Comparasional Analysis of Statistical Results Using the Rasch and Anates Model in Chemistry Learning

Fatma Harian Dini[1], Siti Khadijah Dalimunthe[2], Ayi Darmana[3], Retno Dwi Suyanti[4], Zainuddin Muchtar[5], Ajat Sudrajat[6]

{ fatmadhezar@gmail.com}

Universitas Negeri Medan

Abstract. The purpose of this study was to describe the comparative results of the analysis of student scores through the Rasch and Anates model approach in terms of the validity, reliability, difficulty level, and distinguishing power of the questions. This research is a qualitative research with descriptive method. The data were obtained through 46 students with a total of 40 items. Furthermore, the data obtained were analyzed through the Rasch and Anates model approach. The results showed that the quality of the instrument for measuring concept understanding through the Rasch and Anates model approach was in a good category. Reliability analysis through both applications is in the high category of 0.75 through Anates and 0.76 Person Reliability with the high category, 0.90 Reliability Items with a very high category in the Rasch model. The average level of difficulty index in Anates is with five levels of categories, while the results of the analysis of the level of difficulty based on the Rasch model show that there are four categories of item difficulty levels, namely 7 questions which are very easy, 12 easy questions, 11 difficult questions, and 10 very difficult questions. difficult. The discriminatory power of items using Anates is in the bad category of 3 questions, the sufficient category is 4 questions, and 9 items are in the good category. The discriminatory power of items from Rasch modeling obtained three groups of items based on the item separation index (H=2.73).

Keywords: Item Analysis, Anates, Rasch Model, Reliability, Validity, Difficulty Level, Distinguishing Power

1 Introduction

In learning at school, teachers have goals and paths that must be taken by their students based on established curriculum standards. In its implementation in the classroom, the teacher uses a variety of tools, media, models, methods and strategies to support the achievement of the goals set in learning and to see the learning outcomes that students have been able to achieve. One way that can be used to see the extent of student achievement is by conducting assessment or evaluation of student learning outcomes.

Evaluation is a process of measuring and assessing, these two things are very closely related to the evaluation system. Measurement is a process of comparison between measuring objects and certain measuring instruments that are carried out systematically. While the assessment is the interpretation of the measuring results. The measuring results are obtained from the instrument used or in the form of a test instrument. The form of the test instrument given can be in the form of objective test instruments and subjective test instruments. The test instrument

AISTEEL 2023, September 19, Medan, Indonesia

DOI 10.4108/eai.19-9-2023.2340586

is said to have good quality if it has high validity and reliability. The higher the value of the validity and reliability of an instrument, the more precise the data obtained from a study (Hayati & Lailatussaadah, 2016).

Item analysis needs to be done to test the quality of each item and a set of questions in various aspects. Item analysis can be done qualitatively or quantitatively. The main purpose of item analysis is to obtain information about the characteristics of each item, either through item analysis or empirical analysis. The results can be used to determine the quality of the questions and the quality of student learning from the analysis of exam results. Test questions should have discriminated power between students who are good at students who are not very good at it. In addition, it also has a level of difficulty because it is the purpose of the test or assessment4. Tools for analyzing item items have been developed, including SPSS, RASCH Model, Anates, Iteman, Bilog. different power.

In the 1960s an expert developed an analytical model using ministep applications and accurate analysis results, namely the Rasch Model test theory. The Rasch model is very easy to do and apply with accurate analysis results, as well as looking at the opportunity to answer the questions correctly by comparing students' abilities with the level of difficulty of the questions (Khotimah & Sri, 2014). The Rasch model has a fixed variable difficulty level, regardless of the sample involved in the initial variation (Wei et al, 2012). Rasch developed a data measurement model that can determine the relationship between the level of the student's own ability (person ability) and the item difficulty level (item difficulty) by using a logarithmic function to be able to produce measurements with the same interval values (Bambang, 2014). The advantages of the Rasch model according to Sumintono & Widhiarso (2013), the Rasch model overcomes grain problems, is quite resistant to missing data, and has fulfilled objective measurements.

Analysis using the Anates V4 program is a simple computer application that is easy to implement, fast and accurate. This application was designed by Karno and Yudi Wibisono which is software specifically developed to analyze multiple choice tests and essay tests. The benefit of Anates is that it can analyze the items automatically checking the right and wrong answers quickly and practically. The advantages of this program are easy to understand because the instructions for running the program are in Indonesian and the results of the analysis can be transferred to Ms Excel to calculate the value. Anates' ability to find out the analysis of the items which include: reliability, superior and asor groups, discriminating power, level of difficulty, correlation of item scores with total scores and the quality of distractors (Wiguna et al, 2018).

