



# Journal of Educational Sciences

Journal homepage: <https://jes.ejournal.unri.ac.id/index.php/JES>



P-ISSN  
2581-1657

E-ISSN  
2581-2203

## Development of Test Instrument to Assess Students' Understanding on Macroscopic, Sub-Microscopic, and Symbolic Levels of Acid-Base Topics Using Rasch Model

Arifah Sulvianti\*, Faizah Qurrata Aini

Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Padang, 25131, Indonesia

### ARTICLE INFO

#### Article history:

Received: 08 Aug 2022

Revised: 02 April 2023

Accepted: 04 April 2023

Published online: 24 April 2023

#### Keywords:

Acid-Base;

Multiple Representation;

Test Instrument;

Rasch Model

### ABSTRACT

Students' conceptual understanding on acid-base topic has not been assessed comprehensively in multiple representation. The test instrument is needed to determine levels of students' understanding on acid-base topic comprehensively. This research developed an instrument to assess students' understanding on acid-base concept based on macroscopic, sub-microscopic and symbolic levels that is claimed valid, reliable, had good difficulty index and discriminatory item. This research is Research and Development (R&D) using Rasch model. The research procedure adopted ten stages by Wei, et al in 2012. The research instrument is questionnaire sheet given to subjects from chemistry lecturers and chemistry teacher. The data were analyzed using Rasch model assisted by minifacet and ministep software. The exact agreement value is 89,2% which have small range from the expect agreement with a percentage of 89,5%. It's mean that, there is a fit between the results of expert's agreement and the results of estimated model. The pilot test results showed that all items have reached valid criteria based on the MNSQ, ZSTD, and Pt-Mean Corr. The reliability value is 0,80 in the good category. Difficulty index has variations of easy, medium, and difficult items. The discriminatory item is able to distinguish three levels of student ability.

## 1. Introduction

One of the primary subjects for seniors high school is chemistry. A multirepresentational approach that incorporates macroscopic, sub-microscopic, and symbolic levels that have linkages to each other and are related in chemistry learning might help students to understand the numerous abstract concepts that make up chemistry (Zahro' & Ismono, 2021). Only macroscopic level of three representation levels can be directly observed by the senses. Concepts at the level

\* Corresponding author.

E-mail: [faizah\\_qurrata@fmipa.unp.ac.id](mailto:faizah_qurrata@fmipa.unp.ac.id)

Doi: <https://doi.org/10.31258/jes.7.2.p.282-295>

of particulates, such as molecules, atoms, or particles, are explained at the sub-microscopic level. On the other hand, the symbolic level describes chemical phenomena in the form of symbols like reaction equations, diagrams, etc. If students can comprehend and connect the three levels, their understanding will be thorough and in-depth (Jariati & Yenti, 2020). In other words, all chemical concept and theories are built by three levels of representation which can not be separated from each other to achieve a comprehensive understanding. This is the characteristic of chemistry that causes majority of students think that chemistry is a difficult subject. One of the subject matter of chemistry with abstract and fundamental concepts is acid-base matter.

Preliminary research at four schools in Padang city indicates that student have trouble understanding about acid and base matter. The majority of student failing to reach the minimal completion requirement (KKM) on the daily acid-base test value serves as evidence of this. It can be assumed that students have not been able to fully describe chemical concepts at three levels of representation and their understanding is more dominant at symbolic than macroscopic and sub-microscopic levels. Besides that, most teachers have taught the concept of acid and base through a multiple representation approach.

The learning media used such as power point, student worksheet and textbooks have also been equipped with pictures and explanations at the macroscopic, sub-microscopic, and symbolic levels. However, the test instruments used by teachers have not been able to assess students' understanding comprehensively at all three levels of representation. The test instrument are only focused on symbolic understanding. So that the assessment process carried out by the teacher becomes not in related with what has been taught by students. Through assessment, information can be obtained about how well successful students are in learning and teachers in teaching students, so that it can be a feedback among them. In this case, assessment can be viewed as quality control of the learning activities carried out (Sudarmin, 2015).

Due to the unavailability of test instruments to test students' understanding at all three levels of representation, a test instrument was developed that could asses students' understanding of acid-base material comprehensively. In order to produce a good and quality test instrument in terms of validity, reliability, difficulty index, and discriminatory item, a Rasch model analysis is needed that can provide information about the characteristics of the items in the developed test instrument. The Rasch model was chosen because it has the advantage of accommodating measurement objectives more accurately (Bohori & Liliawati, 2019). This study aims to produced an instrument to asses students' understanding on acid-base concept based on macroscopic, sub-microscopic and symbolic levels using Rasch model that is claimed valid, reliable, had good difficulty index and discriminatory item.

