Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index

Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

Development of a Higher-Order Thinking Skills (HOTS) Test Instrument on Electrochemical Material for High School Students

Nuzulul Nugrahnastiti, Agus Kamaludin*
Chemistry Education Department, Faculty of Tarbiyah and Teacher Training
Universitas Islam Negeri Sunan Kalijaga, Indonesia.
*Corresponding Author. Email: aguskamaludin@gmail.com

Abstract: This research aims to develop a HOTS test instrument on electrochemistry topics that meets valid and reliable criteria. Apart from that, the level of difficulty, distinguishing power, ability level, and detection of respondent bias were also analyzed. This research is development research using the Tessmer model, which includes preliminary and formative evaluation stages. The HOTS questions consist of 15 multiple-choice questions. The validation data from experts is measured using a Likert scale. Coefficient Aikens' V is then used to measure the validation data from reviewers. Experts and reviewer assessments indicated that the HOTS question instrument developed was very good. Analysis of questions using the Rasch model in a field testing of 33 high school students using a purposive sampling technique resulted in 14 valid questions and 1 invalid question with a reliability value of 0,69. There were 3 questions in the easy category, 7 in the moderate category, 1 in the hard category, and 4 in the extremely hard category. The analysis of the distinguishing power of the questions obtained 9 questions with very good criteria, 3 questions with good criteria, and 2 with fair criteria. The results of the validity of the students` responses showed that there were 4 invalid responses. The ability level shows that 5 students have high ability, 7 have moderate ability, and 21 have low ability. In conclusion, the developed HOTS test instrument for electrochemical material has a good quality and is suitable for evaluation.

Article History
Received: 14-03-2024
Revised: 25-04-2024
Accepted: 12-05-2024
Published: 22-06-2024

Key Words:
HOTS, Instruments; Test; Electrochemistry.

How to Cite: Nugrahnastiti, N., & Kamaludin, A. (2024). Development of a Higher-Order Thinking Skills (HOTS) Test Instrument on Electrochemical Material for High School Students. *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran, 10*(2), 451-466. doi:https://doi.org/10.33394/jk.v10i2.11370

https://doi.org/10.33394/jk.v10i2.11370

Introduction

Science literacy is fundamental to achieving educational success in the 21st century (Budiarti & Tanta, 2021). Science literacy is the ability to comprehend how science, technology, and society interact and to use that understanding to solve issues in everyday life (Toharudin et al., 2011). Science literacy is needed to develop science thinking skills to understand advances in technology and science (Pratiwi et al., 2019). Education is essential to ensure students have good science literacy competencies (Andrian & Rusman, 2019). Through science literacy, students will develop their critical and logical thinking skills, creativity, and communication and teamwork abilities. (Sibarani & Tenriawaru, 2019; Yuliati, 2017). However, students' science literacy competency in Indonesia is still classified as low (Herlanti et al., 2019). According to a PISA assessment from 2022, students in Indonesia had an average score of 383 for science literacy, indicating the poor level of science literacy (OECD, 2023a). This score is still at level 1, the lowest level in the PISA scoring system (OECD, 2023b).

The government is making efforts to increase students' scientific literacy in Indonesia by implementing an independent curriculum (Merdeka Curriculum) (Angga et al., 2022;

Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

Pratikno et al., 2022). The independent curriculum is designed by expanding the content of the material to suit needs so that students can improve their higher-order thinking Skills (HOTS) (Tasrif, 2022). In addition, teachers are required to use higher-order thinking (HOT)-based assessments in order to enhance students' abilities in this domain (Febriyani et al., 2020). Students' scientific literacy competencies can be supported through HOTS questions. (Yuriza & Sigit, 2018). HOTS is an assessment used to measure high-level thinking skills which include analyze (C4), evaluate (C5), and create (C6) (Setiawan et al., 2021). This method encourages students to achieve higher abilities in the form of problem-solving, decision-making, critical thinking, and creativity (Andini & Supardi, 2018). HOTS enables students to analyse, relate, and evaluate all elements of situations and problems that occur in daily life. (Rofifah et al., 2013).

Teachers must have pedagogical competencies in order to improve students' HOTS abilities (Aquami, 2018). This competency allows teachers to manage learning in the classroom starting from the planning, implementation, and evaluation stages of student learning processes and outcomes (Emiliasari, 2018). The teacher's learning planning ability takes the form of preparing a learning module as a guide for implementing learning for each learning process activity, which is developed based on the syllabus (Wati et al., 2021). Furthermore, throughout the learning process, teachers must engage in appropriate communication and interaction so that the material presented by the teacher may be well understood by students (Nasution & Mursell, 2008). During the assessment phase, an in-depth assessment is conducted to evaluate the input, process, and learning outcomes. Assessment aims to measure various student skills in various contexts that reflect situations in the world (Mulia, 2020). Increasing teacher pedagogical competence in the evaluation domain can improve students' abilities to solve HOTS-based questions (Astutik & Roesminingsih, 2021). However, high school chemistry teachers still need more understanding of the HOTS question instrument (Driana et al., 2021). Teachers tend to only provide question instruments in the Lower Thinking Skills (LOTS) category (Khaldun et al., 2020).

One way to improve teachers' abilities in designing HOTS questions is by understanding the aspects that must be fulfilled in the HOTS question instrument (Puspitasari et al., 2021). Brookhart (2010) states that the HOTS question instrument must fulfill three aspects, namely: (1) the ability to connect learning material with elements other than those studied; (2) students' ability to reason, reflect, and make the right decisions; (3) the ability to identify and solve problems. According to Widana (2017), the HOTS test instrument must also apply contextual problems related to various aspects of life so that students are able to connect, interpret, apply, and integrate science in learning with problem-solving in everyday life. However, the availability of HOTS question references in chemistry subjects still needs to be improved (Zohrany et al., 2018). This reference's availability can help teachers design HOTS question instruments and train students to get used to working on HOTS questions (Syahana & Andromeda, 2021; Intan et al., 2020). Thus, it is necessary to develop HOTS-based chemistry question instruments (Jayanti, 2020).

Chemistry is a part of natural science that studies the characteristics of matter and the energy of its changes (Istijabatun, 2008). Chemistry is closely related to everyday life (Roy, 2016). Chemistry subject is contextual and has great potential for developing HOTS aspects (Lilia & Widodo, 2014; Harta et al., 2020). Students may find it challenging to fully understand chemistry due to its broad scope, complex concepts, and three levels of representation—macroscopic, submicroscopic, and symbolic (Erlina, 2011; Stojanovska et al., 2012; Talanquer, 2011). In order to help students learn chemistry subjects more simply, teachers might apply their knowledge of chemistry to everyday situations (Sunyono et al., 2015). Electrochemistry is a chemistry topic that has three levels of representation and is extremely complex. It is challenging for teachers to connect electrochemistry concepts to real-world situations (Sukmawati, 2019).

