Vol 13 (1), 2024, 1-16 DOI: 10.23960/jppk.v13.i1.2024.01
Jurnal Pendidikan dan Pembelajaran Kimia
e-ISSN: 2714-9595| p-ISSN 2302-1772
http://jurnal.fkip.unila.ac.id/index.php/JPK/index

Developing the Test Items Based on High Order Thinking Skills in Chemical Equilibrium Materials Using Rasch Model

Ade Ariyani [1], Ratu Evina Dibyantini [2], and Ayi Darmana [3]

[1, 2, 3] Chemistry Education, Magister Program, State University of Medan
Jalan Willem Iskandar Medan Estate, North Sumatra, Indonesia.

*Correspondinge-mail: adeariyani2@gmail.com

Received: December 20th, 2023 Accepted: January 17th, 2024 Online Published: April 19th, 2024

Abstract: Development the Test Items Based on High Order Thinking Skills in Chemical Equilibrium Materials Using Rasch Model. Objectives: This study aims to develop HOTS (High Order Thinking Skill) based items on chemical equilibrium material using the Rasch model. The type of research used is research and development through the ADDIE model. Respondents in the study were 25 students of class XII at the Superior High School in North Sumatra, distinguished by gender and grade X. The test items developed consisted of 10 questions that had gone through the validation stage by experts and then tested and analyzed using the Winsteps software. Content validation obtained 100% fit indexed items; the reliability of the person category is sufficient and the item is in the good category with Cronbach α=0.72 (good); the value of separation person (1.27) is in the weak category and item (2.49) is in the good category; Item difficulty levels are sorted from very easy: medium: difficult: very difficult. Respondents with codes 02LO and 24LO (logit +4.36) are respondents with the highest ability, codes 18PR and 19PR (logit -1.97) are respondents with the lowest ability. Guttman Matrix Scalogram, two respondents have the potential to guess the answer and four respondents have the potential to commit fraud.

Keywords: HOTS Items, Chemical Equilibrium, Rasch Model.

Abstrak: Pengembangan Item Tes Berbasis High Order Thinking Skill pada Materi Kesetimbangan Kimia Menggunakan Permodelan Rasch. Tujuan: *Studi ini bertujuan untuk mengembangkan item berbasis HOTS (High Order Thinking Skill) pada materi kesetimbangan kimia menggunakan permodelan Rasch. Jenis penelitian yang digunakan adalah penelitian dan pengembangan melalui model ADDIE. Responden dalam penelitian adalah 25 siswa kelas XII Sekolah Menengah Atas Unggulan di Sumatera Utara yang dibedakan ebrdasarkan gender dan asal kelas X. Item tes yang dikembangkan sebanyak 10 soal telah melalui tahap validasi oleh ahli kemudian diujikan dan dianalisis menggunakan software Winsteps. Validasi isi diperoleh 100% item terindex fit; reliabilitas person kategori cukup dan item dalam kategori baik dnegan Cronbach α=0,72 (baik); nilai saparation person (1.27) kategori lemah dan item (2.49) dalam kategori baik; tingkat kesekaran item diurutkan dari sangat mudah:sedang:sukar:sangat sukar. Responden kode 02LO dan 24LO (logit +4.36) adalah responden dnegan kemampuan tertinggi, kode 18PR dan 19 pr (logit -1.97) adalah responden dengan kemampuan terendah. Scalogram Matriks Guttman, dua responden berpotensi menebak jawaban serta empat responden berpotensi melakukan kecurangan.*

Kata kunci: *Item HOTS, Kesetimbangan Kimia, Rasch Model.*

▪ INTRODUCTION

The 21st century, marked by the industrial revolution 4.0, is considered a century of knowledge full of competition in all fields, including education (Sutiani et al., 2022). This century requires students to think critically in solving problems, communication and collaboration skills, the ability to create and update, information and communication technology literacy, contextual learning skills, and information skills in media literacy (BSNP, 2020). Higher order thinking skills or High Order Thinking Skills are the transfer of knowledge, HOTS as critical-creative thinking, and HOTS as problem solving (Hirza et al., 2023). This ability also allows students to apply new knowledge in new contexts, which allows them to build values and knowledge across fields (Kim How et al., 2022). Thus, the ability to think at a higher level is an indication of the success of increasing human resources in the field of education. However, students' high-level thinking skills in Indonesia are still relatively low. This is evidenced by the results of a 2018 International Institute study from the Program for International Student Assessment (PISA) which showed that scientific literacy achieved by students in Indonesia was ranked 67th out of 81 participating countries. This ranking increased from the 2018 PISA results, namely that the level of scientific literacy of students in Indonesia was ranked 71st out of 79 participating countries (OECD, 2023). The results of this research show that the HOTS abilities of students in Indonesia have increased, but are still relatively low. So efforts are needed to improve the quality of subsequent PISA learning to show significant improvements. One thing that can be done is to develop HOTS questions based on scientific literacy and familiarize students with working on questions with this type of HOTS.