---

## 2. Methodology

Type of research is research and development (R&D) using Rasch model (Wei et al., 2012). The stage of this study adopted ten stages of instrument development by Wei et al in 2012. The stages are (1) determine construct; (2) identify specified construct; (3) determine item result; (4) conduct pilot testing; (5) analyze data with Rasch model; (6) review item fit; (7) review Wright Map; (8) repeat steps 4–7 until it fits the model; (9) establish claims about the quality of items; and (10) Develop a documentation (Wei et al., 2012).

The subjects in this study consist of five chemistry lecturers from FMIPA UNP, one chemistry teacher from high school 3 Padang, and 30 students 12<sup>th</sup> grade of high school 3 Padang who had studied acid-base material. The objects of this study is the quality of instrument in terms of validity, reliability, difficulty index, and discriminatory items produced. The type of data is primary data obtained directly through content validity tests by validator/experts and question trial data directly to students. The research instrument is a content validity questionnaire in the form of Guttman scale with a choice "Yes" or "No". The data analysis technique of results was carried out using Rasch model with help of minifacet and ministep programs.

The quality of items in terms of four criteria, namely validity, reliability, difficulty index, and discriminatory item. The validity of item is reviewed from Outfit Mean Square (MNSQ) value with a range accepted from 0.5 to 1.5, the Outfit Z-standard (ZSTD) value from -2.0 to +2.0, and the Point Measure Correlation (Pt Mean Corr) value accepted from 0.4 to 0.85. In measuring the reliability of items, it is based on the value of item reliability with a good category  $\geq 0.8$ . The difficulty index give information about standard deviation and logit value for each item which can be seen from distribution of difficulty item on Wright map, specifically can be seen in output item measure. The discriminatory item can be known from the separation value. The separation (H value) can be determined by the following equation.

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

If the separation value is bigger, so the discriminatory items and respondents is better, because it can distinguish between items group and respondents.

## 3. Results and Discussion

### *Define Construct*

The first stage is carried out by making a learning progression, which includes the analysis of basic competencies (KD) from 2013 Curriculum which is derived into several competency achievement indicators (IPK) on acid-base material. The basic competency is KD 3.10, explain the concept of acids and bases and their

---

strength and ionization equilibrium in solution. And then the analysis of multiple representation aspects to find out the macroscopic, sub-microscopic, and symbolic representations contained in acid-base concept to be assessed. Table 1 shows the analysis of learning progression.

Table 1. Learning Progression

Main IPK	Cognitive level	Representation
3.10.1 Explain the concept of acid-base according to Arrhenius. (Mind IPK)	C2	<i>Sub-microscopic</i> , the ionization process of acids in water so as to produce hydronium ions ( $\text{H}_3\text{O}^+$ ); the ionization process of bases in water so as to produce hydroxide ions ( $\text{OH}^-$ ). <i>Symbolic</i> , the combination of acid and base ionization reactions in water.
3.10.2 Explain the concept of acid-base according to Bronsted-Lowry. (Mind IPK)	C2	<i>Symbolic</i> , the combination of acid and base reactions involving the transfer of protons.
3.10.3 Explain the concept of acid-base according to Lewis. (Mind IPK)	C2	<i>Symbolic</i> , the combination of reactions involving the transfer of electron pairs in the formation of coordination covalent bonds.
3.10.4 Explain the effect of acid or base ionization on water equilibrium. (Mind IPK)	C2	<i>Macroscopic</i> , pH value; changes in the pH of the water after adding acids and bases. <i>Sub-microscopic</i> , the effect of addition of acid or base to water equilibrium; difference in the number of $\text{H}^+$ ions and $\text{OH}^-$ ions in acid and base solutions. <i>Symbolic</i> , equation of water equilibrium reactions before and after the addition of acidic and base substances; comparison of $\text{H}^+$ and $\text{OH}^-$ concentrations before and after the addition of acids to water; comparison of $\text{H}^+$ and $\text{OH}^-$ concentrations before and after the addition of bases to water.
3.10.5 Explain the strength of acids and bases. (Mind IPK)	C2	<i>Macroscopic</i> , measuring the strength of acids and bases using an electrolyte tester (ex: strong acids and bases are able to flame the lamp brightly and produce many gas bubbles, weak acids and weak bases are able to dim the lamp and produce a small amount of gas bubbles). <i>Sub-microscopic</i> , ionization process of acids or bases in water (strong acid and base are 100% ionized in water, weak acids and bases are partially ionized in water) <i>Symbolic</i> , ionization reaction equation of strong acids and strong bases in water; ionization reaction equation of weak acids and weak bases in water ( <i>forward and reverse reaction</i> ; acid ionization constant ( $K_a$ ), base ionization constant ( $K_b$ )).