Electrochemistry is one of the chemistry subjects that studies electrically conducting instruments (electrodes) in which electrons are transferred from the positive electrode to the negative electrode with electricity as the energy source (Harahap, 2016). The electrochemistry process is based on a reduction-oxidation reaction in an electrolyte solution (Sukmawati, 2020). Electrochemistry topics are related to various human needs such as dry cell batteries and components used in flashlights (Rudibyani et al., 2020). Therefore, electrochemistry is a contextual chemistry subject and has the potential to develop its HOTS aspects (Agustina et al., 2021). However, there is still a limited number of HOTS test instruments available for electrochemistry topics in schools. To date, the development of HOTS test instruments using Rasch model analysis has only been available on stoichiometric (Abdullah et al., 2022), salt hydrolysis (Lubis et al., 2022), and acid-base reaction (Aprilia et al., 2021).

Based on the problems above, research is needed regarding the development of instruments for Higher-Order-Thinking Skill (HOTS) questions on electrochemistry topics. It is hoped that this research can improve students' science literacy and high-level thinking skills so that they are able to solve HOTS-based questions effectively and accurately.

Research Method

The study used a research and development method with a Tessmer model, which consists of preliminary and formative evaluation stages, was used in this research. The formative evaluation stages include self-evaluation, prototyping (expert review, one-to-one, and small group), and field testing (Tessmer, 2013). Data collection instruments in this study include documentation, validation sheets, and test questions. The research product is a set of fifteen (15) HOTS question instruments on electrochemistry topics that have been validated by two experts: one in materials and the other in evaluation. The product is then revised and assessed by two reviewers who are high school chemistry teachers. The revised product is then tested on 33 students at one of Yogyakarta's Islamic public high schools.

Product Validation

Product validation data contains suggestions and input from materials experts, evaluation experts, and reviewers. This data is used as guidance to improve the HOTS test instrument. The feasibility and quality of the HOTS questions were evaluated by a qualitative examination of the experts. The data is gathered in the form of a quality assessment, which is then measured using a Likert scale (1 to 5). The questions' feasibility can be determined by calculating the average value. If the score percentage meets the criteria of < 21% it is considered very poor, 21 – 40 % is considered poor, 41 – 60% is considered fair, 61 – 80% is considered good, > 81% is considered very good criteria (Arikunto, 2013; Ernawati, 2017). Aiken's V formula was used to analyze the collected data from reviewers (Likert scale 1 to 5) and determine the content validity of each question item (Aiken, 1985).

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

The Winstep software was then used to conduct a Rasch modelling analysis on the responses provided by students during field testing.

Validity Test

The validity test conducted in this research includes testing the validity of test items and student responses. The validity of the question items and response with the Rasch Model can be seen from the MNSQ, ZSTD outfit, and Pt Measure Corr values (Sumintono & Widhiarso, 2015). Question items are accepted if the MNSQ outfit values are in the range 0,5 to 1,5. ZSTD outfit values are in the range of -2,0 to +2,0. Pt Measure Corr values are in the range 0,4 to 0,85. If an item is found where the Z-STD value meets the criteria but the MNSQ and Pt Measure Corr fail to meet the requirements, the item remains considered valid (fit) (Bond et al., 2020; W. J. Boone, 2016). When it comes to the three criteria of MNSQ, ZSTD, and Pt Measure Corr, students' answers are considered fit if they meet at least two of the values (W. Boone, 2013). The response must be further analysed in the scalogram if it does not meet these two criteria.

Reliability Test

An instrument is reliable if repeated measurements consistently produce the same information. Reliability testing using the Rasch Model must meet Person or Item Reliability value criteria, as stated by Sumintono & Widhiarso (2015). If the value meets the criteria of < 0,67 it is considered weak, 0,67 – 0,80 is considered fair, 0,81 – 0,90 is considered good, 0,91 – 0,94 is considered very good, > 0,94% is considered excellent criteria (Perera et al., 2018). According to Sumintono & Widhiarso (2013), the Alpha Cronbach value is the coefficient that measures instrument reliability. If the value of Alpha Cronbach meets the criteria of 0,00 – 0,49, it is considered very poor; 0,50 – 0,59 is considered poor; 0,60 – 0,69 is considered fair; 0,70 – 0,79 is considered good; 0,80 – 1,00 is considered very good criteria.

Item Difficulty Level

The Item Measure table in the Winsteps program can be used to determine the test item difficulty level. The logit Measure and Standard Deviation (SD) values are used to categorize the test items' difficulty levels. The value of the SD of this research is -1,62. The item logits are divided into four categories, which can be seen in Table 1 (Maryati et al., 2019; Sumintono, 2018).

Table 1. Item difficulty level category

Measure (logit) Value	Category
Measure logit < -1,62	Easy
-1,62 ≤ Measure logit ≤ 0	Moderate
0 ≤ Measure logit ≤ 1,62	Hard
Measure logit > 1,62	Very Hard

Item Discrimination Power

The item discrimination power in the Winsteps program can be observed through the Pt Measure Corr values in the item fit order table (Sumintono, 2018). Categories for selecting items based on discrimination power shown in Table 2.

Table 2. Items discrimination power criteria

Pt Measure Corr Value	Interpretation
> 0,4	Very Good
0,39 – 0,30	Good
0,29 – 0,20	Fair
0,19 – 0,00	Cannot discriminate
< 0,00	Requires inspection of items

Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

Students' Ability Level

Person measures in the Winsteps program are used to identify the level of students' ability to answer questions. The logit means and Standard Deviation (SD) values are used to categorise the students` ability levels. According to Sumintono & Widhiarso, (2015), Table 3 shows the criteria for determining students` ability.

Table 3. Logit Value for students` abilities

Measure (logit)	Category
> SD logit	High
Mean logit – SD logit	Moderate
< Mean logit	Low

Respondent Bias Detection

Bias analysis is conducted to determine whether test items exhibit bias in specific respondent categories. In the Winsteps program, information regarding item bias can be observed through the DIF (Differential Item Functioning) table, between/within. Items with P (PROB.) values below 0.05 indicate bias in certain groups (Sumintono & Widhiarso, 2015).

