Vol 13(1), 2024, 77-90 DOI: 10.23960/jppk.v13.i1.2024.08
Jurnal Pendidikan dan Pembelajaran Kimia
e-ISSN: 2714-9595| p-ISSN 2302-1772
http://jurnal.fkip.unila.ac.id/index.php/JPK/index

Analysis and Development of HOTS-Based Reaction Rate Problem Instruments Using the Rasch Model

Santri Angelia Damanik[1*], Inovayani Saragih[2], Ayi Darmana[3]

[1,2,3] Master's Program in Chemistry Education, State University of Medan.

*Correspondinge-mail: santriangelia@gmail.com

Received: November 16th, 2023 Accepted: April 2nd, 2024 Online Published: April 19th, 2024

Abstract: Analysis and Development of HOTS-Based Reaction Rate Problem Instruments Using the Rasch Model. The research carried out was research and development to obtain instruments for HOTS-based reaction rates. The development of this item instrument is carried out first by following the steps for preparing HOTS questions which of course also pay attention to the characteristics of the HOTS questions. The next step is to analyze the item items based on the response pattern of the answers given by the students using the Rasch model assisted by the WINSTEPS 3.73 application. In this study the results of student answers were analyzed based on eight (8) types of output tables. Based on the analysis using the Rasch model, the following results are obtained: 1) Results of the Output tables item fit order obtained Outfit values: MNSQ (0.70 – 1.61); ZSTD (-1.3 – 2.0) and Pt Measure Corr (0.29 – 0.64). 2) Summary statistics: Person Reliability (0.74 & 0.76) enough category; Item Reliability (0.71 & 0.73) is sufficient category, and Cronbach Alpha Reliability (0.81) is very good category. The final result is concluded, as many as 1 question out of 20 questions developed (5%) fall into the misfit category.

Keywords: Reaction Rate Question Instrument, HOTS, Rasch Model.

Abstrak: Analisis dan Pengembangan Instrumen Soal Laju Reaksi Berbasis HOTS dengan Menggunakan Rasch Model. *Penelitian yang dilakukan adalah penelitian dan pengembangan (research & development) untuk memperoleh instrumen soal laju reaksi yang berbasis HOTS. Pengembangan instrument soal ini dilakukan terlebih dahulu dengan cara mengikuti langkah-langkah penyusunan soal HOTS yang tentunya juga memperhatikan karakteristik dari soal HOTS tersebut. Langkah selanjutnya adalah menganalisis item soal berdasarkan pola respon jawaban yang diberikan siswa menggunakan Rasch model berbantuan aplikasi WINSTEPS 3.73. Dalam penelitian ini hasil jawaban siswa dianalisis berdasarkan delapan (8) jenis output tables. Berdasarkan analisis menggunakan Rasch model diperoleh hasil sebagai berikut : 1)Hasil Output tables item fit order diperoleh nilai Outfit : MNSQ (0,70 – 1,61); ZSTD (-1,3 – 2,0) dan Pt Measure Corr (0,29 – 0,64). 2) Summary statistics : Person Reliability (0,74 & 0,76) kategori cukup; Item Reliability (0,71 & 0,73) kategori cukup, dan Cronbach Alpha Reliability (0,81) kategori bagus sekali. Hasil akhir diambil kesimpulan, sebanyak 1 soal dari 20 soal yang dikembangkan (5 %) masuk kedalam kategori misfit.*

Kata kunci: *Instrumen Soal Laju Reaksi, HOTS, Rasch Model.*

▪ INTRODUCTION

We are in a new era, the era of digital industrialization where industrial activities are integrated through the massive use of wireless technology and big data. Currently, various kinds of human needs have widely applied the support of the internet and the digital world as a vehicle for interaction and transactions. Sharing economy, e-education, egovernment, cloud collaborative, marketplace, smart city are the face of today's world that is increasingly complex, rapidly changing, and both challenging and threatening. The world of education needs to prepare students to face the increasingly complex challenges of the 21st century. Education is not enough to equip learners with knowledge and simple thinking processes as known so far, but also needs to prepare them to have and be able to develop essential skills for this century (Kemendikbud; 2019).

21st century education focuses on creativity and innovation skills. This can be realized by providing knowledge at each level of education and training them to have the ability to solve problems, think critically and creatively. These traits will grow if trained, students are accustomed since childhood to explore, inquiry, discovery and solve problems (Alam, S.; 2019). By thinking critically, students will be able to identify, analyze and examine a problem sharply so that they can find space to find the best solution to the problem. This does not rule out the possibility that students will find and even create new solutions that have not been known before. Thus new ideas will always appear and make students more creative (Susilowati, Y., & Sumaji, S.; 2021).

Critical thinking is one of the skills of high order thinking skills. As previously stated, this skill will grow and stick well if it is made a habit. Teachers can play a role in forming students' high-level thinking habits through various learning activities in the classroom, one example is by facilitating students to deal with questions that have high order thinking skills (HOTS) categories. These skills can also be developed cumulatively as students progress through their courses and subjects and other experiences that their institutions provide. In addition, engaging their subjects through problem solving, critical thinking and decision-making activities helps students improve their higher order thinking skills (Abosalem, Y.; 2016).