### *Identify Defined Constructs*

This stage is carried out involves creating a test instrument grids that includes KD, IPK, and item indicators that have chemical multiple representation elements and choosing the form of questions to be developed. The item indicators are designed based on main IPK which has been analyzed for its chemical multiple representation aspects. There are 24 item indicators based on IPK analysis, and the form of question chosen is an essay test. The essay test was chosen because of avoiding the possibility of guessing answer by students and essay questions also require the ability of students to interpret their ideas, so that their cognitive understanding can be tested on aspects of chemical multirepresentation in more depth. Then, assessment in the form of essay (description) can produce higher item information function values (Diputera, 2019).

### *Define Outcome Space of Items*

The design of items and assessment rubrics is carried out based on item indicators of test instruments and level understanding of chemical multirepresentation, which is an important point for making this instrument. The analysis of the grid with 10 items discourses produced in 24 sub-question items which have macroscopic, sub-microscopic, or symbolic related to each question. One set of test instruments consist of question discourse, sub-questions, and assessment rubrics. In the assessment rubric, there are answer keys, the achievement of students' level understanding of chemical multirepresentation, and scoring guidelines. Table 2 is a description of level understanding chemical multirepresentation.

Table 2. Level of Understanding Chemical Representation

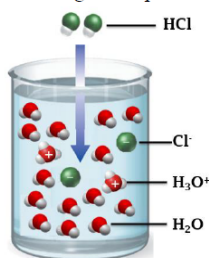
<b>Level of understanding</b>	<b>Information</b>
Level 1: linking symbolic understanding with macroscopic (Identifying colors, phenomena occurring and; writing the symbols of the reaction equation correctly).	Students are able to answer questions at the level of macroscopic and symbolic understanding.
Level 2: understanding sub-microscopic understanding with symbolic (Knowing the shape of the structure of atoms, molecules, and ions; writing the symbols of the reaction equation correctly).	Students are able to answer questions at the level of sub-microscopic and symbolic understanding.
Level 3: understand and interpret macroscopic, sub-microscopic and symbolic understanding with macroscopic (Identifying colors, phenomena occurring and knowing the shape of the structure of atoms, molecules, and ions and being able to write symbols of reaction equations correctly; stating and explaining macroscopic phenomena and processes from sub-microscopic).	Students are able to answer questions at the level of macroscopic, sub-microscopic, and symbolic; or macroscopic and sub-microscopic understanding.

(Wang et al., 2017)

Each question has an interconnected level chemical representation in the sub-items of question. Figure 1 is an example of question design that tests understanding of sub-microscopic and symbolic levels. The IPK achieved was

explaining the concept of Arrhenius acids and bases. Sub-point 1a tests conceptual understanding at the sub-microscopic level. Students are required to explain about Arrhenius acids theory based on sub-microscopic illustration of acid ionization process given. Whereas question 1b requires students to write down the equation for the ionization reaction of Arrhenius acid, which is knowledge at the symbolic level. Through these two sub-items, students' understanding can be tested comprehensively on Arrhenius acids and bases concept especially.

1. Perhatikan gambar proses ionisasi suatu zat di dalam air berikut ini!



Sumber: Tro, N (2012)

a. Apakah zat tersebut asam atau basa? Jelaskan alasannya menurut teori Arrhenius!

.....  
 .....

b. Tuliskanlah persamaan reaksi ionisasi dari zat tersebut!

.....  
 .....

Figure 1. The Sample of Item Developing

Furthermore, the instrument was validated by six validators who were subject matter experts. Content validity uses Guttman scale with choice “Yes or No” for the assessment aspects given. The data of validity were analyzed using Rasch model with minifacet software program. The analysis of content validity by experts can be seen in Figure 2.