This study tries to analyze the comparison of the quality of test instruments on the elements of validity, reliability, difficulty level and item discrimination using the two approaches described above, namely Nodel Rasch and Anates. The test instrument used in this study was the pretest instrument for the National Examination scores in chemistry learning. Analysis of the items on the pretest instrument in chemistry learning is carried out as an important effort in improving the quality of evaluation tools, developing evaluation tools, and increasing the objectivity of a test in measuring students' abilities as feedback on learning activities. In addition, it is very important to maintain and improve the quality of research instruments to avoid defects.

2 Method

This study uses secondary data as a result of measuring the results of the 2012 National Examination questions to measure understanding of the concept of chemical material in one of the public school classes in Medan City which was obtained around 2021. Secondary data was obtained through the documentation method, then written documents were obtained in the form of multiple choice instrument questions totaling 40 questions and 46 written answer sheets. This research is a qualitative research with descriptive method. This research is intended to find information and data that can be used to empirically describe the quality of test instruments based on the elements of validity, reliability, level of difficulty, and the differentiating power of questions through the Anates approach which is processed with the help of the Anates V4 application program and the Rasch model with Winsteps software.

3 Results And Discussion

3.1 Test Validity Test Result

Test the validity of the test can be measured by looking at the correlation of item scores (item scores) with the total score. The following is processed data in the form of correlation and significance of the item tips using anates. Analysis of the quality of the questions for each item can be seen from the aspects of validity, reliability, and the level of difficulty of the questions. The validity of the items was tested using the product moment correlation method. The question is said to be valid if the correlation coefficient is rxy > rtable and if rxy ≤ rtable then the question can be considered invalid (Syofian, 2015). In Anates test, the validity of the test can be measured by looking at the correlation between the item scores (item scores) and the total score. In the Rasch Model, to see the quality of the items from the aspect of validity is if they meet the following criteria (Sumintono & Widhiarso, 2015).

a) The value of the Outfit MNSQ (Mean Square) received is: 0.5 < Outfit – MNSQ < 1.5
b) The accepted Outfit ZSTD (Z – Standard) value is: –2.0 < ZSTD < +2.0
c) Value of Pt Measure Correlation (Point Measure Correlation): 0.4 < Point Measure Corr < 0.85

After testing the validity aspects of each item both theoretically using the Anates test and the Rasch model, the results of a comparison of the results of the analysis of the validity of the items using the Anates and Rash models were obtained, as can be seen in Table 1.

Table 1. Comparison of the results of the analysis of the validity of the items using the theoretical approach of the Anates test and the Rasch model

Result	Question Number Anates	Rasch Model
Valid	1,5,6,7,10,11,12,13,15,16,17 ,18,19,20,22,23,24,25,26,30	1,5,6,7,8,16,18,19,20 ,24,25,26,28
Invalid	2,3,4,8,9,14,21,27,28,29,31, 33,35,36,37,38,39,40	9,4,3,40,39,12,15,17, 14,2,4,10,11,32,34,3 6,33,38,35,27,30,37, 29,21,23,22,31

Based on the results of the analysis of the quality of the questions in terms of the validity of the items, information was obtained that theoretically the Anates test contained 23 questions which were stated to be valid and 18 questions included in the invalid category. Whereas in the analysis of item quality using Rasch modeling, 13 items were said to be valid and 27 questions were said to be invalid because they did not meet the requirements for the MNSQ outfit, ZSTD outfit, and Point Measure Correlation (Pt Measure Corr).

Based on the results of the quality analysis of the items on the validity aspect, it is known that there are differences in the results of the analysis which are considered valid by the two approaches, both theoretically the Anates test and the Rasch modeling. In theory, the Anates test obtained more valid items than valid items in the Rasch modeling. Eleven valid questions from Rasch's modeling are also valid according to Anates' test theory approach. The quality of the items through quality testing is more in Rasch modeling than with the Anates approach because in Rasch modeling, the analysis of the items with the Rasch modeling approach is felt to be more accurate because when the item meets 3 (three) criteria, namely Outfit MNSQ score, Outfit ZSTD score, as well as the Point Measure Correlation value, the question is considered valid.

3.2 Reliability Of Question Item

Table 2. Comparison of the results of the analysis of the validity of the items using the theoretical approach of the Anates test and the Rasch model.