Exams that are held such as college entrance exams contain questions based on HOTS because the HOTS concept is an important part of the 2013 Curriculum in Indonesia and higher-order thinking skills are an important thing in the educational process (Harta et al., 2020). One way to facilitate students' higher-order thinking skills is through the use of HOTS questions. HOTS questions are used as parameters to measure higher-order thinking skills (Lubis et al., 2021). Teachers also need to prepare good tests to measure students' HOTS abilities (Widyaningsih et al., 2021). Based on the results of interviews with chemistry teachers at the Superior High School in Deli Serdang, it was revealed that the initial abilities of Class XII students were relatively good, but there were still scores that were below average. In chemistry lessons, teachers have developed HOTS questions and tested these questions in the Mid-Semester and End-Semester Assessments, although the number of items given is not many and the intensity is not too frequent. This is due to the limitations of the teacher's skills in developing HOTS-based questions and the unavailability of a valid HOTS question bank owned by the school (Hirza et al., 2023).

Thus, the researcher will develop and create a High Order Thinking Skill tool using the Rasch modeling. Rasch modeling can be used to evaluate students' scientific abilities by assessing their performance on chemistry tests. By analyzing the difficulty of the problems and the estimates of students' ability obtained from this model, it becomes possible to measure the level of students' scientific understanding of chemical concepts, find areas that need to be improved, and choose appropriate educational interventions (Bond et al., 2020). Research conducted by Darmana states that the Rasch model can very well analyze the quality of the national exam instrument (Darmana et al., 2021). The researcher hopes that this instrument will help teachers design HOTS-based item items that will be tested on students in analyzing how much students are capable.

▪ METHOD

Research design

This research focuses on development through the ADDIE model (Analysis, Design, Development, Implementation, and Evaluation) which is described in Figure 1 (Syathroh & Rizkiani, 2023). In the ADDIE model, the analysis phase is defining the problem, finding the problem, and determining the best solution (Shakeel et al., 2022). In this stage, the problem analysis was carried out, namely the students' low HOTS ability and finding solutions to develop HOTS-based test items. At the design stage, mapping of basic competencies and indicators is carried out and the preparation of a grid of questions that will be developed. At the development stage, multiple choice tests are developed based on indicators and validation of the contents of the test items by experts. The implementation phase is testing questions to respondents then the implementation results are analyzed using the Rasch modeling. The results of the analysis using the Rasch modeling are used as the Evaluation stage, if a question is found that does not fit, it is an indication that students have misconceptions (Dalimunthe & Tarigan, 2023). This information is very useful for teachers to improve the quality of their teaching (Sumintono & Wahyu Widhiarso, 2015).

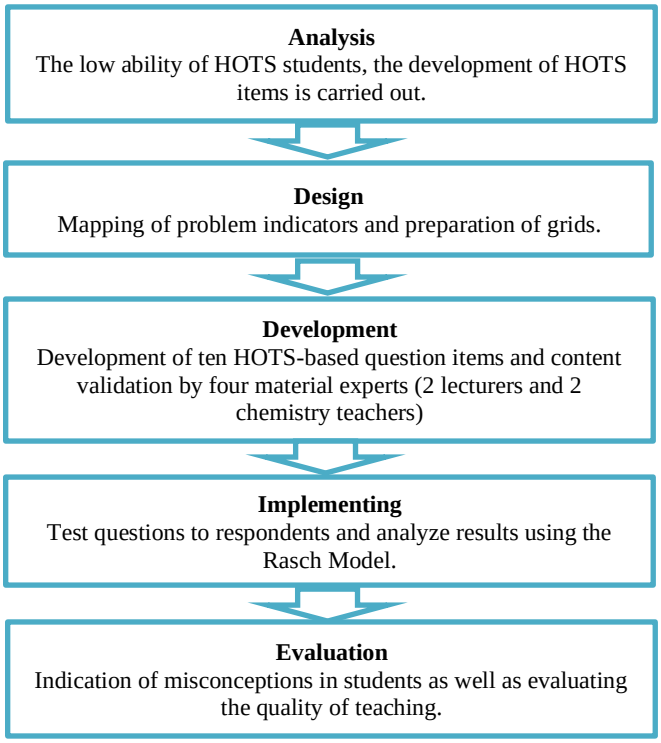

Figure 1. Steps in research design

Samples and Locations

The participants in this study were 25 class XII students of the Superior High School in Manunggal Village, Labuhan Deli District, Deli Serdang Regency, North Sumatra Province, Indonesia. The map of Labuhan Deli District can be seen in Figure 2. The sample consisted of 8 male students (coded “L”) and 17 female students (coded “P”). When data was collected in May for the 2022/2023 School Year, their average age was 17 years.

Data Analysis

In this study, Rasch modeling analysis was used to analyze the data. The Rasch instrument includes item and individual measurements, level of difficulty and discrimination, and outliers or discrepancies in question items (Hale et al., 2016). In science assessment, the analysis of the Rasch model is a probabilistic model used to calibrate the respondent's ability and item difficulty placed on the same interval logit scale. It has been proven that this model can produce evidence of the technical quality of assessment instruments (Chi et al., 2022). Ten items were assessed using the WINSTEP scoring version 3.73. Each question is given a value of one (1) for a correct answer, and a score of zero (0) for an incorrect answer. Students are divided into two groups based on differences in gender and class X origin.

Figure 2. Labuhan Deli District, North Sumatra
Source: https://goo.gl/maps/e8ZGTZWD5hbcHEbj7

Grouping Students Based on Gender

Students are classified based on their gender. Male students are given the code "L", while female students are given the code "P".