Results and Discussion

The aspects assessed in validating LKPD include format, content, language, presentation, and how LKPD supports innovations and improves the Teaching and Learning Activities (KBM). Before implementation, the developed LKPD must be validated by experts in Focus Group Discussion (FGD) activities. Validation was carried out by three experts in educational evaluation, biodiversity biology material, and language. The validation results by experts show that the model validation sheet developed was valid or suitable for use. The validation results of LKPD are listed in Table 1 below:

The Preliminary Stage begins with collecting relevant literature such as books, articles, and information regarding HOTS that students should possess in the independent curriculum. Next, a literature review related to HOTS is conducted, focusing on the urgency of HOTS and a review of relevant previous research. The following step, called self-evaluation, creates test questions that measure higher-order thinking skills. Preparing the HOTS question grid for Electrochemistry topics refers to the syllabus of twelfth-grade chemistry and the analysis of Learning Outcomes (CP) for the F phase in the chemical understanding element. Then, based on its cognitive components, this learning achievement is developed into Learning Objectives to produce HOTS questions (Anggraena et al., 2022).

Table 4. Learning Objectives Opportunities to form HOTS questions

Learning Obejctives	Opportunities
Analyzing the concept of redox reactions and balancing its reaction	High
Analyzing the process of chemical energy transformation that occurs in voltaic cells and its application in everyday life	High
Analyzing the process of chemical energy transformation that occurs in electrolytic cells and its application in everyday life	High

Interviews were conducted with two chemistry teachers in twelfth grade from two different high schools in Yogyakarta. HOTS-based learning processes have been implemented but are limited to contextual stimulus-based learning on specific topics. Teachers only provide 30% of HOTS-based questions out of the total number of questions tested in daily and summative assessments. The rarity of HOTS-based questions in assessments is due to the fact that they require many references and more time to create. This phenomenon has led to a lack of availability of good HOTS question instrument references (Kurniasi et al., 2020). Additionally, based on research by Nadliroh et al. (2019), the

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan
di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index
Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

availability of HOTS questions in school textbooks is still very minimal. The development of HOTS-based questions is crucial for teachers' assessment purposes. Based on these findings, a product was developed in the form of HOTS test instrument for Electrochemistry topics, referring to the question grids that had been created. The developed HOTS questions are multiple-choice questions consisting of 15 items with five answer options.

The product was then validated by materials experts, evaluation experts, and reviewers. Feedback and suggestions from the validation results were then used as a reference for the improvement of the product of HOTS test instruments. Subsequently, the product was tested one-on-one with three twelfth-grade science students. The three students were selected as subjects due to their varying ability levels, which were high, moderate, and low. Students were asked to submit feedback and suggestions after responding to the questions. Based on the criticism and suggestions, the product proceeded to small group testing. The product was tested in a small group consisting of 6 students. Then, the HOTS Electrochemistry test instrument proceeded to the field-testing stage.

Product Validation

The HOTS question instruments developed in this research consist of 15 multiple-choice questions. According to Hartini & Sulitdjo (2015), multiple-choice questions are widely applicable and have high objectivity and validity. The purpose of validation is to evaluate the quality of the HOTS test instrument's initial product, which is then improved in response to feedback and suggestions from evaluation and materials experts (Widhiyani et al., 2019). **Figures 1** and **Figure 2** show the results of the validation analysis and evaluation by experts in materials and evaluation.

Figure 1. Analysis result of material expert's validation

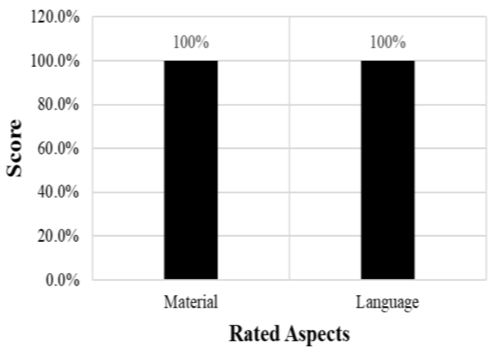
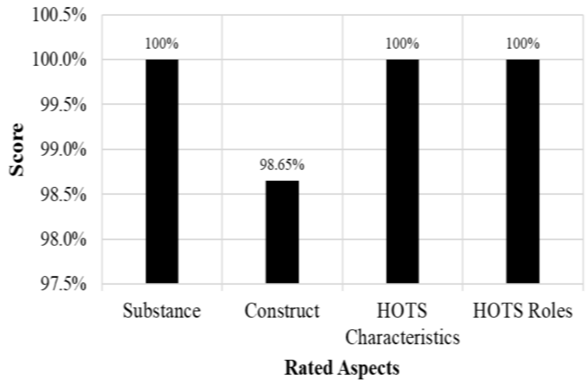


Figure 2. Analysis result of evaluation expert's validation



The validation results from subject matter experts and evaluation experts indicate that the developed questions are in the category of very good (Anggraini et al., 2020a) The product is therefore ideal to be implemented as an instrument for assessing students' higher-order thinking skills. Reviewers were then involved in the validation and assessment process to evaluate the developed product's content validity. Using Aiken's validation analysis, the reviewer assessment scores were examined. The analysis results can be considered valid if they reach the 0.80 Aiken's V coefficient limit (Aiken, 1985). Figure 3 shows the results of the reviewer assessments.

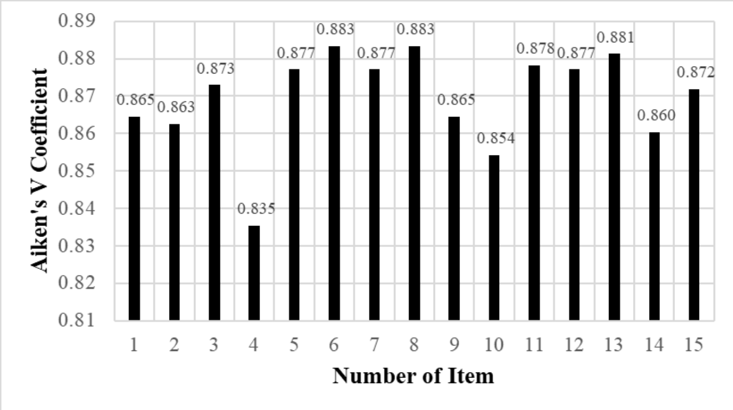


Figure 3. Analysis of Reviewer Assessment Results

According to the reviewer's assessment, each item has an overall coefficient of ≥ 0.80, with a specific value of 0.869. This suggests that the Higher Order Thinking Skills (HOTS) test instruments that were developed are valid (Azwar, 2015). This corresponds with the research by Rahielanamy & Kamaludin (2022), which found that the HOTS-based question were valid and had an Aiken's V value of 0.879. As a result, the developed product can be used as an instrument for assessments that evaluate students` higher-order thinking skills. The criteria for HOTS questions used in this study include measuring high-level thinking abilities, being context-based, stimulus-based, and having novelty in their questions (Awaliyah, 2018). Higher-order thinking skills can include analyzing, evaluating, and creating. Being context-based means that the questions are linked to everyday life. The stimuli presented in the questions can be in graphs, tables, text stories, or images.