Reliability analysis using Anates		Analysis using the Rasch Model			
Reliability	**Category**	**Person Reliability**	**Category**	**Reliability Item**	**Category**
0,75	High	0,76	High	0,90	Very high

Table 2 provides information on the results of the instrument analysis on the reliability element through Anates in terms of the reliability value. The reliability value obtained is 0.75 with a high interpretation (based on the criteria used to interpret the reliability correlation value of the instrument using Anates). While the analysis using the Rasch model obtained an item reliability value of 0.90 (very high) in terms of the reliability criteria of the item through the Rasch model. This means that the reliability of the questions through Anates analysis and the Rasch model is reliable for use on the same subject even though at different times, places and by different people. In addition to the reliability of the questions, the analysis of the Rasch model can also be seen that the value of person reliability or student reliability is equal to 0.76, meaning that the consistency of student answers is high. out of 46 students who can answer questions consistently. Based on the results of the analysis of the two approaches, seen from the value and reliability category of the questions, it shows that the analysis using the Anates approach and the Rasch model has the same analysis results at high criteria.

Code. Program listings or commands in the text are set in typewriter font (CMTT10 or Courier) and referred to in the text.

3.3 Difficulty Level of Items

The difficulty level of the item indicates the probability of how many respondents can answer a question item correctly. In the theory of measuring the items in Anates, the difficulty index of the items is interpreted according to the following criteria (Lestari & Yudhanegara, 2017):

Table 3. Criteria and Results of the Distribution of Item Difficulty Levels with the Anates Test Theory Approach

Problem Difficulty Index Value	Difficulty Index Interpretation	Number of Question Items
DI = 0,00	Very Difficult	6
0,00 < DI ≤ 0,30	Difficult	5
0,30 < DI ≤ 0,70	Currently	15
0,70 < DI ≤ 1,00	Easy	9
DI = 1	Very easy	3

In the Rasch modeling, the difficulty level of the items is categorized based on the Logit Measure and the Logit Item Standard Deviation (SD) value and is divided into four categories as follows (Sumintono & Widhiarso, 2015):

Table 4. Criteria and Results of Item Difficulty Level Distribution with Rasch Modeling

Measure Value (Logit)	Interpretation of Item Difficulty	Number of Question Items
Measure logit < -1,12	Very easy items	7
-1,12 ≤ measure logit ≤ 0,00	Easy items	12
0,00 ≤ measure logit ≤ 1,12	Difficult items	11
Measure logit > 1,12	Very difficult items	10

Based on Table 3 and Table 4 it can be seen the comparison between the results of the analysis of the questions on the aspect of difficulty level through the theoretical approach of the Anates test and Rasch modeling. Based on these two tables it can be seen that most of the difficulty levels of the items analyzed with the classical test theory approach are in all categories. With an average of the most questions are questions with moderate criteria, namely as many as 15 questions.

Different results were obtained after the items were analyzed using Rasch modeling. The level of difficulty or difficulty of the items in Rasch modeling can be seen from the measure value in the logit unit of each item. Based on the criteria in table 4 it can be seen that from the Measure (logit) value the items can be grouped into 4 (four) categories of item difficulty levels, namely very easy items, easy items, difficult items, and very difficult items with an almost even distribution.

TABLE 3.1 D:\MAGISTER\SEMESTER 2\STATISTIK KEPEN ZOU446WS.TXT\ May 23 2021 18:45RN.pr
INPUT: 46 Person 25 Item REPORTED: 46 Person 25 Item 2 CATS MINISTEP 4.8.1.0

SUMMARY OF 46 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	16.4	25.0	.96	.55	.97	-.08	1.10	.02
SEM	.7	.0	.18	.02	.05	.19	.13	.19
P.SD	4.4	.1	1.23	.11	.36	1.29	.87	1.25
S.SD	4.5	.1	1.24	.12	.36	1.30	.88	1.26
MAX.	24.0	25.0	4.05	1.07	1.86	2.45	5.25	2.78
MIN.	5.0	24.0	-1.82	.46	.43	-2.49	.21	-1.93
REAL RMSE	.60	TRUE SD	1.07	SEPARATION	1.80	Person	RELIABILITY	.76
MODEL RMSE	.56	TRUE SD	1.09	SEPARATION	1.94	Person	RELIABILITY	.79
S.E. OF Person MEAN = .18								

Person RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .81 SEM = 1.96
STANDARDIZED (50 ITEM) RELIABILITY = .88