Table 1. Distribution of Students Based on Gender

Gender Code	Description	Amount
L	Male	8 students
P	Female	17 students
Amount		25 students

Grouping of Students Based on School Origin

Apart from that, students are grouped based on class X. Regular class students are given the code "R", while superior or Olympic class students are given the code "O".

Table 2. Distribution of Students Based on Class X Origin

Class Origin Code	Description	Amount
R	Reguler	5 students
O	Olimpiade	20 students
Amount		25 students

▪ RESULT AND DISCUSSION

HOTS-Based Item Development Stages

Analysis Step

In the analysis step, based on the results of the researcher's evaluation for approximately five years of teaching at school. It was identified that students had low high-level thinking abilities or HOTS. Based on research conducted by Napitupulu and Yusuf in 2022, the HOTS instrument developed can determine students' high-level thinking abilities (Napitupulu & Yusuf, 2022). On this basis, in order to assess students' HOTS abilities, HOTS-based question items will be developed which can be used as a formative instrument for chemical equilibrium material.

Design Steps

In the design step, basic competency, material and indicator mapping was carried out for each question item developed, namely 10 questions, on chemical equilibrium material. The material mapping for each question item is presented in Table 3 below:

Table 3. Question Grid

No.	Question Indicator	Item Number
3.8 Explain equilibrium reactions in terms of the relationship between reactants and reaction products.		
3.8.1	Given an equilibrium graph profile, students can analyze the equilibrium reaction equation correctly based on the graph presented.	1
3.8.2	Given several Kc values, students can evaluate the most appropriate Kc value based on the reaction concentration, reaction coefficient, and ionization degree values presented.	2
	Given the moles added during the reaction, students can properly evaluate the composition of each gas once the new equilibrium is reached.	3
3.8.3	Given a reaction and moles of each substance at equilibrium, as well as the total pressure of the gas, students can analyze the value of the partial pressure of a gas correctly.	4
	Presented a reaction and moles at equilibrium state of one of the substances in the reaction. Students can evaluate the total system pressure value correctly based on the Kp value and the degree of dissociation.	5
3.8.4	Given several reactions along with their Kc values. Students can compare the value of the equilibrium constant K for the reaction in question correctly.	6
3.9 Analyze the factors that influence shifts in the direction of equilibrium and their application in industry.		
3.9.1	Given five reactions which are exothermic and endothermic reactions. Students can analyze reactions that produce a lot of reaction results or products, if the pressure is reduced and the temperature is increased appropriately based on the reaction coefficient and the value of the enthalpy change.	7
	Presented with discourse and reactions, students can determine the shift in equilibrium and the effect of decreasing and increasing reaction pressure correctly based on the reaction coefficient.	8
3.9.2	A table of varying temperatures is presented with data on the equilibrium number of moles of reagents and products. Students can draw conclusions based on the concept of the influence of temperature on the value of K correctly.	9
3.9.2	Equilibrium constant data and an explanation of the Haber Bosch process are provided. Students can determine the correct statement based on the reaction enthalpy change, decomposition enthalpy change, and equilibrium constant.	10

Development Steps

At the development stage, the question items are arranged in multiple choice form with five answer choices, namely A, B, C, D, and E, as shown in Figure 3.

3. CO_2 gas reacts with H_2 gas to form an equilibrium system producing carbon monoxide gas (CO) and water vapor according to the following reaction equation:

CO_2 (g) + H_2 (g) ⇌ CO (g) + H_2O (g)

Gas	Volume	Temperature	Moles in Equilibrium
CO_2 (g)	10 L	T°C	4 mol
H_2 (g)	10 L	T°C	4 mol
CO (g)	10 L	T°C	4 mol
H_2O (g)	10 L	T°C	4 mol

Based on the table above, this system takes place in a container with a volume of 10 L at a temperature ToC, where there are 4 moles of these gases in equilibrium. If at a constant temperature, 4 moles of CO_2 gas and 4 moles of H_2 gas are added, the composition of each gas after reaching the new equilibrium is ....

A. 2 mol
B. 4 mol
C. 6 mol
D. 8 mol
E. 10 mol

Figure 3. Developed Question Items

At this stage, content validation was also carried out by chemists, namely two lecturers and two chemistry teachers. The aspects assessed by expert validators are the material aspect consisting of three statements, the constructive aspect consisting of nine statements, and the linguistic aspect (Suhardi & Budiyaningsih, 2022).

assessment criteria for the three aspects assessed by expert validators regarding the question items developed are on a scale of 1 to 4. Figure 4 below presents the results of validation of HOTS-based question items by experts before and after revisions. From the graph, it can be seen that there is an increase in the quality of the material, construction and language of the question items after revision by experts. Design validation is carried out to evaluate the feasibility of an instrument product based on input from experts. Product validation analysis is carried out quantitatively, with the following V Aiken formula:

$$V = \frac{\sum s}{n(c-1)}$$

Source: (Silviariza et al., 2023).

