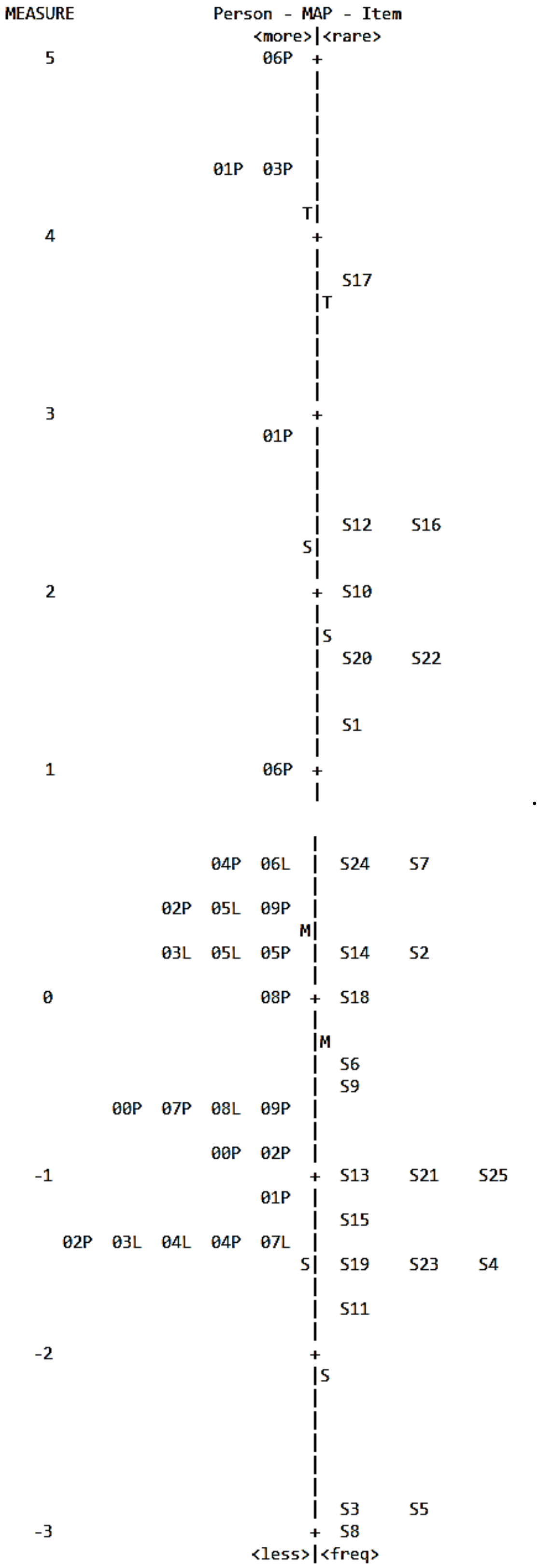

Figure 4. *Person Measure*

Figure 4 depicts the Analyses of data distribution ability level of the difficulty items while the left part is distribution students' abilities. The capacity of students to solve issues analyzed with the help of Winstep software in the form of a wright map can be seen in Figure 4. From the wright map obtained student data with codes (P = Female) and (L = Male) and the number given is the sequential number of student data when testing in the table. Problem number 17 is the most difficult and problem number 8 is the easiest, the student with highest ability is 06P while students who have the lowest skill 02P, 03L, 04L, 04P, and 07L. It is clear that students 06P, 01P, and 03P can answer correctly the most difficult item, item number 17. Another analytical tool used is a scologram that proves the methodical response patterns between people who took the test (sorted from high to low) ability vertically, or downward) and questions (sorted horizontally from easiest to hardest, from left to right).

```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |   1 121122  1 1 2 221111
       |8351493531596824741020267
       |-------------------------
    26 +1111111111111111111111111  06P
    11 +1111111111111111111111110  01P
    23 +1111111111111111111101111  03P
     1 +1111101011111111111110111  01P
     6 +1111111111111101010000010  06P
    14 +1111110111111101011000000  04P
    16 +1111111111110001010101000  06L
     2 +1111111111111000010001000  02P
    15 +1111110111111010110000000  05L
    19 +1111111111101100100010000  09P
     5 +1111101111011110001000000  05L
    13 +1111110101110110001000000  03L
    25 +1111010111111110001000000  05P
    18 +1111111111111000100000000  08P
     7 +1111101001111000000000100  07P
     8 +1111111111000010000000000  08L
     9 +1111111100010101000000000  09P
    10 +1111111111110000000000000  00P
    12 +1010011100001111100000000  02P
    20 +1111110101001100000000000  00P
    21 +1001010110000100100011000  01P
     3 +1110010000000011100010000  03L
     4 +1110101010100000010000000  04L
    17 +1100010100000001010101000  07L
    22 +1110100100110000000100000  02P
    24 +1111001101100000000000000  04P
       |-------------------------
       |   1 121122  1 1 2 221111
       |8351493531596824741020267
```

Figure 5. Scologram

In Figure 4 above the items are on the left. Student 01P is included in the category of careless students where the second easiest item (no. 3) and the third (no. 5), the student cannot do it correctly, even though the more difficult problem can be done, namely item no. 10. In the scologram above there is no similar response pattern so there is no indication of mutual cooperation between students. In the assessment analysis, this instrument meets the requirements for fit with the Rasch model as evidenced by the MNSQ infit value between 0.5-1.5, namely 1.11 and the instrument reliability value stated in Cronbach's alpha is 0.86. Whereas in previous research by (Edi Istiyono, 2013) the MNSQ infit value was 0.81 and Cronbach's alpha value was 0.97.