Higher Order Thinking Skills (HOTS) are the ability to think beyond recall, restate, or refer without processing (recite). HOTS in the context of assessment measures the ability to: 1) Transfer one concept to another, 2) Processing and applying information 3) Find links between different information, 4) Use information to solve problems, and 5) Critically examine ideas and information. (Ismafitri, R., Alfan, M., & Kusumaningrum, S. R.; 2022).

Specifically, Brookhart (2010) states that HOTS-based tests are tests that are directed so that students are able to respond or answer test items: a. transfer one concept to another; b. processing and applying information; c. look for connections between different pieces of information; d. use information to solve problems; e. critically examine ideas and information.

The characteristics of the test imply that the development of HOTS-based tests cannot be done haphazardly, but requires careful planning by taking into account important indicators of higher order thinking skills. Indicators of higher order thinking skills are: 1. **Transfer of learning**. Transfer of learning is the ability of learners to utilize previously learned knowledge and skills to be applied to solve problems in new learning contexts or situations. 2. **Critical thinking**. Critical thinking refers to the intellectual process possessed by learners to actively and skillfully conceptualize, apply, analyze, synthesize and evaluate various information gathered through observation, experience, reflection, reasoning and communication. 3. **Problem solving.** Problem-solving skills refer to learners' ability to solve problems effectively, and in a timely manner. It involves learners' ability to identify and define problems, generate alternative solutions, evaluate and select reasonable alternative solutions to a particular problem. 4. **Logical thinking.** Just like critical thinking, logical thinking also requires reasoning skills to study a problem objectively, which enables learners to make rational conclusions.

From the characters and indicators described above, of course this type of HOTS question will be able to train students to think critically and creatively.

▪ METHOD

This research is a type of research & development. The development of HOTS based reaction rate question instrument first follows the following steps: (Ministry of Education and Culture; 2019)

1. Determine the basic competencies and materials to be assessed.
   Researchers analyzed cognitive processes, knowledge dimensions, and materials on basic competencies in the curriculum that allow higher order thinking skills questions to be made.
2. Developing a grid.
   Researchers ensure that all components contained in the grid are consistent, aligned, and can be made into higher order thinking skills questions.
3. Formulate question indicators.
   To produce questions that measure higher order thinking skills, the formulation of indicators needs to meet the principles of assessment of these skills, namely the need for stimulus, new contexts, and higher order thinking processes. Researchers use stimuli that are related to real life everyday and in accordance with the level of cognitive development of students. Contextual stimulus will make it easier for learners to transfer things that have been learned so that positive attitudes arise and appreciate things that have been learned. Stimulus with a context that is not in accordance with the development of learners will be difficult to digest so that it does not support the development of higher order thinking skills.
4. Writing questions according to the rules of writing questions.
   To ensure the quality of questions so that they provide valid information, questions need to fulfill the rules of question writing from the aspects of construction, substance, and language.

The development of HOTS questions in accordance with the steps mentioned above can be implemented using the Rasch model. One of the most important uses of the Rasch model in educational research is to help guide the construction of test instruments and evaluate their functionality. The construction of measurement instruments with Rasch models is "a systematic process in which items are deliberately constructed according to theory and empirically tested through Rasch models to produce a set of items that define a linear measurement scale." Liu (2020) includes the following steps

1. Define constructs that can be characterized by linear attributes.
2. Identify behaviors that correspond to different levels of the specified construct.
3. Determine the result space of the behavior (set of items).
4. Field test with a representative sample of the target population. The test was conducted on 36 students of XI MIPA class of Eka Prasetya Medan High School, on Monday, April 17, 2023 using the google form link.
5. Perform Rasch modeling.
   Since the question developed is in the form of multiple choice with a score of 1 for the correct answer and 0 for the wrong answer, the dichotomous Rasch model is used.

The next step is to analyze the question items based on the answer response patterns given by students using the Rasch model assisted by the WINSTEPS 3.73 application. Based on the criteria of Boone et al. (2014), the criteria used to check the suitability of items and individuals who do not fit (outliers or misfits) are: the accepted Outfit mean square (MNSQ) value: 0.5 < MNSQ < 1.5; ZSTD value (-2.0 < ZSTD < + 2.0), and Point Measure Corrlation (Pt Mean Corr) value is 0.4 < Pt. Measure Corr < 0.85. In addition, the detection of biased questions through the gender variable (male and female) of the sample in the study. An item is said to be biased if the probability of the item is below 5% (0.05). However, in this study the results of student answers were analyzed based on eight (8) types of output tables.

6. Review item fit statistics and revise items if it is necessary
7. Review Wright map results.
   The Wright map provides an overview of multiple choice test items by placing the difficulty of the test items on the same scale of measurement as the candidates' abilities. It provides users with a comparison of candidates and items, to better understand how precisely the test is measured, so that items can be added or removed. (Andrich, D.; 2010).
8. Repeat step (4) through (7) until a set of items is Rasch-compliant.
9. Establish validity and reliability claims for measurement instruments.
10. Develop documentation for measurements in the instrument. After all these stages are completed, the question instrument on HOTS-based Reaction Rate material is ready for use.