Information:
s = r – lo
n = number of expert validators
lo = lowest validity rating
c = highest validity rating
r = number given by the appraiser

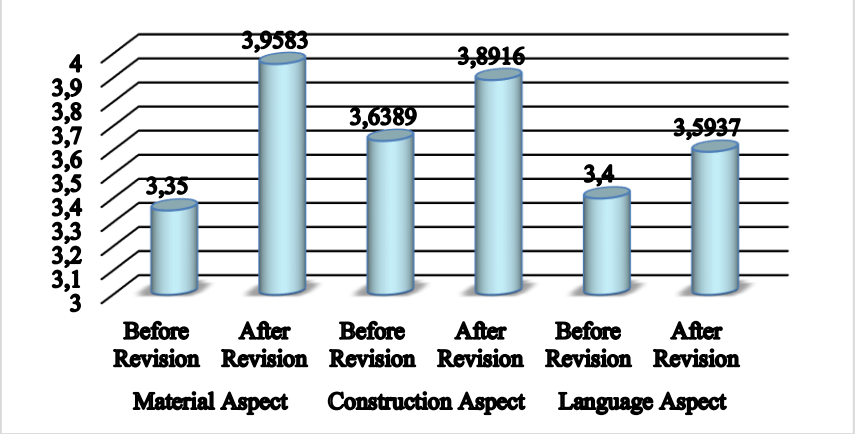

Figure 4. Diagram of Expert Validation Results

Test items are declared valid if their value is equal to or greater than the Aiken table (Astari et al., 2023).With the number of experts = 4 and the n value = 4, the value of the validity coefficient (V) is 0.92, which has a significance of 5%. Therefore, the ten test items are valid, because the validity coefficient (V) value obtained according to the Aiken formula is greater than 0.92. The analysis results are presented in Table 4 below:

Table 4. Validation Results Using V Aiken

Item Number	V Value	Description
1	0.94792	Valid
2	0.9375	Valid
3	0.9375	Valid
4	0.94792	Valid
5	0.92708	Valid
6	0.94792	Valid
7	0.95313	Valid
8	0.9375	Valid
9	0.9375	Valid
10	0.95833	Valid

ImplementingStep

The test items were 10 questions which had gone through the validation stage by experts and undergone a revision stage and had been declared valid, then tested on 25 class XII high school students and then analyzed using Winsteps 3.73 software. The analysis stages using the Rasch model are as follows: 1). Trials are carried out to obtain student answers; 2). Tabulate data into excel; 3). Data conversion using WINSTEP application version 3.73 into interval data that has the same measurement scale; 4). Measuring the effectiveness of the instrument is seen from the validity and reliability values of people and items; 5). Determine item validation, reliability, and level of difficulty; 6). Analyzing the distribution of item difficulty levels and respondents' abilities; 7). Analyze the results of the Scalogram or Guttman Matrix to see the potential for cheating and guessing in answering test items; 8). Item bias analysis based on differences in respondent gender; 9). Item bias analysis based on differences in class X origins of respondents (Sukmawati & Zulherman, 2023).

Analysis Results Using Winsteps 3.73 Software

Validity

The criteria for a fit question are if it meets the following three criteria. Firstly, the item limit is considered appropriate if the MNSQ outfit value is between -0.5 and 1.5. The second condition is that the ZSTD outfit value must be between -2.0 and 2.0. Finally, the point measure correlation value, or the correlation of items with the total score, must be between 0.4 and 0.85 (Yusmaita et al., 2022). Based on the statistical item output analysis: misfit order shows that ten questions meet the three criteria and are said to be fit. The analysis results are presented in table 5 below.

Table 5. Output Item Statistics: Misfit Order

Item	Outfit		Pt Measure Correlation	Category
	MNSQ	ZSTD		
S1	0,45	-1.0	0.63	Fit
S2	0,80	-0.3	0.50	Fit
S3	1,34	0.9	0.46	Fit
S4	0,45	-0.2	0.40	Fit
S5	0,49	-1.0	0.77	Fit
S6	1.25	0.8	0.56	Fit
S7	0,38	-1.6	0.82	Fit
S8	1.35	0.9	0.37	Fit
S9	1,07	0.4	0.58	Fit
S10	1.47	0.8	0.14	Fit
MEAN	**0.91**	**0.00**		

Instrument Reliability

Next, determine the reliability of the instrument by paying attention to the separation of person and separation of items. A reliability index greater than 2 is considered satisfactory (Darmana et al., 2021). The results of person and item reliability are presented in Table 6.

Table 6. Output summary statistics for person and item reliability instruments

Criteria	Parameter	
	Persons (25)	Items (10)
Measure	0.07	0.00
MNSQ	0.91	0.91
ZSTD	0.00	0.00
Reliability	0.69 (Enough)	0.86 (Good)
Separation	1.27	2.49
Cronbach Alpha	0.72 (Good)	

Based on Person Reliability Values

This research uses person reliability values with the following criteria in Table 7:

Table 7. Person / Item Reliability Criteria

No	Reliability Value	Category
1	< 0.67	Weak
2	0.67 – 0.80	Enough
3	0.8 – 0.9	Good
4	0.91 – 0.94	Very Good
5	> 0.94	Special

Source: (Lestari et al., 2020).

Based on the results of the Rasch Model analysis, the person reliability value is 0.69, which shows that the category is sufficient, indicating a relatively small diversity of respondents. This is a common situation because the number of test takers is only 25 and the data type is dichotomous; at least 150 people are needed to demonstrate the stability of data from multiple choice tests.