Question Validity

Validity testing of question items aims to determine the suitability of the instrument in measuring variables (Ghazali, 2016). Response validity testing aims to ascertain the stability of students' thinking in answering questions and to detect any cheating (W. Boone, 2013). The validity results of the questions on item fit order using Rash model analysis are shown in Table 5.

Table 5. Item Fit Order Result

Question Number	Outfit MNSQ	Outfit ZSTD	Pt Measure Corr	Discriminating Power Category	Fit/misfit
1	0,82	-0,40	0,50	Very good	Fit
2	0,66	-0,50	0,53	Very good	Fit
3	0,90	0,10	0,34	Good	Fit
4	0,51	-0,80	0,61	Very good	Fit
5	0,85	-0,40	0,51	Very good	Fit
6	1,19	0,50	0,49	Very good	Fit
7	0,7	-0,30	0,45	Very good	Fit
8	1,02	0,20	0,34	Good	Fit
9	3,61	2,80	0,31	Good	Misfit
10	1,70	1,30	0,28	Fair	Fit
11	0,71	-0,90	0,59	Very good	Fit
12	0,30	-0,30	0,52	Very good	Fit
13	1,21	0,50	0,23	Fair	Fit
14	1,41	1,10	0,31	Good	Fit
15	0,96	0,00	0,46	Very good	Fit

Item number 9 is considered invalid (misfit) because it has an outfit Z-STD value greater than 2.0. The other items remain valid (fit) because they meet the criteria for Z-STD values (Bond et al., 2020; W. J. Boone, 2016). Item number 9 should be discarded or revised

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan
di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index

Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

for further testing. According to the findings of Ridwan et al. (2023) items that do not meet the outfit MNSQ and Z-STD criteria are considered unfit and must be improved. As a result, the final 14 items are able to evaluate students' higher order thinking skills in electrochemistry (Astuti et al., 2022). Rasch modeling can also detect students' responses that do not match their abilities and the ideal model. The validity results of response in person fit order are shown in Table 6.

Table 6. Person Fit Order Results

Fit/Misfit	Total	Respondent Code
Fit	29	01AP, 02HL, 03YL, 04ML, 05MP, 06ML, 07CP, 08NP, 09SP, 10XL, 11VL, 12AP, 13OP, 14AP, 15SP, 16ML, 17ML, 18RP, 20NP, 21NP, 22FP, 23NP, 24DP, 25ML, 27FP, 28SP, 29RP, 32PL, 33ML
Misfit	4	19AP, 31EL, 30RL, 26AL

Out of 33 respondents, 4 students have invalid responses (misfit) because they do not meet two of three criteria: the outfit Mean Square value, the outfit Z-standard value, and the Pt Measure Corr value. Misfit responses show inconsistent response patterns for every item. Since misfit responses may cause noise and biases in the measurement process, it is important to identify it (Ocy et al., 2023). Therefore, each misfit response must be analyzed with a scalogram to determine the cause of the response not aligning with Rasch modeling (Damanik et al., 2023). The scalogram developed by Louis Guttman provides information about why response patterns do not conform to the model (Sumintono & Widhiarso, 2015). The results of the scalogram are shown in Figure 4 and Figure 5.

Figure 4. Scalogram for misfit response

```
26 +110000100110000   26AL
30 +000001001100110   30RL
31 +100000010011100   31EL
 7 +011001010000000   07CP
12 +111001000000000   12AP
20 +101010100000000   20NP
 8 +110000000000000   08NP
19 +000000100000100   19AP
   |---------------
   |1      1  11  11
   |337428515106942
```

Figure 5. Scalogram for cheating response

```
   |1      1  11  11
   |337428515106942
   |---------------
16 +111111111101111   16ML
11 +111111111111100   11VL
14 +111111111111100   14AP
17 +111111111100010   17ML
33 +111111110110001   33ML
22 +011111111100000   22FP
```

From the figure, it can be observed that from left to right represents the sequence of questions from the easiest to the hardest. The four students with invalid responses were unable to answer the easiest questions but could correctly answer the more difficult ones. Misfit responses occur due to lucky guessing and lack of carefulness in answering questions. Misfit responses indicate that they have response patterns that cannot be predicted by Rasch modeling (Smith Jr., 2001). Furthermore, scalogram analysis can also be used to detect cheating among students based on similar matrix patterns, as seen in the case of students 11VL and 14AP.

Item Discriminating Power Analysis

Item discrimination is used to determine the ability of items to differentiate between respondents with high and low levels of ability (Wulandari et al., 2022). Items number 1, 2, 4, 5, 6, 7, 11, 12, and 15 meet the very good category, items number 3, 8, 9, and 14 meet the good category, and items number 10 and 13 meet the fair category. Items with very good and good category indicate that they do not mislead respondents, so low-ability respondents choose incorrect answers and high-ability respondents choose the correct ones (Ocy et al., 2023). On the other hand, items categorized as fair indicate that only some respondents are confused when selecting the answers. This corresponds with the research conducted by Ngadi

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index
Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

(2023) which states that items with Pt Measure Corr > 0.4 have good discriminating power, enabling them to differentiate between groups of respondents with high and low abilities.

Question Reliability

According to the reliability test findings, the product fits fair category with a Cronbach's alpha score of 0.69, meaning that the instrument adequately measures the interaction between students and the overall test instrument. The person reliability value of 0.69 is also categorized as fair, indicating that respondents' answers are fairly consistent and students' abilities vary moderately (Taber, 2018). The item reliability value of 0.90 is categorized as good, meaning the test instrument demonstrates good consistency in measuring students' high-level thinking skills. Therefore, this test instrument can be used for assessment because it can be relied upon to provide consistent results if reused under the same conditions (Tavakol & Dennick, 2011; Yusup, 2018).