▪ RESULT AND DISCUSSION

The basic competencies and materials assessed in the development of this instrument are Reaction Rate material with:
3.6 Explaining factors affecting reaction rate using collision theory; and
3.7. Determine the reaction order and reaction rate constant based on experimental data.

The preparation of question grids is consistent, aligned, and can be made into higher order thinking skills questions. There are 11 indicators of questions from both basic competencies, accompanied by stimuli related to real life and in accordance with the level of cognitive development of students. Furthermore, 20 multiple-choice HOTS questions were developed. To ensure the quality of the questions so that they provide valid information, the questions need to fulfill the rules of writing questions from the aspects of material, construction, substance, and language.

Validation was conducted by 4 expert validators (2 chemistry lecturers and 2 chemistry teachers). The results of validation from experts were then evaluated using Aiken's formula. Aiken (1985) formulated Aiken's V formula to calculate the content validity coefficient which is based on the results of the assessment of an expert panel of n people on an item in terms of the extent to which the item represents the measured construct. The formula proposed by Aiken is as follows: $V = \sum s / [n(c-1)]$ (in Hendryadi; 2014) S = r - lo Lo = lowest validity score C = highest validity rating number R = the number given by the rater The item was assessed by 4 raters with 4 scale options of 4, then if we refer to the table the minimum acceptable V value with a 5% error rate is 0.92.

The results of the validation analysis are presented in table 1.

Table 1. Validation Analysis (Aiken's V)

No. item	V	Description
1	0,94	Valid
2	0,96	Valid
3	0,96	Valid
4	0,96	Valid
5	0,96	Valid
6	0,96	Valid
7	0,94	Valid
8	0,96	Valid
9	0,97	Valid
10	0,95	Valid
11	0,95	Valid
12	0,94	Valid
13	0,94	Valid
14	0,94	Valid
15	0,97	Valid
16	0,97	Valid
17	0,92	Valid
18	0,93	Valid
19	0,95	Valid
20	0,97	Valid

Since all questions have been declared valid, a field test was conducted with a representative sample of the target population. The test was conducted on 36 students of class XI MIPA SMA Eka Prasetya Medan, on Monday, April 17, 2023 using the google form link. The results of student answers were then analyzed using Rasch modeling with the following results:

1. Summary statistics

Summary statistics is a part of descriptive statistics that summarizes and provides the gist of information about the sample data. Statisticians commonly try to describe and characterize the observations by finding: a measure of location, or central tendency, such as the arithmetic mean. Summary statistics from this data can be seen in Figure 1 below.

```
     SUMMARY OF 36 MEASURED (EXTREME AND NON-EXTREME) Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      12.9       20.0         .87     .61                                |
| S.D.       4.3         .0        1.33     .24                                |
| MAX.      20.0       20.0        4.54    1.84                                |
| MIN.       5.0       20.0       -1.25     .48        .65   -2.2    .30   -1.9 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .68 TRUE SD    1.14  SEPARATION  1.69  Person RELIABILITY  .74 |
|MODEL RMSE    .66 TRUE SD    1.15  SEPARATION  1.76  Person RELIABILITY  .76 |
| S.E. OF Person MEAN = .22                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .97
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .81

     SUMMARY OF 20 MEASURED (NON-EXTREME) Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      23.2       36.0         .00     .41       1.01     .1    .94    -.1 |
| S.D.       5.0         .0         .80     .02        .14     .8    .26     .8 |
| MAX.      28.0       36.0        1.95     .44       1.32    1.6   1.61    2.0 |
| MIN.      11.0       36.0        -.80     .39        .77   -1.4    .67   -1.3 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .43 TRUE SD     .67  SEPARATION  1.58  Item   RELIABILITY  .71 |
|MODEL RMSE    .42 TRUE SD     .68  SEPARATION  1.64  Item   RELIABILITY  .73 |
| S.E. OF Item MEAN = .18                                                     |
-------------------------------------------------------------------------------
```

Figure 1. Summary statistics

From the summary statistics results we can see three (3) types of reliability, person reliability, item reliability and Cronbach Alpha with the following conditions whose categories we can see in the following tables 2 and 3.

Table 2. Person reliability and Item reliability categories

No	Person reliability and Item reliability values	Category
1	< 0,67	Weak
2	0,67 - 0,80	Simply
3	0,80 - 0,90	Good
4	0,91 - 0,94	Very good
5	>0,94	Special

As for the provisions of Cronbach Alpha, we can see in table 3.

Table 3. Categories of Cronbach Alpha Value (reliability) of question items

No	Cronbach Alpha value (reliability)	Category
1	< 0,50	Bad
2	0,50 - 0,60	Ugly
3	0,60 - 0,70	Simply
4	0,70 - 0,80	Good
5	>0,80	Very good

Based on these criteria, we can state that :

a. Person Reliability 0.74 and 0.76 categories are **simply**
b. Item Reliability 0.71 and 0.73, **simply** category
c. Cronbach Alpha Reliability 0.81, **very good** category

2. Item Fit Order

From the results shown by the Item Fit Order, we can see the level of item suitability (validity) which is used to explain whether the items function normally to make measurements or not. According to Boone et al (2014), the criterion used to check the suitability of outlier or misfit items is :

a. Accepted Outfit mean square (MNSQ) value: 0.5 < MNSQ < 1.5
b. Accepted Z-standardized Outfit (ZSTD) values: -2.0 < ZSTD < + 2.0
c. Accepted Point Measure Correlation (Pt Measure Corr) value: 0.4 < Pt Measure Corr < 0.85.