Figure 2 is a Wright map that shows validity content data from expert's assessment with nine aspect criteria items. Based on the picture, there are four columns with different information. The first column is the measure column which shows the logit scale with a values range from -5 to +3. The second column is an explanation about quality of items from validator. It can be read by looking at the order of questions from top to bottom. The top items indicate the best quality, because they are achieve most criteria in assessment aspect. The third column is code for assessment aspect criteria. It describes the order of difficulty aspects. The top aspect is the most difficult for all items to achieve. While the bottom aspects are the easiest for item to achieve according to assessment given by validator. Next, the fourth column is the validator column which explains the order of the validator's assessment.

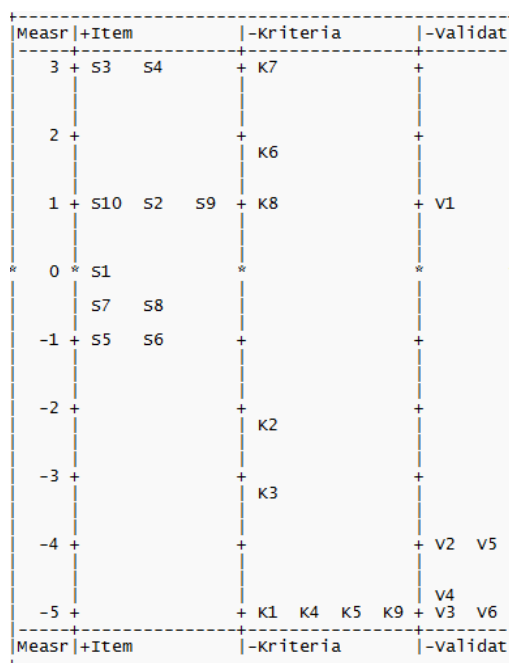


Figure 2. Wright Map of Expert and Items

Based on the results of content validity analysis using Rasch model which consists of 10 essay questions, there are 2 questions with the best quality are obtained which are in the top position, namely item 3 and 4. Item 3 and 4 have reached all the criteria for the assessment aspect because the position is higher than all aspects on the logit scale. Furthermore, in the column aspect criteria, there are several criteria that are positioned higher than some items. It means that these items have not achieved all criteria for assessment aspect. Besides that, it also describes the most difficult criteria to achieve, namely criteria "K7". Besides the difficult criteria, there are four criteria that occupy the lowest position in logit scale, which means that these criteria are the easiest to achieve by all questions, namely the K1, K4, K5, and K9 criteria.

The strata value obtained from the expert’s assessment is equal to 3,37 which indicates that the validator’s assessment achieve the reliable criteria. Meanwhile, the reliability value of the validator obtained was 0.84 with a very good category. This reliability value indicates the reliability of experts in providing assessments on items (Nisa & Yusmaita, 2022). The third column describes exact agreement value with a percentage of 88.5%, which is not so far from expert agreement (model estimate) with a percentage of 88.8%. This means that there is a fit between the results of expert assessment and estimated by the model (Eliza & Yusmaita, 2021). Table 3 shows a summary of the result from expert assessment analysis.

Table 3. Summary of The Result From Expert Assessment Analysis

Strate value	Reliability	Exact agreement	Expert agreement
3,37	0,84	88,5%	88,8%

### **Conduct Pilot Testing**

The validated products, were tested to 30 students 12<sup>th</sup> grade high school 3 Padang who had learned acid-base material. According to (Sumintono & Widhiarso, 2013), this sample size achieved the minimum requirements for testing an essay test using the Rasch model. In order to avoid high bias scores, before students are given test, they are first taught the acid-base material so that students can remember the material again. If the students are ready, so they are asked to answer on 24 sub-questions of essay within 60 minutes individually. The student's answer data is corrected and given a score based on the reference of multirepresentation understanding and scoring guidelines that have been created.

### **Analyze Data using Rasch Model**

Student's score data were analyzed using Rasch model with the helping of ministep software. The quality of items reviewed included validity, reliability, difficulty index and item discriminatory.

#### **1. Validity of Items**

The items validity can be determined from output items fit order with 3 fit criteria, namely Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point Measure Correlation (Pt-Mean Corr). According to (Planinic et al., 2019), from these three criteria, not all of them must achieve the "accepted" value for an item to be said as "valid". If only one criteria is achieved, so the item can still be said to be valid. However, in the results of this analysis there are no items that only achieve one criteria. The minimum criteria to be achieved are as many as 2 from 3 fit criteria for each items. Meanwhile, the most difficult criteria to achieve were Outfit Pt-Mean Corr. Even though there are 15 items less than acceptable limit for this Outfit, but the Outfit ZSTD and MNSQ Outfit scores can be achieved by other items. Table 4 is the item fit order which shows the result of content validity analysis according Rasch model.