Based on Item Reliability Value

According to the criteria listed in Table 8, the Reliability item value has a value of 0.86, which indicates that the Reliability item value is in the Good category. Therefore, the stability of the instrument can provide good results in various research environments. Based on Cronbach's Alpha Value Rasch modeling uses person and item separation, as well as Cronbach's reliability index, to demonstrate the reliability of the developed instrument. Person separation indicates how consistently estimates of a student's abilities fit the data, while item separation indicates how reliably the instrument can differentiate between different students based on their abilities (He et al., 2022). Cronbach α value criteria with the following conditions:

Table 8. Cronbach's α value

No	Cronbach's α Value	Categoty
1	< 0.5	Bad
2	0.5 – 0.6	Not Good
3	0.6 – 0.7	Enough
4	0.7 – 0.8	Good
5	> 0.8	Very Good

Source: (Lestari et al., 2020).

The results of the Rasch analysis show that the items in the "Good" category have high reliability, with a Cronbach Alpha value of 0.72. Item reliability indicates the possibility that the results can be repeated in follow-up tests (Hale et al., 2016).

Based on Separation Value

The abilities of the examinees do not vary much, according to the separation person value (1.27), which when rounded together indicates that the data analyzed is not very reliable (the highest value is 2). The item separation value (2.49) is included in the good category, because according to Linacre (2016), the separation index value is considered good if it is greater than 2.0. This shows that each item has a different level of difficulty, ranging from the easiest to the most difficult. However, a narrow research sample cannot be distributed from the most intelligent to the least intelligent (Linacre, 2016). Another equation used to look at grouping more closely is called person and item separation:

$H = \frac{[(4 \text{ x } \text{SEPARATION})+1]}{3}$ Source: (Yusmaita et al., 2022).

Based on Table 7, the separation person value is 1.27. then the value of H is: $H = \frac{[(4 \text{ x } 1.27)+1]}{3} = 2.027$, the number 2,027 is rounded to 2, which means there are two groups of people, which can be interpreted as smart and not smart students, meaning the sample cannot be spread from the most intelligent to the least intelligent. Meanwhile, based on Table 7, the item separation value is 2.49. then the value of H is: $H = \frac{[(4 \text{ x } 2.49)+1]}{3} = 3.653$, the number 3,653 is rounded to 4, which means there are four

groups of items, each classified as very difficult, difficult, medium and very easy. The level of difficulty of each question will be discussed further below.

Item Difficulty Level

Using the Rasch Model, analysis of item difficulty levels is based on the very difficult category with a measuring item value > 1; difficult category ($0.5 \leq b < 1$); medium category ($-0.5 \leq b < 0.5$); easy category ($-1 \leq b < -0.5$), and very easy category ($b \leq -1$) (Darmana et al., 2021). Table 9 below shows the results of the analysis of the difficulty level of the question items.

Table 9. Results of analysis of item difficulty levels

Logit Item	Category	Item Number	%
>1	Very Difficult	9, 5, and 7	30
0.5≤b<1	Difficult	3 and 6	20
-0.5≤b<0.5	Moderate	8 and 2	20
≤ -1	Very Easy	1, 10, and 4	30

Distribution of Student Ability Levels and Difficulty Levels of Question Items

Wright's map, one of the most useful Rasch model results to date, shows that person and item estimates are on the same logit (Oliva & Blanco, 2022). The results of the Wright Map analysis can be seen in Figure 5. Figure 5 shows the distribution of respondents' abilities (left side of the map) and the difficulty of the questions on the right side of the map (Wahyudi et al., 2023). The distribution of student ability levels has a range of logit values between the lowest being below -2 and the highest being a logit of +3.