Measurement of item fit order results are shown in the figure 2.

```
        Item STATISTICS:  MISFIT ORDER

-------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|------------------------------------+----------+----------+-----------+-----------+------|
|    17     14     36     1.44    .40|1.32   1.6|1.61   2.0|A .29   .52| 71.4  73.7| IT17 |
|     6     21     36      .37    .39|1.19   1.1|1.47   1.7|B .35   .49| 71.4  71.7| IT6  |
|     7     23     36      .06    .40|1.25   1.4|1.28   1.0|C .31   .47| 60.0  72.4| IT7  |
|    18     11     36     1.95    .42|1.22   1.1|1.21    .7|D .38   .51| 62.9  75.8| IT18 |
|    16     27     36     -.61    .43|1.12    .6| .93    .0|E .36   .41| 65.7  76.5| IT16 |
|     8     23     36      .06    .40|1.06    .4|1.10    .4|F .43   .47| 65.7  72.4| IT8  |
|    20     27     36     -.61    .43|1.05    .3| .91   -.1|G .40   .41| 77.1  76.5| 1T20 |
|    14     25     36     -.26    .41|1.02    .2| .95    .0|H .44   .44| 74.3  74.1| IT14 |
|    11     26     36     -.43    .42|1.00    .1| .96    .0|I .43   .43| 77.1  75.1| IT11 |
|     5     22     36      .22    .39| .98   -.1| .88   -.4|J .51   .48| 74.3  71.7| IT5  |
|    13     28     36     -.80    .44| .98    .0| .73   -.5|j .44   .39| 77.1  78.4| IT13 |
|    10     28     36     -.80    .44| .96   -.1| .78   -.3|i .43   .39| 82.9  78.4| IT10 |
|     1     28     36     -.80    .44| .95   -.2| .67   -.6|h .46   .39| 77.1  78.4| IT1  |
|    15     28     36     -.80    .44| .94   -.2| .77   -.4|g .44   .39| 77.1  78.4| IT15 |
|     3     28     36     -.80    .44| .92   -.3| .67   -.6|f .47   .39| 77.1  78.4| IT3  |
|     2     18     36      .83    .39| .89   -.6| .82   -.8|e .59   .51| 74.3  72.6| IT2  |
|    19     16     36     1.13    .39| .88   -.7| .78   -.9|d .61   .52| 77.1  73.1| IT19 |
|     4     27     36     -.61    .43| .87   -.6| .71   -.6|c .50   .41| 82.9  76.5| IT4  |
|    12     23     36      .06    .40| .83  -1.0| .77   -.8|b .58   .47| 82.9  72.4| IT12 |
|     9     21     36      .37    .39| .77  -1.4| .70  -1.3|a .64   .49| 82.9  71.7| IT9  |
|------------------------------------+----------+----------+-----------+-----------+------|
| MEAN    23.2   36.0      .00    .41|1.01    .1| .94   -.1|           | 74.6  74.9|      |
| S.D.     5.0     .0      .80    .02| .14    .8| .26    .8|           |  6.5   2.5|      |
-------------------------------------------------------------------------------------
```

Figure 2. Item Fit Order

In accordance with the criteria put forward by Boone et al, we can see in the table 4, some question items that do not meet the criteria.

Table 4. Interpretation of Fit Order Items

No. Item	MNSQ	ZSTD	Pt. Measure Corr
17	Does not meet	Meet	Does not meet
6	Meet	Meet	Does not meet
7	Meet	Meet	Does not meet
18	Meet	Meet	Does not meet
16	Meet	Meet	Does not meet

Based on the results obtained, all question items are declared reliable, these questions have appropriate consistency. Based on the interpretation of the Item Fit Order results, several question items were found to be less in accordance with the criteria, however, if there is one item where the MNSQ and Pt Measure Corr values do not meet the criteria but the ZSTD value meets the criteria then the item is still considered fit, meaning that the item can still be maintained. However, since the questions developed are HOTS-based, it would be better if question item number 17 is replaced. This is also almost in accordance with the analysis value in Aiken's formula, where question number 17 has a value of V = 0.92, which is the lower limit value that can be accepted in the formula.

3. Wright Map

When it comes to presenting the results of dichotomous or polytomous item response models, Wright Maps are commonly used. The WrightMap package offers functions that allow you to easily create these beautiful Wright Maps from item parameters and person estimates stored as R objects as can be seen in Figure 3 below.