Table 4. Item Fit Order

Number	Number of item	Outfit		Pt Mean Corr	Information
		MNSQ	ZSTD		
1	8c	1.37	0.80	0.21	Accepted
2	10b	0.83	-0.11	0.40	Accepted
3	5c	1.39	1.52	0.48	Accepted
4	6c	1.28	1.01	0.43	Accepted
5	4b	1.13	0.46	0.50	Accepted
6	3b	1.18	0.52	-0.02	Accepted
7	8b	1.13	0.68	0.05	Accepted
8	4a	1.11	0.48	0.31	Accepted
9	5a	0.94	0.06	0.16	Accepted
10	6b	1.03	0.21	0.14	Accepted
11	6a	0.98	0.20	0.07	Accepted
12	8a	0.99	0.01	0.22	Accepted
13	9b	0.83	-0.25	0.27	Accepted
14	9a	0.96	-0.29	0.30	Accepted
15	7b	0.88	-0.29	0.33	Accepted
16	7c	0.72	-0.40	0.39	Accepted



17	1a	0.80	-0.46	0.40	Accepted
18	7a	0.88	-0.24	0.37	Accepted
19	3a	0.88	-0.36	0.39	Accepted
20	2a	0.66	-0.52	0.45	Accepted
21	10a	0.84	-0.64	0.47	Accepted
22	1b	0.65	-0.53	0.48	Accepted
23	2b	0.31	-0.54	0.51	Accepted
24	5b	0.75	-0.69	0.20	Accepted

**2. Reliability of Items**

Reliability analysis on Output Summary Statistics in Rasch model provides information on the Cronbach's Alpha value of 0.83. Cronbach's Alpha value is used to measure overall reliability by looking at the interaction between person and item (Widhiarso & Sumintono, 2015). Reliability explains whether an instrument provides the same or consistent information if repeated measurement tests are carried out (Pratama, 2020). With a Cronbach Alpha value of  $\geq 0.8$ , it means that the reliability of the instrument is in very good category. In addition, to know the reliability of items, it can see at the "item reliability" value. Figure 3 shows the result of Rasch's analysis about item reliability.

CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .83 SEM = 3.18								
SUMMARY OF 24 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	68.2	27.8	.00	.46	1.02	.14	.94	.03
SEM	6.6	.5	.24	.05	.04	.11	.05	.12
P. SD	31.7	2.6	1.16	.24	.19	.53	.24	.57
S. SD	32.4	2.6	1.19	.24	.19	.54	.24	.58
MAX.	127.0	30.0	2.65	1.03	1.55	1.33	1.39	1.52
MIN.	21.0	21.0	-2.13	.17	.78	-.75	.31	-.69
REAL RMSE	.52	TRUE SD	1.04	SEPARATION	1.99	Item	RELIABILITY	.80
MODEL RMSE	.51	TRUE SD	1.04	SEPARATION	2.03	Item	RELIABILITY	.80
S.E. OF Item	MEAN = .24							

Figure 3. Summary Statistic

**3. Item Discriminatory**

The discriminatory item explains how good the level of item to compare individuals who have high and low abilities. The different power of items is also analyzed using the Output Summary Statistics in Figure 3. The grouping of different power items can be viewed from the value of separation. If the separation value is greater, so the quality of differentiating power of all items and respondents is also better, because it can distinguish groups of items and respondents (Widhiarso & Sumintono, 2015). With a separation value of 1.99 then  $H = [(4 \times 1.99) + 1] / 3 = 2.98$ . The number 2.98 is rounded to 3. This means that there are 3 groups of items, namely easy, medium, and difficult.

**4. Difficulty Index**

Analysis of difficulty index can be viewed from the Output Item Measure which is presented in Figure 4. The item difficulty index is obtained from a combination of the mean value and the Standard Deviation (SD) value. The average logit value is 0.00 and the standard deviation is 1.16. Then the average value of the logit measure is 1.16 logit. Figure 4 shows the group of item difficulty index.