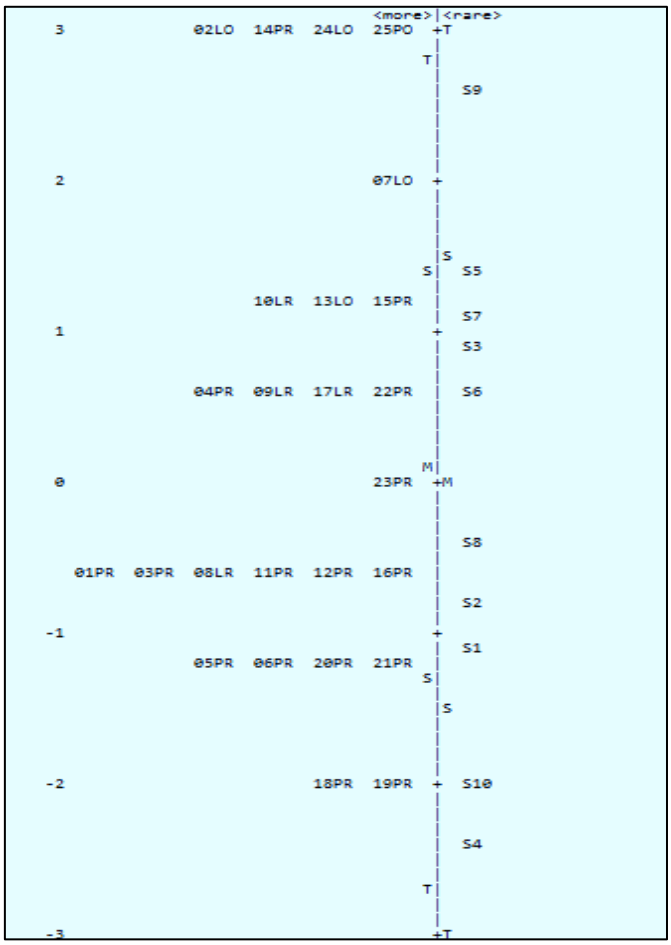

Figure 5. Distribution of Respondents' Ability and Item Difficulty

There were 13 respondents (52%) who had the ability to have a logit value greater than or equal to "0". Meanwhile, 12 other respondents (48%) had logit values below 0. Respondents with codes 02LO and 24LO had logit prices of +4.36 (answered 10 question items correctly), and respondents with codes 14PR and 25PO had logit prices of +2.95 (answered 9 question item correctly) indicates the respondent with the highest ability. Meanwhile, respondents with codes 18PR and 19PR logit -1.97 indicate respondents with the lowest ability (answering 2 questions correctly). In the results of the analysis, no respondents were found in the outlier category or respondents who were below T (twice the standard deviation) (Liu & Jurich, 2023). The difficulty level of the items can be seen on the right of the diagram, item S9 is a question in the very difficult category with a logit of +2.59; while item S4 is at the bottom of the Wright Map and is a question with a very easy level of difficulty with a logit of -2.36. If you look at each HOTS-based item developed, it has a different logit value, so it can be said that these items can measure each respondent's ability well (Siregar et al., 2022).

Scalogram or Guttman Matrix

Figure 6 below shows the results of the Scalogram or Guttman Matrix from the Rasch model analysis carried out:

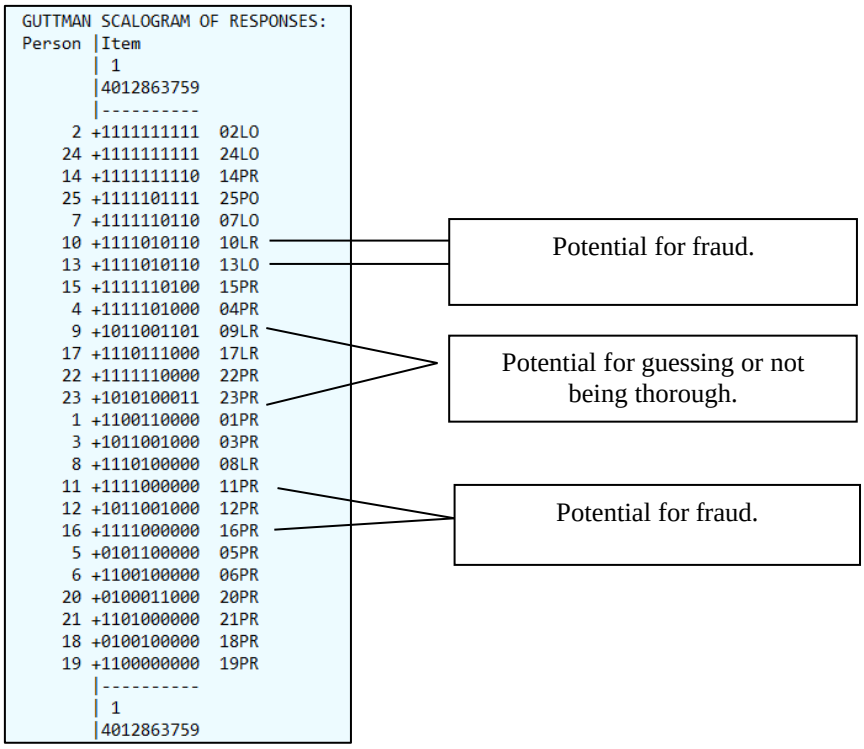

Figure 6. Student Response Patterns Based on the Guttman Scalogram or Matrix

The Guttman Matrix can show respondents' guesses or inaccuracies in answering the test items given (Darmana et al., 2021). This scalogram data is arranged based on the easy level to the most difficult level (from item 4 to item 9), and simultaneously respondents are sorted from the most capable rank to the lowest (from person 2 to 19). In Figure 6, respondents with codes 09LR and 23PR answered incorrectly for very easy questions (item 10) and answered correctly for very difficult questions (item 9). The Guttman Matrix has the ability to analyze indications of fraud in addition to finding guesses. Of the 25 respondents, four respondents had the same answer pattern. Namely, respondents with codes 10 LR and 13 LO as well as 11PR and 16PR are indicated to have cheated or collaborated when taking the test.

Item Bias Analysis Based on Gender

Item bias means test conditions that are unfair, inconsistent, and polluted by factors beyond the ability of the factor to be measured. Item bias also means a test that is discriminatory or favors certain groups, the causes of which can be seen from various aspects, such as gender, ethnicity, culture, region, etc. (Darmana et al., 2021). If items are assigned more favorably to students with certain attributes than to students with other attributes, the items can be considered biased. Students are given tools to see the gender of male (L) and female (P). If the probability value of an item is below 5% (0.05), the item can be considered biased (Lestari et al., 2020). Figure 7 below shows the results of item bias analysis based on gender.

Person CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
1	.0000	0	1.0000	.0000	.0000	1	S1
2	1.0392	1	.3080	.4123	-.0711	2	S2
2	.9790	1	.3224	.5133	.0486	3	S3
1	.0000	0	1.0000	.0000	.0000	4	S4
2	.6970	1	.4038	.3747	-.1205	5	S5
2	.4729	1	.4917	.2359	-.3393	6	S6
2	1.9697	1	.1605	1.1080	.5451	7	S7
2	2.1104	1	.1463	1.0259	.4896	8	S8
2	.0016	1	.9684	.0008	-1.4568	9	S9
2	.5052	1	.4772	.2360	-.3389	10	S10

Figure 7. Results of Item Bias Analysis Based on Gender

From Figure 7, no items were found that had a probability value <0.05, so it is said that there is no gender bias in the items being developed.