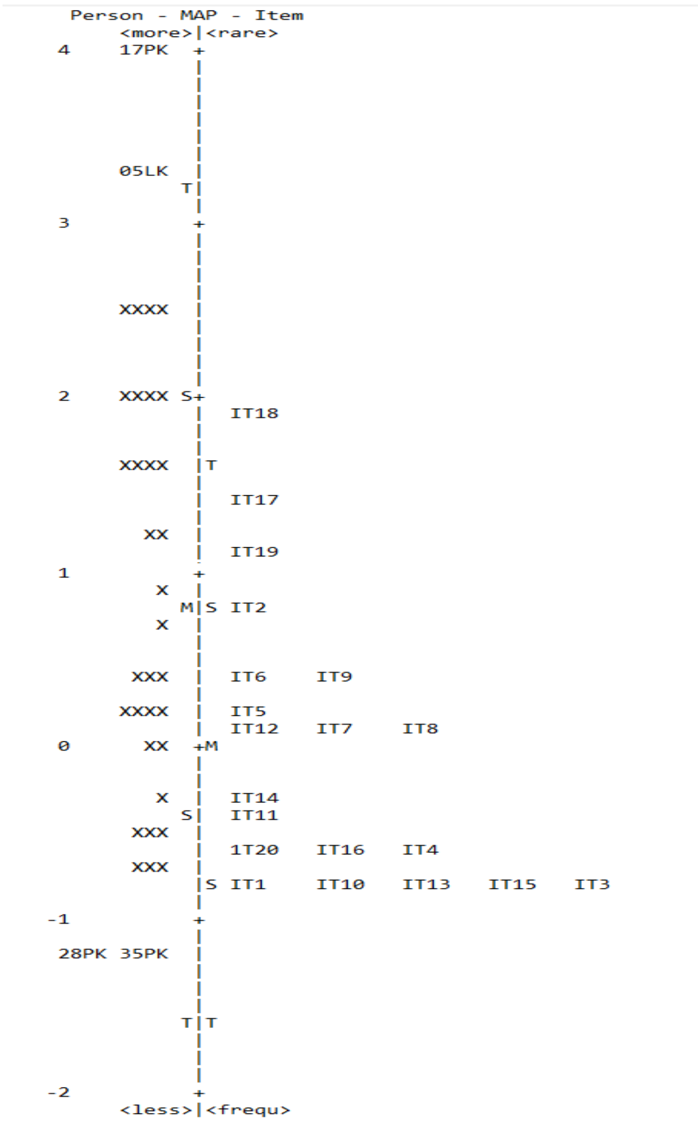

Figure 3. Wright Map

The results shown by the wright map are as follows:

a. The right part of the wright map above is the distribution of item difficulty levels, while the left part is the distribution of student abilities.
b. Problem number 18 is a question that has a high level of difficulty, while question items number 1, 10, 13, 15 and 3 have the same level of difficulty and are easy questions.
c. The highest-ability student is 17PK, while the lowest-ability students are 28PK and 35PK.

Based on the results we obtained from the wright map, all question items are within the standard deviation (T), where there are no questions that are too easy, nor questions that are too difficult. Students 17PK, 05PK, 04LD, 10LK, 14LK and 23PK, have abilities above the difficulty level of question number 18, which means that these students are able to answer question number 18 correctly, and also they still have the ability to answer questions that are more difficult than question number 18.

4. Item Measure (level of item difficulty)

To find out the level of difficulty of the items (item measure), it can be seen from the logit value of each item in the measure column. A high logit value indicates the highest level of difficulty of the question as can be seen in Table 5 below.

Table 5. Categories of question groups based on difficulty level

No.	Logit Value	Category
1	>+1.37 SD	Very difficult
2	0.0 logit +1.37 SD	Difficult
3	0.0 logit -1.37 SD	Medium
4	< -1.37 SD	Easy

Based on table 5, the question items can be categorized as follows:

Table 6. Difficulty Level of Questions

No	Score	Count	Measure	Category
18	11	36	1.95	Very difficult
17	14	36	1.44	Very difficult
19	16	36	1.13	Difficult
2	18	36	0.83	Difficult
6	21	36	0.37	Difficult
9	21	36	0.37	Difficult
5	22	36	0.22	Difficult
7	23	36	0.06	Difficult
8	23	36	0.06	Difficult
12	23	36	0.06	Difficult
14	25	36	-0.26	Medium
11	26	36	-0.43	Medium
4	27	36	-0.61	Medium
16	27	36	-0.61	Medium
20	27	36	-0.61	Medium
1	28	36	-0.80	Medium
3	28	36	-0.80	Medium
10	28	36	-0.80	Medium
13	28	36	-0.80	Medium
15	28	36	-0.80	Medium

The item measure shown in table 6 relates to the wright map discussed earlier. The level of difficulty of the items (item measure) is sorted from high to low difficulty, from very difficult to easy. Item number 18 is an item that is very difficult for students to do, this can also be seen from the number of students who can answer the question correctly as many as 11 people (30.56%). The same logit value for each item shows that the difficulty level of each question is not much different.

5. Person Measure (student ability)

This person measure is used to identify the level of student ability in answering questions. Students' ability levels are ranked from highest to lowest based on the logit value of each person. The standard deviation (SD) value in this person measure is 1.80. The determination criteria can be seen in table 7.

Table 7. Logit Value of Student Abilities

No	**Logit score of student ability**	**Category**
1	>1,80	High
2	-1,29 – 1,80	Medium
3	< -1,29	Low

Based on the criteria for the student ability category, the results of student grouping are as table 8 follows.