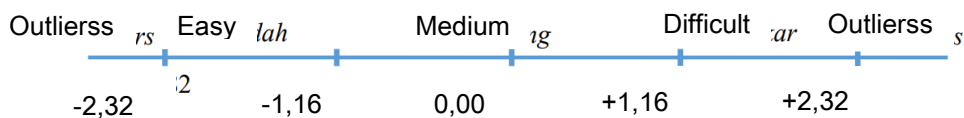


Figure 4. Variation of Item Difficulty Index

Based on the results of item measure analysis, one outlier item was obtained, item 7a with a logit value of +2.65 because it exceeded the +2.32 logit value. Besides that, no other outlier items were found, so the difficulty index of the items can be grouped, consisting of variations that are easy, medium, difficult, until very difficult (Palimbong et al., 2018). Figure 5 is item measure that shows the difficulty index analysis by Rasch model.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.	INFIIT MNSQ	INFIIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Item	G
15	28	23	2.65	.52	.90	-.26	.88	-.24	.37	.20	78.3	78.3	7a	B
8	62	21	2.13	.22	1.25	.80	1.13	.46	.50	.43	14.3	35.5	4b	E
18	41	30	1.90	.39	1.00	.05	.99	.01	.22	.22	66.7	64.7	8a	B
5	104	23	.97	.17	.89	-.40	.88	-.36	.39	.53	26.1	26.2	3a	E
21	46	29	.97	.39	.97	-.26	.96	-.29	.30	.24	72.4	62.3	9a	B
14	111	24	.79	.17	1.13	.60	1.28	1.01	.43	.51	16.7	22.5	6c	E
19	21	30	.43	.41	1.07	.48	1.13	.68	.05	.22	73.3	70.8	8b	A
10	98	29	.36	.22	.78	-.75	.75	-.69	.20	.33	48.3	41.3	5b	C
11	116	30	.33	.18	1.30	1.33	1.39	1.52	.48	.45	13.3	31.2	5c	D
23	50	29	.32	.43	.88	-.59	.84	-.64	.47	.22	75.9	73.0	10a	B
7	110	28	.29	.19	1.00	.06	1.11	.48	.31	.46	42.9	33.1	4a	D
16	24	30	-.14	.47	.94	-.12	.88	-.29	.33	.20	80.0	80.0	7b	A
13	127	29	-.26	.23	.88	-.30	1.03	.21	.14	.41	37.9	46.9	6b	D
20	103	28	-.31	.31	1.55	1.25	1.37	.80	.21	.33	71.4	73.0	8c	C
1	53	29	-.33	.50	.91	-.17	.80	-.46	.40	.19	82.8	82.8	1a	B
24	113	30	-.56	.35	1.44	.96	.83	-.11	.40	.29	83.3	81.8	10b	C
22	56	30	-.64	.55	.98	.08	.83	-.25	.27	.18	86.7	86.7	9b	B
17	51	27	-.78	.62	.92	-.02	.72	-.40	.39	.13	88.9	88.9	7c	B
9	51	27	-.85	.62	1.03	.21	.94	.06	.16	.17	88.9	88.9	5a	B
2	55	29	-.93	.62	.87	-.14	.65	-.53	.48	.16	89.7	89.7	1b	B
3	55	29	-.94	.62	.88	-.11	.66	-.52	.45	.16	89.7	89.7	2a	B
12	56	29	-1.31	.74	1.02	.24	.98	.20	.07	.11	93.1	93.1	6a	B
6	49	25	-1.96	1.03	1.04	.35	1.18	.52	-.02	.11	96.0	96.0	3b	B
4	57	29	-2.13	1.02	.87	.16	.31	-.54	.51	.10	96.6	96.6	2b	B
MEAN	68.2	27.8	.00	.46	1.02	.1	.94	.0			67.2	68.0		
P. SD	31.7	2.6	1.16	.24	.19	.5	.24	.6			26.9	23.9		

Figure 5. Item Measure

**Review Item Compatibility**

A review of the items is carried out based on analysis four criteria in previous stage. In the validity analysis, all items are fit and in accordance with the model. While the reliability analysis with a reliability item value of 0.8, it means that item reliability is good. Difficulty index analysis provides information that there are variations of items that are easy, medium, difficult, and one item is very difficult. For discriminatory items have three different power which are considered appropriate and good according to the model.

**Review Wright Map**

Wright map analysis provides comprehensive information about distribution abilities of students and the level of difficulty items on same scale. The left side of Wright map is the distribution of students' abilities, and the right side is the distribution of difficulty levels of items. There are two students with the highest position which means they have the highest ability, namely L05 and P18 with

logit values  $> +2$ . Meanwhile, students with the lowest abilities were in the lowest position, namely P12 with a logit value of 0. Besides that, the most difficult item was item 7a because it was in the top position and outside the T (outlier) limit. The items with the lowest logit value ( $\leq -2$  logit) are items 3b and 2b, but not outliers. In this case, it means that students have more correct answers to these items (Sabekti & Khoirunnisa, 2018). Figure 6 shows the distribution Wright map of students and items.