Item Difficulty Level Analysis

The difficulty level in Rasch modeling is based on the logit measure value, which is the probability that respondents will answer the question (Mulyanti et al., 2022). The difficulty level of the instruments is yet to have a balanced proportion. The moderate category has the highest percentage, 7 items (46%) in questions 2, 4, 8, 15, 1, 11, and 5. The very hard category consists of 4 items (26%) in questions 12, 6, 9, 14. The easy category consists of 3 items (20%) in questions 7, 3, 13, and the hard category only has 1 item (0.06%) in question 10. Items should ideally have a balanced proportion of difficulty levels, with easy items at 25%, moderate items at 50%, and hard items at 25% (Amalia & Widayati, 2012). Additionally, question items are considered good quality if they are neither too easy nor too difficult. This is because overly easy items cannot stimulate effort in answering, while overly difficult items can lead to frustration due to being beyond the respondent's capability (Iskandar & Rizal, 2018).

Students` Ability Levels Analysis

The Winstep program's person-measure results are used to determine the students' ability to answer questions. The students` ability levels are ranked from highest to lowest based on the logit measure value of each response. The analysis's findings indicate that the mean logit value is -0.16 and the standard deviation (SD) is 1.35. There are 5 students with high abilities, 7 students with moderate abilities and the majority of students, totaling 21, have low abilities. The category of students` ability levels is shown in **table 7.**

Table 7. Students` ability levels category

Measure (logit)	Category	Respondent Codes
> 1,32	High	16ML, 11VL, 14AP, 17ML, 33ML
-0,16 – 1,32	Moderate	22FP, 24DP, 02HL, 23NP, 27FP, 28SP, 32PL
< -0,16	Low	03YL, 04ML, 05MP, 09SP, 13OP, 15SP, 18RP, 21NP, 25ML, 29RP, 01AP, 06ML, 10XL, 26AL, 30RL, 31EL, 07CP, 12AP, 20NP, 08NP, 19AP

Respondent Bias Detection Analysis

Bias detection analyzed in this instrument is gender-based. Bias detection analysis in the Rasch model uses the concept of Differential Item Functioning (DIF). Gender is used to detect bias because there is a relationship between gender and someone's cognitive abilities (Chung & Chang, 2017; Halpern et al., 2007). The results of respondent bias analysis can be seen in Table 8, where the probability values for all items are above 0.05, indicating that gender does not discriminate against the test items.

Table 8. Differential Item Functioning (DIF) Results

Summary DIF Chi-Square	Probability	Between		Question Number
		MNSQ	Z-STD	
0,6393	0,4240	0,3236	-0,1935	1
0,1373	0,7109	0,0683	-0,7828	2
0,3369	0,5616	0,1685	-0,4845	3
0,5201	0,4708	0,2603	-0,2955	4
2,8873	0,893	1,5432	0,8015	5
0,4271	0,5134	0,2201	-0,3692	6
1,1917	0,2750	0,6015	0,1400	7
3,3635	0,0667	1,8031	0,9320	8
0,4271	0,5134	0,2201	-0,3692	9
1,3209	0,2504	0,6792	0,2148	10
0,296	0,8634	0,0145	-1,1321	11
0,0000	1,000	0,0000	0,0000	12
1,0015	0,3169	0,4780	0,0000	13
0,4271	0,5134	0,2202	-0,3692	14
0,1145	0,7351	0,0569	-0,8338	15

The HOTS test instrument that has been developed can be used to train students to solve HOTS questions, which enhances their critical thinking and problem-solving skills. Thus, it promotes the enhancement of students` science literacy. This research will also provide insights into creating more valid and reliable assessments that truly measure higher-order thinking skills rather than rote memorization.

Conclusion

The Higher Order Thinking Skills (HOTS) test instrument in electrochemistry topics for high school students was developed with evaluation from experts and reviewers, indicating that all items are valid and suitable for use. Analysis of questions using the Rasch model in a field test of 33 high school students using a purposive sampling technique resulted in 14 valid questions and 1 invalid question. The Cronbach's Alpha reliability value is 0.69, indicating fair reliability. The item reliability value is 0.90, indicating a good category and the person reliability value is 0.69, indicating a fair category. The difficulty level of the 15 items is 3 items categorized as easy, 7 as moderate, 1 as hard, and 4 as very hard. Item discrimination analysis results in 9 items categorized as very good, 3 as good, and 2 as fair. The validity response results indicate that 4 students have invalid responses (misfit). Scalogram analysis reveals the response misfit due to lucky guessing and careless responding. The ability level shows 5 students with high abilities, 7 students with moderate abilities, and 21 with low abilities. Therefore, the developed HOTS test instrument for electrochemistry materials has a good quality and is suitable for evaluation.

Recommendation

For further instruments developments further test instrument development, it is recommended that field testing in this research be conducted with a larger sample of students with more diverse abilities. Invalid student responses (misfit) could be removed as they can cause noise and bias the measurement. It will result in more accurate validity and reliability. Furthermore, chemistry teachers could start ensuring students' ability to solve HOTS questions by implementing learning strategies that promote active learning, critical thinking, and problem-solving.

References

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan
di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index

Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

Abdullah, Ramisweti, & Ardiansyah. (2022). Preliminary analysis: Development of higher order thinking skills problem on stoichiometry topic using the rasch model. *Jurnal Pendidikan Sains*, *10*(2), 32–43. https://doi.org/10.26714/jps.10.2.2022.32-43

Agustina, L., Feronika, T., & Yunita, L. (2021). Analysis of higher Order Thinking Skills Questions in the Brookhart Category in High School Chemistry Textbook: Curriculum 2013. *Journal of Educational Chemistry (JEC)*, *3*(1), 23–34. https://doi.org/10.21580/jec.2021.3.1.6546

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Amalia, A. N., & Widayati, A. (2012). Analisis butir soal tes kendali mutu kelas XII SMA mata pelajaran ekonomi akuntansi di kota Yogyakarta tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, *10*(1). https://doi.org/10.21831/jpai.v10i1.919

Andini, D. M., & Supardi, E. (2018). Kompetensi Pedagogik Guru Terhadap Efektivitas Pembelajaran dengan Variabel Kontrol Latar Belakang Pendidikan Guru. *Jurnal Pendidikan Manajemen Perkantoran*, *3*(1), 1–7. https://doi.org/10.17509/jpm.v3i1.9450

Andrian, Y., & Rusman, R. (2019). Implementasi Pembelajaran Abad 21 dalam Kurikulum 2013. *Jurnal Penelitian Ilmu Pendidikan*, *12*(1), 14–23. https://doi.org/10.21831/jpipfip.v12i1.20116

Angga, A., Suryana, C., Nurwahidah, I., Hernawan, A. H., & Prihantini, P. (2022). Komparasi Implementasi Kurikulum 2013 dan Kurikulum Merdeka di Sekolah Dasar Kabupaten Garut. *Jurnal Basicedu*, *6*(4), 5877–5889. https://doi.org/10.31004/basicedu.v6i4.3149

Anggraena, Y., Ginanto, D., Felicia, N., Andiarti, A., Herutami, I., Alhapip, L., Iswoyo, S., Hartini, Y., & Mahardika, R. L. (2022). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Menengah*. Badan Penelitian dan Pengembangan dan Perbukuan 2021.