Table 8. Student Abilities Categories

No	Total Score	Total Count	Measure	Category
17	20	20	4,54	High
5	19	20	3,26	High
4	18	20	2,46	High
10	18	20	2,46	High
14	18	20	2,46	High
23	18	20	2,46	High
6	17	20	1,95	High
11	17	20	1,95	High
26	17	20	1,95	High
32	17	20	1,95	High
1	16	20	1,56	Medium
8	16	20	1,56	Medium
13	16	20	1,56	Medium
27	16	20	1,56	Medium
15	15	20	1,23	Medium
33	15	20	1,23	Medium
19	14	20	0,94	Medium
2	13	20	0,68	Medium
7	12	20	0,43	Medium
18	12	20	0,43	Medium
25	12	20	0,43	Medium
9	11	20	0,20	Medium
16	11	20	0,20	Medium
20	11	20	0,20	Medium
29	11	20	0,20	Medium
22	10	20	-0,03	Medium
30	10	20	-0,03	Medium
24	9	20	-0,26	Medium
3	8	20	-0,49	Medium
21	8	20	-0,49	Medium
31	8	20	-0,49	Medium
12	7	20	-0,72	Medium
34	7	20	-0,72	Medium
36	7	20	-0,72	Medium
28	5	20	-1,25	Medium
35	5	20	-1,25	Medium

The logit value shows the student's ability to answer questions. The student who has the highest ability to answer questions is student number 17 with a logit value of 4.54.

The student who has the lowest ability to answer questions is student number 35 with a logit value of -1.25.

6. Person Fit Order

This person fit order is used to detect individuals whose response patterns do not fit (different), meaning that there is a mismatch in the answers given based on their abilities compared to the ideal model. According to Boone et al (2014), the criteria used to see the level of individual suitability (person fit) are:

a. Accepted Outfit mean square (MNSQ) value: 0.5 < MNSQ < 1.5
b. Accepted Z-standardized Outfit (ZSTD) values: -2.0 < ZSTD < + 2.0
c. Accepted Point Measure Correlation (Pt Measure Corr) value: 0.4 < Pt Measure Corr < 0.85 The results of the Person fit order are shown in Figure 4.

Person STATISTICS: MISFIT ORDER

ENTRY	TOTAL	TOTAL		MODEL	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Person
18	12	20	.43	.49	1.53	2.5	1.60	2.5	A-.26	.36	50.0	68.2	18PK
31	8	20	-.49	.48	1.17	1.1	1.46	1.8	B .07	.32	50.0	64.7	31PD
35	5	20	-1.25	.54	1.10	.5	1.46	1.1	C .07	.26	75.0	74.9	35PK
27	16	20	1.56	.60	1.31	.9	1.44	1.0	D-.03	.34	75.0	80.9	27PK
15	15	20	1.23	.55	1.24	.9	1.43	1.2	E .04	.36	75.0	77.2	15LK
1	16	20	1.56	.60	1.32	.9	1.21	.6	F .04	.34	75.0	80.9	01LD
7	12	20	.43	.49	1.25	1.3	1.28	1.3	G .07	.36	60.0	68.2	07LD
29	11	20	.20	.48	1.16	.9	1.25	1.3	H .15	.36	60.0	65.9	29LK
3	8	20	-.49	.48	1.22	1.4	1.13	.6	I .10	.32	40.0	64.7	03LK
25	12	20	.43	.49	1.18	1.0	1.19	.9	J .16	.36	60.0	68.2	25PK
36	7	20	-.72	.49	1.03	.3	1.18	.7	K .23	.31	70.0	65.8	36PD
20	11	20	.20	.48	1.10	.6	1.10	.5	L .25	.36	60.0	65.9	20LK
5	19	20	3.26	1.05	1.09	.4	.95	.4	M .11	.21	95.0	95.0	05LK
33	15	20	1.23	.55	1.08	.4	1.06	.3	N .28	.36	75.0	77.2	33PK
21	8	20	-.49	.48	1.05	.4	1.01	.1	O .27	.32	60.0	64.7	21PK
16	11	20	.20	.48	1.04	.3	1.03	.2	P .32	.36	60.0	65.9	16LK
13	16	20	1.56	.60	1.01	.1	.89	-.1	Q .36	.34	85.0	80.9	13PD
4	18	20	2.46	.78	1.00	.2	.79	.0	R .32	.28	90.0	89.9	04LD
34	7	20	-.72	.49	.98	.0	.89	-.3	q .35	.31	70.0	65.8	34PK
12	7	20	-.72	.49	.94	-.3	.86	-.4	p .40	.31	70.0	65.8	12LK
32	17	20	1.95	.66	.92	.0	.68	-.4	o .46	.32	85.0	84.9	32PK
2	13	20	.68	.50	.91	-.3	.89	-.4	n .47	.37	75.0	71.1	02LK
30	10	20	-.03	.48	.85	-1.0	.87	-.6	m .52	.35	80.0	64.9	30PK
26	17	20	1.95	.66	.86	-.2	.67	-.4	l .50	.32	85.0	84.9	26PK
8	16	20	1.56	.60	.84	-.4	.67	-.6	k .55	.34	85.0	80.9	08LK
28	5	20	-1.25	.54	.83	-.6	.67	-.8	j .50	.26	75.0	74.9	28PK
6	17	20	1.95	.66	.83	-.3	.58	-.6	i .55	.32	85.0	84.9	06LD
24	9	20	-.26	.48	.81	-1.4	.76	-1.2	h .57	.34	75.0	64.6	24PK
9	11	20	.20	.48	.80	-1.3	.77	-1.2	g .60	.36	80.0	65.9	09LK
19	14	20	.94	.52	.78	-.9	.75	-.8	f .62	.36	85.0	74.0	19PK
11	17	20	1.95	.66	.75	-.5	.58	-.6	e .61	.32	85.0	84.9	11PK
10	18	20	2.46	.78	.72	-.4	.35	-.8	d .64	.28	90.0	89.9	10LK
22	10	20	-.03	.48	.70	-2.2	.67	-1.9	c .71	.35	90.0	64.9	22PD
14	18	20	2.46	.78	.65	-.5	.30	-1.0	b .71	.28	90.0	89.9	14LK
23	18	20	2.46	.78	.65	-.5	.30	-1.0	a .71	.28	90.0	89.9	23PK
MEAN	12.9	20.0	.87	.61	.99	.1	.94	.0			74.6	74.9	
S.D.	4.3	.0	1.33	.24	.21	.9	.33	.9			13.3	9.6	