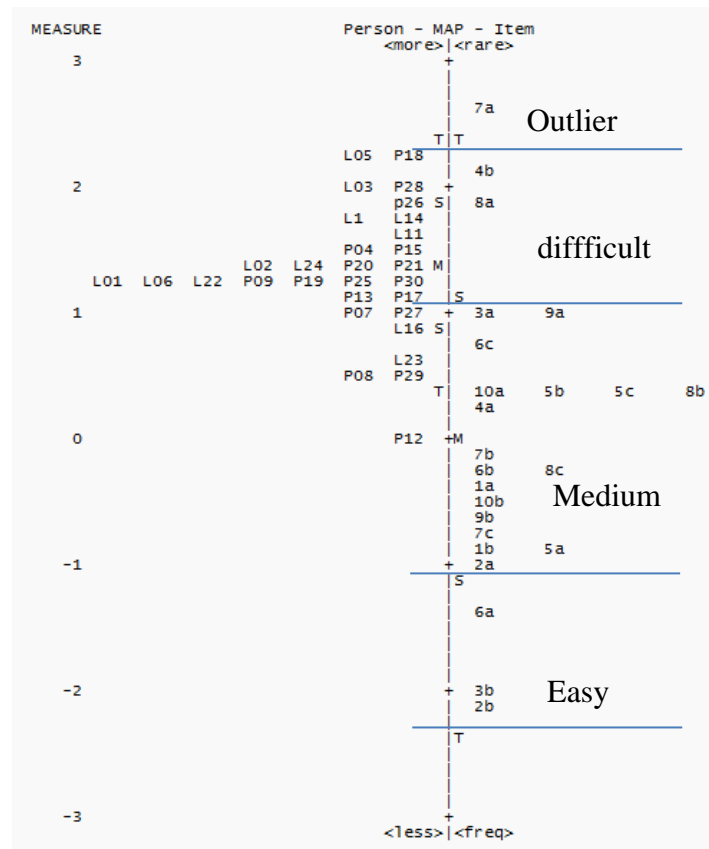


Figure 6. Wright Map of Students and Items

Based on the Wright map analysis, it was found that one item was distributed outside the T limit, namely item 7a. Item 7a assess understanding to macroscopic level. While the item that are classified as difficult include item 8a and 4b. Item 8a tests sub-microscopic understanding, and item 4b tests symbolic understanding. The items that are relatively easy are item 2b, 3b, and 6a with a logit below -1.16. Item 2b and 3b test symbolic understanding, and item 6a tests macroscopic understanding. This is indicated by the item of students who answer incorrectly on item 2b, only one subject, item 3a only four subjects, item 6a only one subject who answered incorrectly, and a small number of students answered incompletely and most answered correctly. Most of the items were considered fit and had a good and even distribution of difficulty levels. However, there is one item at very top, item 7a (outlier) which is considered for revising the items and re-testing. Even so, the distribution of the items displayed on the Wright map is

mostly in the good category, so deviation of one item (outside the T limit) in the Wright map can still be tolerated.

### ***Repeat Steps 4-7 until All Items Fit***

The expected quality of the instrument has been achieved with evidence the results of analysis validity, reliability, difficulty index and discriminatory of items that are good and in accordance with the model. However, when viewed from the Wright map analysis, there is one item outside the T (outlier) limit, namely item 7a. however, item 7a is considered still suitable for use because the distance is not too far from the previous item and person, at +2.65 logit. Although this item needs to be considered for revision and re-testing, the second trial and re-testing were not carried out due to limited research time.

### ***Establish Test Instrument Quality Claims***

The validity of all items has been detailed in Table 4. All items can be claimed to be valid because they achieved at least 2 from 3 criteria for MNSQ, ZSTD, and Pt-Mean Corr in the Item Fit Order table. As for the reliability aspect of the test instrument, it can be claimed to be reliable based on the summary statistical data in Figure 3 with item reliability value = 0,80 and Cronbach's Alpha = 0,83 in very good category. Likewise with the difficulty index and discriminatory items obtained which have 3 variations level of questions from easy, medium until difficult questions.