Aprilia, N., Susilaningsih, E., Sudarmin, S., Sumarni, W., Mahatmanti, F. W., & Adhelia, N. U. (2021). Application of the rasch model on the test instrument to analyze the problem-solving ability of students in acid-base solutions materials. *EDUSAINS*, *13*(2), 106–118. https://doi.org/10.15408/es.v13i2.21375

Aquami, A. (2018). Hubungan Kompetensi Guru dan Peran Orang Tua Terhadap Terhadap Hasil Belajar Siswa di MIN Se-Kota Palembang. *JIP: Jurnal Ilmiah PGMI*, *4*(1), 1–12. https://doi.org/10.19109/jip.v4i1.2211

Arikunto, S. (2013). *Prosedur Penelitian Suatu Pendekatan Praktek*. Rineka Cipta.

Astuti, S. R. D., Sari, A. R. P., Amelia, R. N., & Karlinda, K. (2022). Instrumen penilaian terintegrasi: Pengukuran kemampuan berpikir kritis dan keterampilan sains siswa pada materi larutan elektrolit dan nonelektrolit. *JURNAL TARBIYAH*, *29*(2), 340. https://doi.org/10.30829/tar.v29i2.1858

Astutik, K. F., & Roesminingsih, E. (2021). The Improvement of Teachers' Professional Competency Through Hots-Based Training. *JPI (Jurnal Pendidikan Indonesia)*, *10*(1), 145–151. https://doi.org/10.23887/jpi-undiksha.v10i1.22407

Awaliyah, S. (2018). Penyusunan Soal HOTS Bagi Guru PPKN dan IPS Sekolah Menengah Pertama. *Jurnal Praksis dan Dedikasi Sosial*, *1*(1), 46–53. http://dx.doi.org/10.17977/um032v0i0p48-55

Azwar, S. (2015). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (Fourth edition). Routledge.

Boone, W. (2013). *Rasch analysis in the human sciences*. Springer Berlin Heidelberg.

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE—Life Sciences Education*, *15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Brookhart, S. M. (2010). *How to assess high-er order thinking skills in your classroo*. ASCD.

Budiarti, I. S., & Tanta, T. (2021). Analysis on Students' Scientific Literacy of Newton's Law and Motion System in Living Things. *Jurnal Pendidikan Sains Indonesia*, *9*(1), 36–51. https://doi.org/10.24815/jpsi.v9i1.18470

Chung, L.-Y., & Chang, R.-C. (2017). The Effect of Gender on Motivation and Student Achievement in Digital Game-based Learning: A Case Study of a Contented-Based Classroom. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(6). https://doi.org/10.12973/eurasia.2017.01227a

Damanik, S. A., Saragih, I., & Darmana, A. (2023). Analisis dan Pengembangan Instrumen Soal Laju Reaksi Berbasis HOTS dengan Menggunakan Rasch Model. *Prosiding Seminar Nasional Ilmu Pendidikan*, *2*(1), 448–458. http://e-jurnal.fkip.unila.ac.id/index.php/psnip/article/view/286

Driana, E., Susilowati, A., Ernawati, E., & Ghani, Abd. R. A. (2021). Assessing Students' Higher-Order Thinking Skills: Knowledge and Practices of Chemistry Teachers in Vocational Senior Secondary Schools. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *27*(1), 37–47. https://doi.org/10.21831/jptk.v27i1.32882

Emiliasari, R. N. (2018). An Analysis of Teachers' Pedagogical Competence in Lesson Study of MGMP SMP Majalengka. *ELTIN JOURNAL, Journal of English Language Teaching in Indonesia*, *6*(1), 22–33. https://doi.org/10.22460/eltin.v6i1.p22-33

Erlina. (2011). Deskripsi kemampuan berpikir formal mahasiswa pendidikan kimia universitas tanjungpura. *Jurnal Visi Ilmu Pendidikan*, *6*(3), 631–640. https://jurnal.untan.ac.id/index.php/jvip/article/download/56/55

Ernawati, I. (2017). Uji kelayakan media pembelajaran interaktif pada mata pelajaran administrasi server. *Elinvo (Electronics, Informatics, and Vocational Education)*, *2*(2), 204–210. https://doi.org/10.21831/elinvo.v2i2.17315

Febriyani, R. A., Yunita, W., & Damayanti, I. (2020). An Analysis on Higher Order Thinking Skill (HOTS) in Compulsory English Textbook for the Twelfth Grade of Indonesian Senior High Schools. *Journal of English Education and Teaching*, *4*(2), 170–183. https://doi.org/10.33369/jeet.4.2.170-183

Ghazali, I. (2016). *Aplikasi Analisis Multivariete Dengan Program IBM SPSS 23* (VIII). Badan Penerbit Universitas Diponegoro.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Harahap, M. R. (2016). Sel Elektrokimia: Karakteristik dan Aplikasi. *CIRCUIT: Jurnal Ilmiah Pendidikan Teknik Elektro*, *2*(1), 177–180. https://doi.org/10.22373/crc.v2i1.764

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

Hartini & Sulitdjo. (2015). Pengembangan higher order thinking multiple choice test untuk mengukur keterampilan berpikir kritis IPA Kelas VII SMP/MTs. *Jurnal Inovasi Pendidikan IPA*, *1*(1), 86–101. https://doi.org/10.21831/jipi.v1i1.4535

Herlanti, Y., Mardiati, Y., Rahmawati, R., Putri, A. M. K., Jamil, N., Miftahuzzakiyah, M., Sofyan, A., Zulfiani, Z., & Sugiarti, S. (2019). Finding Learning Strategy in Improving Science Literacy. *Jurnal Penelitian Dan Pembelajaran IPA*, *5*(1), 59–71. https://doi.org/10.30870/jppi.v5i1.4902

Intan, F. M., Kuntarto, E., & Alirmansyah, A. (2020). Kemampuan Siswa dalam Mengerjakan Soal HOTS (Higher Order Thinking Skills) pada Pembelajaran Matematika di Kelas V Sekolah Dasar. *JPDI (Jurnal Pendidikan Dasar Indonesia)*, *5*(1), 6–10. https://doi.org/10.26737/jpdi.v5i1.1666

Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(1), 12–23. https://doi.org/10.21831/pep.v22i1.15609

Istijabatun, S. (2008). Pengaruh Pengetahuan Alam Terhadap Pemahaman Mata Pelajaran Kimia. *Jurnal Inovasi Pendidikan Kimia*, *2*(2), 323–329. https://journal.unnes.ac.id/nju/index.php/JIPK/article/download/1261/1312

Jayanti, E. (2020). Instrumen Tes Higher Order Thinking Skill Pada Materi Kimia SMA. *Orbital: Jurnal Pendidikan Kimia*, *4*(2), 135–149. https://doi.org/10.19109/ojpk.v4i2.4368

Khaldun, I., Hanum, L., & Utami, S. D. (2020). Pengembangan Soal Kimia Higher Order Thinking Skills Berbasis Komputer dengan Wondershare Quiz Creator Materi Hidrolisis Garam dan Larutan Penyangga. *Jurnal Pendidikan Sains Indonesia*, *7*(2), 132–142. https://doi.org/10.24815/jpsi.v7i2.14702

Kurniasi, E. R., Y, Y., & Karennisa, F. (2020). Analisis Soal Ulangan Harian Matematika Kelas IX SMP Negeri 1 Toboali. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, *12*(1), 43–52. https://doi.org/10.37640/jip.v12i1.276

Lilia, L., & Widodo, A. T. (2014). Implementasi Pembelajaraan Kontekstual dengan Strategi Percobaan Sederhana Berbasis Alam Lingkungan Siswa Kelas X. *Jurnal Inovasi Pendidikan Kimia*, *8*(2), 1351–1359. https://journal.unnes.ac.id/nju/index.php/JIPK/article/download/4440/3804

Lubis, A. R., Sudrajat, A., Department of Chemistry, Universitas Negeri Medan, Medan 20221, Indonesia, Nugraha, A. W., & Department of Chemistry, Universitas Negeri Medan, Medan 20221, Indonesia. (2022). Development of Higher Order Thinking Skill (HOTS) Assessment Instrument for Salt Hydrolysis. *Jurnal Pendidikan Kimia*, *14*(1), 63–69. https://doi.org/10.24114/jpkim.v14i1.31382

Maryati, M., Prasetyo, Z. K., Wilujeng, I., & Sumintono, B. (2019). Measuring teachers' pedagogical content knowledge using many-facet RASCH model. *Jurnal Cakrawala Pendidikan*, *38*(3), 452–464. https://doi.org/10.21831/cp.v38i3.26598

Mulia. (2020). Pemanfaatan ICT dalam Peningkatan Kompetensi Pedagogik Guru Bersertifikasi di Sekolah Menengah Pertama Aceh Barat. *Jurnal MUDARRISUNA: Media Kajian Pendidikan Agama Islam*, *10*(1). https://doi.org/10.22373/jm.v10i1.6791

Mulyanti, S., Sukmawati, W., & Tarkin, N. E. H. (2022). Development of items in Acid-Base Identification Experiments Using Natural Materials: Validity Test with Rasch Model Analysis. *Phenomenon : Jurnal Pendidikan MIPA*, *12*(1), 17–30. https://doi.org/10.21580/phen.2022.12.1.10703

Nadliroh, N., Saputro, S. D., Sari, A. K., Ningsih, P. R., & Wijaya, E. Y. (2019). Pengembangan Instrumen Penilaian Berbasis HOTS (High Order Thinking Skill) Mata Pelajaran Sistem Komputer Kelas X SMA Negeri 2 Bangkalan. *Reorientasi Profesionalisme Pendidik Dalam Menghadapi Tantangan Revolusi Industri 4.0*. Prosiding Seminar Nasional Pendidikan dan Pembelajaran 2019. https://seminar.umpo.ac.id/index.php/SNPP2019/article/view/334/331

Nasution, S., & Mursell, J. (2008). *Mengajar dengan Sukses*. Bumi Aksara.

Ngadi. (2023). Analisis model rasch untuk mengukur kompetensi pengetahuan siswa SMAKN 1 Kalianget pada mata pelajaran perawatan sistem kelistrikan sepeda motor. *Jurnal Pendidikan Vokasi Otomotif,* *6*(1), 1–19. http://dx.doi.org/10.21831/jpvo.v6i1.63479

Ocy, D. R., Rahayu, W., & Makmuri, M. (2023). RASCH model analysis: Development of HOTS-based mathematical abstraction ability instrument according to Riau Islands culture. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(4), 3542. https://doi.org/10.24127/ajpm.v12i4.7613

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

OECD. (2023b). *PISA 2022 Results (Volume II): Learning During – and From – Disruption*. OECD. https://doi.org/10.1787/a97db61c-en

Perera, C. J., Sumintono, B., & Jiang, N. (2018). The Psychometric Validation Of The Principal Practices Questionnaire Based On Item Response Theory. *International Online Journal of Educational Leadership*, *2*(1), 21–38. https://doi.org/10.22452/iojel.vol2no1.3

Pratikno, Y., Hermawan, E., & Arifin, A. L. (2022). Human Resource 'Kurikulum Merdeka' from Design to Implementation in the School: What Worked and What not in Indonesian Education. *Jurnal Iqra': Kajian Ilmu Pendidikan*, *7*(1), 326–343. https://doi.org/10.25217/ji.v7i1.1708

Pratiwi, S. N., Cari, C., & Aminah, N. S. (2019). Pembelajaran IPA Abad 21 dengan Literasi Sains Siswa. *Jurnal Materi dan Pembelajaran Fisika (JMPF)*, *9*(1), 34–42. https://jurnal.uns.ac.id/jmpf/article/view/31612https://jurnal.uns.ac.id/jmpf/article/view/31612

Puspitasari, L., Alwi, Z., Suhardi, S., & Zulkardi, Z. (2021). Literacy Questions Based on Hots Using Context Covid-19. *Journal of Education Research and Evaluation*, *5*(4), 624–632. https://doi.org/10.23887/jere.v5i4.33870

Rahielanamy, & Kamaludin, A. (2022). Development of Higher Order Thinking Skill Qualified Test on Colligative Properties of Solutions. *Jurnal Pendidikan Sains Indonesia*, *10*(2), 294–311. https://doi.org/10.24815/jpsi.v10i2.23968

Ridwan, I. M., Kaniawati, I., Suhandi, A., Ramlan, T., Rizal, R., & Sujarwanto, E. (2023). *Instrumen Kemampuan Pemecahan Masalah pada Tema Perubahan Iklim: Analisis Rasch Model*.