SUMMARY OF 25 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	30.1	46.0	.00	.40	.99	-.05	1.10	.14
SEM	2.0	.0	.27	.01	.03	.18	.13	.21
P.SD	10.0	.2	1.32	.05	.16	.90	.62	1.05
S.SD	10.2	.2	1.35	.06	.17	.92	.63	1.07
MAX.	41.0	46.0	2.93	.56	1.32	1.71	3.33	2.36
MIN.	8.0	45.0	-1.90	.34	.74	-1.64	.39	-1.38
REAL RMSE	.41	TRUE SD	1.26	SEPARATION	3.03	Item	RELIABILITY	.90
MODEL RMSE	.40	TRUE SD	1.26	SEPARATION	3.13	Item	RELIABILITY	.91
S.E. OF Item MEAN = .27								

Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Fig 1. Summary Statistics Results for 40 items with Rasch modeling

3.4 Discriminating Power of Questions

The item discriminating power is the ability of the item to distinguish students who are able to answer questions or students who have a high level of ability from students who have low ability to answer questions. The results of the calculation of the item discriminating power index based on the classical test theory using Anates can generally be categorized into three categories as shown in Table 5.

Table 5. Item Difference Power Index Through Classical Theory Using Anates

Distinguishing Power (DP)	Interpretation	Number of Question Items
DP ≥ 0,70	Very well	12
0,40 ≤ DP < 0,70	Good	9
0,20 ≤ DP < 0,40	Enough	4
DP < 0,20	Bad	3
D ≤ 0	Very bad	12

Based on Table 5, it can be seen that the results of the analysis of the differentiating power of the questions through the classical test theory approach using Anates showed that most of the differentiating power of the questions were in the good and very good categories with item

items sequentially 5, 8, 11, 14, 17, 22, 23, 24, 28 and 1, 6, 7, 12, 13, 16, 18, 19, 20, 25, 26, 30. The discriminating power of the questions was in the sufficient category with 4 items, 3 questions in the bad category and 12 items in the very bad category.

In contrast to the theoretical approach of Anates' test, in Rasch modeling analysis is used at the level of individual ability as a tool to distinguish the ability of students who are able to answer questions and those who are unable to answer questions. In addition, you can also use a way to identify groups of respondents based on the respondent's separation index. The greater the item separation value, the better the quality of the instrument in terms of all respondents and item items, because it can identify groups of respondents and groups of items (Sumintono & Widhiarso, 2015). Another equation to find out more precise grouping is used the stratum equation (H):

$$H = \frac{[(4\ X\ SEPARATION) + 1]}{3}$$

Based on Figure 1 it is known that the value of the separation of the items is 3.03, so the value of H = 4.37 is rounded to 4, so that there are four groups of items and for the respondents a separation value of 1.80 is obtained with H = 2.73 rounded to 3, shows that the respondent group can be divided into two groups based on the respondent's separation value.

4 Conclusion

The results showed that the quality of the instrument for measuring conceptual understanding through Anates was of good quality in terms of validity while the Rasch model analysis did not have good quality. Reliability analysis through both approaches is included in the high category of 0.75 using Anates and 0.76 in the Rasch model. Based on the difficulty level index, the results using Anates already have good quality, while the results of the analysis using the Rasch model show various levels of difficulty, namely very easy, easy, difficult and very difficult. On the discriminating power index, most of the discriminating power of the instrument items using Anates was in the good category. As with the analysis of the Rasch model, the questions were also included as questions that had not good quality because based on the results of the analysis through the separation values of the respondents, it was obtained 3.

Acknowledgments. Thank you to the Lecturer in Chemistry Education Statistics and friends of the Class A Chemistry Education Postgraduate Program for the data suggestions for the availability of data in supporting this research.

References

[1] Bambang, S. (2014). *Model Rasch untuk Penelitian Sosial Kuantitatif*. Surabaya

[2] Hayati, S., & Lailatussaadah, L. (2016). Validitas Dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif Dan Menyenangkan (Pakem) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika*, 16(2), 169–179

[3] Khotimah, M & Sri W.2014. *Analisis Kemampuan Peserta Didik dengan Model Rasch* : Universitas Negeri Yogyakarta

[4] Lestari, K., & Yudhanegara, M. (2017). *Penelitian Pendidikan Matematika*. Bandung: PT Refika Aditama

[5] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata

[6] Wei, S., Liu, X., Wang, Z, & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, 89(3), 335-345

[7] Widyaningsih, S. W., & Yusuf, I. (2018). Analisis Soal Modul Laboratorium Fisika Sekolah I Menggunakan Racsh Model.Gravity. *Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, *4*(1). https://doi.org/10.30870/gravity.v4i1.3116

[8] Wiguna, Satria, Syaukani, dan Ananda, Rusydi. (2018). Kemampuan Guru PAI dalam Merancang Tes (Analisis Aplikasi Anates Ganda Di Sekolah SMA Negeri 1 Hinai), *EDU Riligia*, 2 (1) : 1-10