### ***Develop Test Instrument Documentation***

Test instrument documentation needs to be developed to provide information for teachers and students in using instrument. So that the information obtained is more complex related to the characteristics of the instrument being developed (Sabekti & Khoirunnisa, 2018). The documents required for this test instrument are in the form of learning progression, test instrument grids, items (including covers and general instructions), assessment rubrics, and guidelines for achieving students' understanding levels in macroscopic, sub-microscopic, and symbolic representations.

## **4. Conclusion**

Based on the results of this study, it can be concluded that instruments for testing understanding at the macroscopic, sub-microscopic and symbolic levels of high school students in acid base material that has been developed using Rasch model achieve the criteria of being valid, reliable, having a good difficulty index and item discrimination, as well as there is a fit between the result of the validity approval by the expert and the estimated model.

---

---

## Acknowledgments

The author would like to thank the Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, and High School 3 Padang for facilitating the author in completing this reserach.

## References

- Bohori, M., & Liliawati, W. (2019). Analisis penguasaan konsep siswa menggunakan Rasch Model pada materi usaha dan energi. *Prosiding Seminar Nasional Fisika*, 0, 138–143. <http://proceedings.upi.edu/index.php/sinafi/article/view/579>
- Diputera, A. M. (2019). Teori Penilaian Tes Essai Atau Uraian. *Journal Resapedia*, 1(1), 1.
- Eliza, W., & Yusmaita, E. (2021). Pengembangan Butir Soal Literasi Kimia pada Materi Sistem Koloid Kelas XI IPA SMA/MA. *Jurnal Eksakta Pendidikan (Jep)*, 5(2), 197–204. <https://doi.org/10.24036/jep/vol5-iss2/621>
- Jariati, E., & Yenti, E. (2020). Pengembangan E-Magazine Berbasis Multipel Representasi untuk Pembelajaran Kimia di SMA pada Materi Larutan Elektrolit dan Non Elektrolit. *Journal of Natural Science and Integration*, 3(2), 138. <https://doi.org/10.24014/jnsi.v3i2.10131>
- Nisa, D. Q., & Yusmaita, E. (2022). Pengembangan Butir Soal Literasi Kimia pada Topik Larutan Elektrolit dan Non Elektrolit Kelas X SMA / MA Development of Chemical Literacy Items in Electrolyte and Non Electrolyte Solution Topic for Class X SMA / MA. *Entalpi Pendidikan KImia*, 17, 49–57.
- Physics Education Journal. (2018). Universitas Papua. *Physics Education Journal*, 1(1), 12–21. [i.yusuf@unipa.ac.id](mailto:i.yusuf@unipa.ac.id)
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 20111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61–70. <https://doi.org/10.32923/tarbawy.v7i1.1187>
- Sabekti, A. W., & Khoirunnisa, F. (2018). Penggunaan Rasch Model Untuk Mengembangkan Instrumen Pengukuran Kemampuan Berikir Kritis Siswa Pada Topik Ikatan Kimia. *Jurnal Zarah*, 6(2), 68–75. <https://doi.org/10.31629/zarah.v6i2.724>
- Sudarmin, S. wardani I. (2015). Pengembangan Asesmen Autentik Pada Materi Interaksi Makhluk Hidup Dengan Lingkungan Untuk Meningkatkan Kemampuan Berpikir Kritis Siswa. *Unnes Science Education Journal*, 4(2), 1253–1260.
- Sumintono, B. . W. W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata.
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Model Rasch Untuk Penelitian*
-

*Ilmu-Ilmu Sosial.*

- Wang, Z., Chi, S., Luo, M., Yang, Y., & Huang, M. (2017). Development of an instrument to evaluate high school students' chemical symbol representation abilities. *Chemistry Education Research and Practice*, 18(4), 875–892. <https://doi.org/10.1039/c7rp00079k>
- Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, 89(3), 335–345. <https://doi.org/10.1021/ed100852t>
- Zahro', S. F., & Ismono, I. (2021). Analisis Kemampuan Multirepresentasi Siswa Pada Materi Keseimbangan Kimia di Masa Pandemi Covid-19. *Chemistry Education Practice*, 4(1), 30. <https://doi.org/10.29303/cep.v4i1.2338>

How to cite this article:

Sulvianti, A., & Aini, F. Q. (2023). Development of Test Instrument to Assess Students' Understanding on Macroscopic, Sub-Microscopic, and Symbolic Levels of Acid-Base Topics Using Rasch Model. *Journal of Educational Sciences*, 7(2), 282-295.

---