Item Bias Analysis Based on Class X Origin

Figure 8 below shows the results of item bias analysis based on class X origin.

Person CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
1	.0000	0	1.0000	.0000	.0000	1	S1
1	.0000	0	1.0000	.0000	.0000	2	S2
2	1.3864	1	.2390	.7926	.3132	3	S3
1	.0000	0	1.0000	.0000	.0000	4	S4
1	.0000	0	1.0000	.0000	.0000	5	S5
2	.1470	1	.7014	.0619	-.8108	6	S6
1	.0000	0	1.0000	.0000	.0000	7	S7
2	.5074	1	.4763	.2407	-.3304	8	S8
2	.0242	1	.8763	.0145	-1.1322	9	S9
1	.0000	0	1.0000	.0000	.0000	10	S10

Figure 8. Results of Item Bias Analysis Based on Class X Origin

From Figure 8, no items were obtained that had a probability value <0.05, so it is said that there is no class X origin bias in the items being developed.

▪ CONCLUSION

The conclusion about the quality of the High Order Thinking Skill-based test items in chemical equilibrium material is as follows: Validation results using V Aiken by experts stated that ten items were valid from the material, construction and language aspects; content validation through statistical item output: misfit order obtained as many as 10 items indexed as fit items (100%); person reliability is 0.69 (fair) while item reliability is 0.86 (good) with Cronbach α 0.72 which includes good criteria; the value of person separation (1.27) is in the weak category and item (2.49) is in the good category; there are four levels of item difficulty, each classified as very difficult, difficult, medium and very easy. (30% : 20% : 20% : 30%); respondents coded 02LO and 24LO (logit +4.36) are respondents with the highest ability, codes 18PR and 19PR (logit -1.97) are respondents with the lowest ability; Guttman Matrix Scalogram, there are two respondents who have the potential to guess the answer and four respondents who have the potential to commit fraud; there were no items that detected bias based on differences in gender and class X origin of the respondents. The results of this study show a relatively small variation in respondents, this is because the number of test participants was only 25

people and the type of data was dichotomous. Take the HOTS chemical equilibrium questions that were developed and tested on a larger number of respondents, namely 150 people, to show the stability of the data from the multiple choice test.

▪ REFERENCES

Astari, M. A., Masykuri, M., Susilowati, E., & Yamtinah, S. (2023). Effectiveness of Online Learning with the Web-based TPACK Scaffolding for Enhancement TPACK Ability of Pre-service Chemistry Teachers. *Journal of Curriculum and Teaching*, *12*(1). https://doi.org/10.5430/jct.v12n1p183

Bond, T., Yan, Z., & Heene, M. (2020). Applying the rasch model: Fundamental measurement in the human sciences. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. https://doi.org/10.4324/9780429030499

BSNP. (2020). *Paradigma Pendidikan Nasional di Abad 21*. Badan Standar Nasional Pendidikan.

Chi, S., Wang, Z., & Liu, X. (2022). Assessment of Context-Based Chemistry Problem-Solving Skills: Test Design and Results from Ninth-Grade Students. *Research in Science Education*, *53*(2), 295–318. https://doi.org/10.1007/s11165-022-10056-8

Dalimunthe, M., & Tarigan, A. T. S. (2023). Development of A Five-Tier Diagnostic Test Instrument on Atomic Structure Material to Identify Misconceptions of SMA Negeri 2 Medan Students. *Jurnal Pendidikan dan Pembelajaran Kimia*, *12*(3), 102–114. https://doi.org/10.23960/jppk.v12.i3.2023.09

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. https://doi.org/10.24815/jpsi.v9i3.19618

Hale, L. V. A., Lutter, J. C., & Shultz, G. V. (2016). The development of a tool for measuring graduate students' topic specific pedagogical content knowledge of thin layer chromatography. *Chemistry Education Research and Practice*, *17*(4), 700–710. https://doi.org/10.1039/c5rp00190k

Harta, J., Rasuh, N. T., & Seriang, A. (2020). Using HOTS-Based Chemistry National Exam Questions to Map the Analytical Abilities of Senior High School Students. *Journal of Science Learning*, *3*(3), 143–148. https://doi.org/10.17509/jsl.v3i3.22387

He, P., Zheng, C., & Li, T. (2022). Upper Secondary School Students' Conceptions of Chemical Equilibrium in Aqueous Solutions: Development and Validation of A Two-Tier Diagnostic Instrument. *Journal of Baltic Science Education*, *21*(3), 428–444. https://doi.org/10.33225/jbse/22.21.428

Hirza, S., Muchtar, Z., Sutiani, A., Dibyantini, R. E., & Sinaga, M. (2023). Development of an Evaluation Instrument to Measure Higher-Order Thinking Skills in Acid-Base Topic. *Jurnal Pendidikan dan Pembelajaran Kimia*, *12*(1), 13–29. https://doi.org/10.23960/jppk.v12.i1.2023.02