Rofifah, E., Aminah, N. S., & Ekawati, E. Y. (2013). Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika Pada Siswa SMP. *Jurnal Pendidikan Fisika*, *1*(2), 17–22. https://jurnal.fkip.uns.ac.id/index.php/pfisika/article/download/2797/1913

Roy, S. (2016). Chemistry in Our Daily Life: Preliminary Information. *International Journal of Home Science*, *2*(3), 361–366. https://www.homesciencejournal.com/archives/2016/vol2issue3/PartF/2-3-26.pdf

Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

Rudibyani, R. B., Perdana, R., & Elisanti, E. (2020). Development of Problem-Solving-Based Knowledge Assessment Instrument in Electrochemistry. *International Journal of Instruction*, *13*(4), 957–974. https://doi.org/10.29333/iji.2020.13458a

Setiawan, J., Ajat Sudrajat, A., Aman, A., & Kumalasari, D. (2021). Development of Higher Order Thinking Skill Assessment Instruments in Learning Indonesian History. *International Journal of Evaluation and Research in Education (IJERE)*, *10*(2), 545. https://doi.org/10.11591/ijere.v10i2.20796

Sibarani, R. A. M., & Tenriawaru, A. B. (2019). Pentingnya Literasi Sains Bagi Siswa di Era Revolusi Industri 4.0. *Prosiding Seminar Nasional FKIP 2019*, 9.

Smith Jr., E. V. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, *2*(3), 281–311.

Stojanovska, M. I., Soptrajanov, B. T., & Petrusevski, V. M. (2012). Addressing Misconceptions about the Particulate Nature of Matter among Secondary-School and High-School Students in the Republic of Macedonia. *Creative Education*, *03*(05), 619–631. https://doi.org/10.4236/ce.2012.35091

Sukmawati, W. (2019). Analisis Level Makroskopis, Mikroskopis dan Simbolik Mahasiswa dalam Memahami Elektrokimia. *Jurnal Inovasi Pendidikan IPA*, *5*(2), 195–204. https://doi.org/10.21831/jipi.v5i2.27517

Sukmawati, W. (2020). *Redoks dan elektrokimia*. Bintang Pustaka Madani.

Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning. *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)*. 1st International Conference on Education Innovation (ICEI 2017), Surbaya, Indonesia. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model rasch: Untuk penelitian ilmu-ilmu sosial*. Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata Publishing House.

Sunyono, Yuanita, L., & Ibrahim, M. (2015). Supporting Students in Learning with Multiple Representation to Improve Student Mental Models on Atomic Structure Concepts. *Science Education International*, *26*(2), 104–125. http://www.icaseonline.net/sei/june2015/p1.pdf

Syahana, S., & Andromeda, A. (2021). Pengembangan Instrumen Tes Two Tier Berbasis Higher Order Thinking Skills pada Materi Sifat Koligatif Larutan untuk Siswa SMA/MA. *EDUKATIF : Jurnal Ilmu Pendidikan*, *3*(3), 1050–1058. https://doi.org/10.31004/edukatif.v3i3.505

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Talanquer, V. (2011). Macro, Submicro, and Symbolic: The Many Faces of the Chemistry "triplet". *International Journal of Science Education*, *33*(2), 179–195. https://doi.org/10.1080/09500690903386435

Tasrif, T. (2022). Higher Order Thinking Skills (HOTS) dalam pembelajaran social studies di sekolah menengah atas. *Jurnal Pembangunan Pendidikan: Fondasi Dan Aplikasi*, *10*(1). https://doi.org/10.21831/jppfa.v10i1.29490

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Vol. 10, No. 2 : June 2024
E-ISSN: 2442-7667
pp. 451-466
Email: jklppm@undikma.ac.id

Tessmer, M. (2013). *Planning and Conducting Formative Evaluations* (0 ed.). Routledge. https://doi.org/10.4324/9780203061978

Toharudin, U., Hendrawati, S., & Rustaman, A. (2011). *Membangun Literasi Sains Peserta Didik*. Humaniora.

Wati, S., Syafryadin, S., & Apriani, E. (2021). Learning module development on compiling exposition and argumentation text using project-based learning. *English Review: Journal of English Education*, *9*(2), 355–366. https://doi.org/10.25134/erjee.v9i2.4370

Widana, W. (2017). *Penyusunan Soal High Order Thinking Skills*. Departemen Pendidikan dan Kebudayaan.

Widhiyani, I. A. N. T., Sukajaya, I. N., & Suweken, G. (2019). Pengembangan soal higher order thinking skills untuk pengkategorian kemampuan pemecahan masalah geometri siswa SMP. *Jurnal Pendidikan dan Pembelajaran Matematika Indonesia*, *8*(2), 68–77. https://doi.org/10.23887/jppm.v8i2.2854

Wulandari, T., Ramli, M., & Muzzazinah. (2022). Analisis Butir Soal Dynamic Assessment untuk Mengukur Pemahaman Konsep Klasifikasi Tumbuhan pada Mahasiswa. *Jurnal Pendidikan Sains Indonesia*, *10*(1), 191–201. https://doi.org/10.24815/jpsi.v10i1.22082

Yuliati, Y. (2017). Literasi Sains dalam Pembelajaran IPA. *Jurnal Cakrawala Pendas*, *3*(2), 21–28. https://doi.org/10.31949/jcp.v3i2.592

Yuriza, P. E., & Sigit, D. V. (2018). Hubungan Antara Kemampuan Berpikir Tingkat Tinggi dan Tingkat Kecerdasan dengan Kemampuan Literasi Sains Pada Siswa SMP. *BIOSFER: JURNAL PENDIDIKAN BIOLOGI (BIOSFERJPB)*, *11*(1), 13–20. https://doi.org/10.21009/biosferjpb.11-1.2

Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah : Jurnal Ilmiah Kependidikan*, *7*(1), 17–23. https://doi.org/10.18592/tarbiyah.v7i1.2100

Zohrany, E., Suryati, & Khery, Y. (2018, September). Pengembangan bahan ajar laju reaksi dengan modelinquiry berbasis android untuk menumbuhkan literasi sains siswa. *Membangun Pendidikan Yang Mandiri Dan Berkualitas Pada Era Revolusi Industri 4.0*. Prosiding Seminar Nasional Lembaga Penelitian dan Pendidikan (LPP) Mandala. http://ejournal.mandalanursa.org/index.php/Prosiding/article/view/396/382