Figure 4. Person Fit Order

The criteria for checking person fit are similar to the criteria for checking item fit order. In this check, the student response pattern is declared fit if it meets at least two of the three criteria, namely MNSQ, ZSTD and Pt. Measure Corr. If a student response pattern is found that does not fit, it can then be checked on the scalogram.

The Person Fit Order results (figure 4) show that students 10LK, 14LK, 23PK, and 18PK did not meet the MNSQ criteria. Student 18PK does not meet the ZSTD criteria. Students 18PK, 31PD, 35PK, 27PK, 15LK, 01LD, 07LD, 29LK, 03LK, 25PK, 36PD, 20LK, 05LK, 33PK, 21PK, 16LK, 13PD, 04LD, 34PK did not meet the Pt criteria. Measure Corr. From these results, it is stated that 1 student (5%), namely student 18PK, has a response pattern that does not fit.

We can analyze the unfit response pattern of student 18PK on the scalogram results and readjust it to the student's position in the wright map. The response pattern is declared not fit because student 18PK is able to answer difficult questions that are above his ability correctly, but the student cannot answer correctly the easy questions whose difficulty level is below the student's ability.

7. Scalogram

Scalogram developed by Louis Guttman. Guttman introduced the ranking of attitude scales from lowest to highest, which was developed into a specific matrix. The results of this research scalogram are shown in Figure 5.

```
Person |Item
       |   111 1211   1     111
       |13035460147825692978
       |--------------------
    17 +11111111111111111111  17PK
     5 +11111111111111011111  05LK
     4 +11111111111011111101  04LD
    10 +11111111111111111010  10LK
    14 +11111111111111111100  14LK
    23 +11111111111111111100  23PK
     6 +11111111111111100110  06LD
    11 +11111111110111111100  11PK
    26 +11111111111101110110  26PK
    32 +11111111110111111001  32PK
     1 +11111111011010011111  01LD
     8 +11111111110111011010  08LK
    13 +11111111101110011110  13PD
    27 +11111110001111111011  27PK
    15 +11110001111111110101  15LK
    33 +11011111110111010101  33PK
    19 +11111011111110101000  19PK
     2 +11101111101101100100  02LK
     7 +00110110110011111100  07LD
    18 +10101100000111111101  18PK
    25 +00111011111000111010  25PK
     9 +11110110111011000000  09LK
    16 +11001110110011010100  16LK
    20 +11011011010010101010  20LK
    29 +11001101101100110001  29LK
    22 +01111111111000000000  22PD
    30 +11111011101000000010  30PK
    24 +01111101011100000000  24PK
     3 +00100100100111110000  03LK
    21 +01010101011100001000  21PK
    31 +00011001110100100001  31PD
    12 +10101101000001100000  12LK
    34 +11100010000111000000  34PK
    36 +10110010011000000010  36PD
    28 +11001010100000000000  28PK
    35 +00010011001000000010  35PK
       |--------------------
       |   111 1211   1     111
       |13035460147825692978
```

Figure 5. Scalogram

From Figure 5, we can see that each item has an order that can be systematically ranked from low to high and also sorted from the easiest to the most difficult. The goal is to analyze, provide explanations and predict at the same time individual abilities as well as the level of difficulty of the items (Sumintono, B., & Widhiarso, W.; 2015).

In addition, from the scalogram results we can also identify if there is cheating, for example students cheating. The first thing we can see is if there is the same person logit value, then we can review the pattern of the student's answers.

8. Item DIF

One of the characteristics of a valid instrument measurement is that the items do not contain bias. A question item is called biased if it is found that one individual with certain characteristics is more favored than individuals with other characteristics. To check whether the question items in the study can be seen from Figure 6.