Kim How, R. P. T., Zulnaidi, H., & Rahim, S. S. B. A. (2022). Development of Higher-Order Thinking Skills test instrument on Quadratic Equations (HOTS-QE) for Secondary School Students. *Pegem Egitim ve Ogretim Dergisi*, *13*(1), 379–394. https://doi.org/10.47750/pegegog.13.01.41

Lestari, W. T., Saputro, S., Masykuri, M., Hastuti, B., Ulfa, M., Mulyani, S., & Yamtinah, S. (2020). Item analysis of teachnological pedagogical content knowledge (TPACK)

in pre-service chemistry teachers using the Rasch Model application. *Journal of Physics: Conference Series*, *1511*(1). https://doi.org/10.1088/1742-6596/1511/1/012043

Linacre, J. M. (2016). *A User's guide to Winsteps-ministep: Rasch-model computer programs. Program Manual*.

Liu, C., & Jurich, D. (2023). Outlier Detection Using t-test in Rasch IRT Equating under NEAT Design. Applied Psychological Measurement. *Applied Psychological Measurement*, *47*(1), 34–47. https://doi.org/10.1177/01466216221124045

Lubis, A. R., Sudrajad, A., & Nugraha, A. W. (2021). Pengembangan Instrument Asessment Higher Order Thinking Skill (HOTS) Untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Pada Materi Hidrolisis Garam. *Prosiding Seminar Nasional Kimia Dan Kependidikan Kimia*, 173–181.

Napitupulu, N. M., & Yusuf, M. (2022). Pengembangan Instrumen Tes Berbasis HOTS (Higher Order Thinking Skills) Pada Materi Laju Reaksi Kimia. *Jurnal Sekolah*, *6*(3), 162–168. https://doi.org/10.24114/js.v6i3.36179

OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Oliva, J. M., & Blanco, Á. (2022). Rasch analysis and validity of the construct understanding of the nature of models in Spanish-speaking students. *European Journal of Science and Mathematics Education*, *11*(2), 344–359. https://doi.org/10.30935/scimath/12651

Shakeel, S. I., Al Mamun, M. A., & Haolader, M. F. A. (2022). Instructional design with ADDIE and rapid prototyping for blended learning: validation and its acceptance in the context of TVET Bangladesh. *Education and Information Technologies*. https://doi.org/10.1007/s10639-022-11471-0

Silviariza, W. Y., Sumarmi, S., Utaya, S., Bachri, S., & Handoyo, B. (2023). Development of Evaluation Instruments to Measure the Quality of Spatial Problem Based Learning (SPBL): CIPP Framework. *International Journal of Instruction*, *16*(2), 413–436. https://doi.org/10.29333/iji.2023.16223a

Siregar, A. U., Lubis, A., Darmana, A., & Suyanti, R. D. (2022). Analysis of Student Cognitive Ability using The K-Means Cluster and Rasch Model (Winsteps) on Chemical Materials through Zoom Meeting. *Jurnal Komunikasi Pendidikan*, *6*(2), 91. https://doi.org/10.32585/jkp.v6i2.2474

Suhardi, I., & Budiyaningsih, H. E. (2022). Validasi Pakar Konten pada Pengembangan Paket Soal Bahasa Indonesia untuk Basis Data Computer Based Testing. *Jurnal MediaTIK : Jurnal Media Pendidikan Teknik Informatika Dan Komputer*, *38*(2). https://doi.org/10.26858/jmtik.v5i2.33784

Sukmawati, W., & Zulherman, Z. (2023). Analysis of Changes in Students' Scientific Literacy Ability After Attending Lectures Using the RADEC Model. *Jurnal Penelitian Pendidikan IPA*, *9*(3), 1039–1044. https://doi.org/10.29303/jppipa.v9i3.2846

Sumintono, B., & Wahyu Widhiarso. (2015). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. TrimKom.

Sutiani, A., Muchtar, Z., Dibyantini, R. E., Sinaga, M., & Purba, J. (2022). Analisis Kemampuan Guru-Guru Kimia SMA Sumatera Utara Dalam Mengintegrasikan TPACK. *Jurnal Inovasi Pembelajaran Kimia (Journal Of Innovation in Chemistry Education)* , *4*(2), 112–131. https://doi.org/10.24114/jipk.v4i2.39259

Syathroh, I. L., & Rizkiani, S. (2023). Developing English Research Article Writing Guide Textbook For Non-English Department Students Using The ADDIE Model. *Indonesian EFL Journal (IEFLJ)*. https://doi.org/10.25134/ieflj.v9i1.7183

Wahyudi, A., Richardo, R., Eilks, I., & Kulgemeyer, C. (2023). Development of Three Tier Open-Ended Instrument to Measure Chemistry Students' Critical Thinking Disposition Using Rasch Analysis. *International Journal of Instruction*, *16*(3), 191–204. https://doi.org/10.29333/iji.2023.16311a

Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The development of the hots test of physics based on modern test theory: Question modeling through e-learning of moodle lms. *International Journal of Instruction*, *14*(4), 51–68. https://doi.org/10.29333/iji.2021.1444a

Yusmaita, E., Yuliani, F., Gazali, F., & Sumintono, B. (2022). Chemical Literacy Instrument on Oxidation and Reduction Reactions Material Through Rasch Model. *Journal of Chemistry Education Research*, 6 (2), 124–30. https://doi.org/10.26740/jcer.v6n2.p124-130