CONCLUSION

The 25 items tested were analyzed using the Rasch Model, and the results show that 23 of them (or 94 percent) are suitable and can be used, while two of them (or 6 percent) are unsuitable and cannot be used, so they need to be replaced. Using the Rasch Model, the item analysis of the survey designer analysis course on student understanding revealed that with an item reliability value of 0.87 and a person reliability value of 0.83, the consistency of responses is sufficient. With an Alpha Cronbach value of 0.86, the items fall into the excellent quality category. whereas the person column's Outfit Mean Square Statistic (Outfitt MNSQ) value is 1.11, the item column's value is -0.08. In the person table, the value of the Outfit Z Standard (Outfit ZSTD) is 1.11, while it is 0.26 in the item table. On the other hand, out of the total of 25 items, 23 were able to fit, while two were unable to.

REFERENCE

Arikunto, S., & Suhramini. (2006). *Prosedur Penelitian Suatu Pendekatan Praktik* (Edisi Revisi VI, Cetakan ke-13). Jakarta: PT. Asdi Mahasarya.

Ardiyanti, D. (2016). *Penerapan Model Rasch pada Pengembangan Skala Efisiensi Diri untuk Tujuan Penggunaan Karir Siswa*. *Jurnal Psikologi*, 43, 248–263.

Azizah, N., & Wahyuningsih. (2020). "Genggam Bola Untuk Mengatasi Hambatan Mobilitas Fisik Pada Pasien Stroke Non Hemoragik," 4(1), hlm.

Friendly, S., Pratikno, A. S., & A.A. (2020). *Pemetaan Ukuran dengan Pemusatan Data*. Pracetak dari OSF, 25(03), 2-3.

Hayati, S., & Lailatunissaadah, L. (2016). *Validitas dan Reliabilitas Aktivasi, Kreativitas, dan Model Metodologi Implementasi Rasch Ilmiah Didaktika Jurnal*, 16(2), 169-179.

Istiyono, E. (2013). *Pengembangkan Tes Kemampuan Berpikir Tingkat Tinggi Fisika Siswa SMA (pysthots). Jurnal Penelitian dan Penilaian Pendidikan UNY.*

Marjiastuti, K., & Wahyuni, S. (2014). *Analisis Model Rasch Kemampuan Siswa*. *Seminar Nasional Evaluasi Pendidikan 2014*, 94-102.

Olsen, L. W. (2003). *Georg Rasch dan Kemajuan yang Dia Buat dalam Statistik*. Di Institute of Economics, University of Copenhagen (tesis PhD).

Pisca, C. C. (2014). *Evaluasi Komparatif Penilaian Butir Ujian Bahasa Prancis di SMAN 10 Yogyakarta Tahun Pelajaran 2013–2014 Dari Perspektif Klasik dan Kontemporer* (Thesis Ph.D.). Perguruan Tinggi Negeri Yogyakarta.

Purnomo, A. R. (2016). *Fondasi Pertumbuhan Indonesia Adalah Ekonomi Kreatifnya*. Surakarta: Visi Media Ziyad.

Ratnaningsih, D. J. (2013). *Berdasarkan Teori Tes Saat Ini, Analisis Tes Objektif Ujian Akhir Semester Mahasiswa Universitas Terbuka*. *Jurnal Pendidikan Terbuka dan Jarak Jauh*, 14(2), 98-109.

Sinaga, N. (2016). *Membuat Penilaian Kemampuan Berpikir Matematis Siswa dan Menyelesaikan Masalah di Kelas VIII*. *PYTHAGORAS: Jurnal Pendidikan Matematika*, 11(2), 169–181.

Sudirman, Wiyanto, Supartono, & Sumarni, W. (2016). *Transformasi Ilmu Pengetahuan Masyarakat Adat Menjadi Pengetahuan Ilmiah tentang Proses Pembuatan Gula Aren*, 13(4), 281-292. *Jurnal Pendidikan Sains Turki*.

Sumarni, W., Sudirman, Wiyanto, & Supartono. (2016). *The Reconstruction of Society Indigenous Science into Scientifie Knowledge in the Production Process of Palm Sugar*. *Jurnal of Tuurkish Science Education*, 13(4), 281-292.

Sumintono, B., & Widhiarso, W. (2014). *Penerapan Model Rasch pada Penelitian Ilmu Sosial* (Edisi Revisi). Bandung: Rumah Penerbitan Komunikata.

Wahyuningsih, T. E. (2015). *Analisis Butir Soal Tes Objektif yang Dikembangkan oleh Pengajar Ulangan Semester Ganjil Kelas X di SMA Negeri 1 Mlati Tahun Pelajaran 2013–2014* (Skripsi PhD). Personalia Aspek Keuangan, Perguruan Tinggi Negeri Yogyakarta.