Person CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
2	1.5799	1	.2088	.6498	.1875	1	IT1
2	.0006	1	.9806	.0001	-1.5378	2	IT2
2	1.5799	1	.2088	.6498	.1875	3	IT3
2	.0027	1	.9586	.0017	-1.3973	4	IT4
2	2.6760	1	.1019	1.1686	.5845	5	IT5
2	.4086	1	.5227	.1559	-.5083	6	IT6
2	.5672	1	.4514	.2184	-.3723	7	IT7
2	3.2840	1	.0700	1.4848	.7702	8	IT8
2	.4086	1	.5227	.1559	-.5083	9	IT9
2	.5154	1	.4728	.1969	-.4158	10	IT10
2	.0241	1	.8765	.0085	-1.2169	11	IT11
2	.0237	1	.8776	.0090	-1.2095	12	IT12
1	.0000	0	1.0000	.0000	.0000	13	IT13
2	1.4994	1	.2208	.6372	.1756	14	IT14
2	.0861	1	.7692	.0329	-.9704	15	IT15
2	.8060	1	.3693	.3188	-.2008	16	IT16
2	.6664	1	.4143	.2713	-.2766	17	IT17
2	.3558	1	.5508	.1464	-.5320	18	IT18
2	1.5373	1	.2150	.6458	.1837	19	IT19
2	.0027	1	.9586	.0017	-1.3973	20	IT20

Figure 6. Item Person DIF

A question item is detected as DIF if it has a probability value of less than 5%. The analysis results from Figure 6 show that there is no probability value less than 0.05 (5%), which means that it can be concluded that the DIF class has no bias.

▪ CONCLUSION

Today, the Rasch measurement model is unquestionably the gold standard for psychometric assessments of outcome scales in the modern era. Performing Rasch analysis offers a strong technique for bringing together important topics like unidimensionality, category ordering, and DIF within the framework of measurement science, whether building a new scale or analyzing and updating current ones. It is crucial to keep in mind that all ordinal scales are nonlinear, and even in cases when the data fit the Rasch model, the raw score stays nonlinear.

Since the Rasch model defines measurement, it offers a model for the proper pattern of responses in the event that a unidimensional scale needs to be built and an ordinal score needs to be converted linearly. This Rasch model can be used to create non-test instruments in addition to multiple choice question instruments. We will build non-test instruments that assess students' critical thinking abilities in order to undertake additional study.

▪ REFERENCES

Andrich, D. (2010). *Rasch models for measurement*. Sage.

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.

Boone, W. J., Staver, J. R., Yale, M. S., Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Person reliability, item reliability, and more. *Rasch analysis in the human sciences*, 217-234.

Boone, W. J., Staver, J. R., Yale, M. S., Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Understanding person measures. *Rasch analysis in the human sciences*, 69-92.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Ascd.

Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, I., & Nurhaswinda, N. (2021). Analysis of RASCH model for the validation of chemistry national exam instruments. *Jurnal Pendidikan Sains Indonesia*, *9*(3), 329-345.

Fahreza, R., Yudha, S., Purwanto, R., Suyanti, R. D., & Darmana, A. (2022). Used of RASCH Model for Analysis of Students' Critical Thinking Skills Test Instruments on Thermochemical Topics. *Jurnal Pendidikan dan Pembelajaran Kimia*, *11*(2).

Harmurni, L. (2019). *Instrumen Penilaian & Validasinya*. Uwais Inspirasi Indonesia.

Hartik, S., Utaminingsih, S., & Madjdi, A. H. (2021, March). A Need Assessment of Integrated Science Teaching Material Based Higher Order Thinking Skills (HOTS). In *Journal of Physics: Conference Series* (Vol. 1823, No. 1, p. 012078). IOP Publishing.

Hendryadi, H. (2014). Content Validity (Validitas isi). *International Encyclopedia of the Social & Behavioral Sciences:*, *1*, 774-777.

Kemendikbud, P. P. P., & Abduh, M. (2019). Panduan penulisan soal HOTS-higher order thinking skills.

Laliyo, L. A. R., Sumintono, B., & Panigoro, C. (2022). Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in Rasch model. *Heliyon*, *8*(3), e09126.

Liu, X. (2020). *Using and developing measurement instruments in science education: a Rasch modeling approach 2nd edition*. IAP.

PUTRA, B. R. PENGEMBANGAN BUKU TEKS BERMUATAN HIGHER ORDER THINKING SKILLS (HOTS) PADA MATERI LAJU REAKSI KIMIA.

Risdiana, A., Erna, M., & Holiwarni, B. (2022). Pengembangan Soal HOTS (Higher Order Thinking Skills) pada Materi Asam-Basa untuk Kelas XI SMA/MA Sederajat. *Jurnal Inovasi Pendidikan Kimia*, *16*(2), 111-117.

Rodríguez-Mora, F., Cebrián-Robles, D., & Blanco-López, Á. (2021). An assessment using Rubrics and the Rasch Model of 14/15-year-old students’ difficulties in arguing about bottled water consumption. *Research in Science Education*, 1-17.

Shahat, M. A., Boone, W. J., Ambusaidi, A. K., Al Bahri, K., & Ohle-Peters, A. (2022). Use of Rasch Analysis to Develop an Arabic Language Survey (STPLTS) to Measure Omani Science Teachers' Views towards the Classroom Application of Pedagogical Learning Theories. *Journal of Baltic Science Education*, *21*(3), 513-527.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.

Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, *16*(2), 92-96.

Werner, S. M., Chen, Y., & Stieff, M. (2021). Examining the psychometric properties of the chemistry self-concept inventory using rasch modeling. *Journal of Chemical Education*, *98*(11), 3412-3420.