JOURNAL
OF BALTIC
SCIENCE
EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

RASCH MODELLING TO EVALUATE REASONING DIFFICULTIES, CHANGES OF RESPONSES, AND ITEM MISCONCEPTION PATTERN OF HYDROLYSIS

Lukman A. R. Laliyo,
Akram La Kilo,
Mardjan Paputungan,
Wiwin Rewini Kunusa,
Lilan Dama,
Citra Panigoro

Abstract. *This study evaluates the difficulties in concept reasoning, changes in response patterns, and item misconception hydrolysis patterns using Rasch modeling. Data were collected through the development of 30 distractor-based diagnostic test items, measuring ten levels of conceptual reasoning ability in three types of salt hydrolysis compounds:* $Na_5P_3O_{10}$, $NaOCl$ *and* $(NH_4)_2SO_4$. *These 30 written test items were completed by 849 students in Gorontalo, Indonesia. The findings show empirical evidence of the reliability and validity of the measurement. Further analysis found that the students' reasoning difficulty levels of the concept of saline solutions were varied; the calculation of saline solution's pH level is the most difficult construct to reason. In particular items, changes in response patterns were found; the misconception curve showed a declining trend and disappeared along with the increase of comprehension along the spectrum of students' abilities. The item misconceptions pattern was found repeatedly in similar items. This finding strengthens the conclusion that resistant misconceptions potentially tend to cause students' conceptual reasoning difficulties and are difficult to diagnose in conventional ways. This study contributes to developing ways of diagnosing resistant misconceptions and being a reference for teachers and researchers in evaluating students' chemical conceptual reasoning difficulties based on Rasch modeling.*

Keywords: *reasoning difficulties, hydrolysis, misconception, Rasch model.*

Lukman A. R. Laliyo, Akram La Kilo, Mardjan Paputungan, Wiwin Rewini Kunusa, Lilan Dama, Citra Panigoro
Gorontalo State University, Indonesia

Introduction

Chemistry learning is not only intended to transfer knowledge and skills but also to build higher-order thinking skills (analytical, creative, critical, synthetic, and innovative) in students. Developing this ability requires correct conceptual mastery of chemistry so that students can use their knowledge to solve problems. Unfortunately, students often experience obstacles in developing these abilities, which tend to be caused by the learning difficulties they experience. Many factors can cause the cause of this difficulty; one of which potentially hinders the conceptual development of students is the difficulty of conceptual reasoning and misconceptions.

Difficulties in concept reasoning are often indicated as one of learning barriers that students find in solving problems due to their lack in utilizing conceptual understanding in an accurate and scientific fashion (Gabel, 1999; Gette et al., 2018). Experts argue that all students – in all educational level – oftentimes do not understand; or only few who understand; or find difficulties in elaborating the linkages between concepts (Johnstone, 1991; Taber, 2019), as well as difficulties in explaining social-scientific problems with the knowledge in chemistry that they have learned in school (Bruder & Prescott, 2013; Kinslow et al., 2018; Owens et al., 2019). These types of difficulties commonly take place due to the students' conceptual understanding that they form according to their own thought process (Ausubel et al., 1978; Yildirir & Demirkol, 2018). This refers to the understanding that is formed based on the sensory impressions, cultural environment, peers, learning media, and learning process in class (Chandrasegaran et al., 2008; Lu & Bi, 2016), that contains misconception (Johnstone, 2006, 2010; Taber, 2002, 2009), and is divergent from scientific concepts (Alamina & Etokeren, 2018; Bradley & Mosimege, 1998; Damanhuri et al., 2016; Orwat et al., 2017; Yaşar et al., 2014).

https://doi.org/10.33225/jbse/22.21.817

Misconceptions that are resistant (Hoe & Subramaniam, 2016) tend to hinder the correct process of conceptual reasoning (Soeharto & Csapó, 2021), as students will find difficulties in receiving and/or even rejecting new insights when they are inconsistent and contrary to their own understanding (Allen, 2014; Damanhuri et al., 2016; Jonassen, 2010; Soeharto et al., 2019). These types of misconceptions come in various forms (Aktan, 2013; Orwat et al., 2017). Therefore, it is crucial to understand how these misconceptions occur in the process of concept reasoning in order to formulate proper strategies to develop students' understanding that is accurate and scientific (Chandrasegaran et al., 2008; Kolomuç & Çalik, 2012; Sunyono et al., 2016).

Salt hydrolysis is one of the concepts in chemistry that students often find it difficult to understand (Damanhuri et al., 2016; Orwat et al., 2017; Tümay, 2016). This issue has been explored by numerous research, and they commonly agree that misconception is one of the contributing factors. Misconceptions in salt hydrolysis are often caused by the difficulties in reasoning the submicroscopic dynamic interaction of buffer solution due to the students' lack of competence in explaining the acid-base concept and chemical equilibrium (Demircioğlu et al., 2005; Orgill & Sutherland, 2008; Orwat et al., 2017); error in interpreting the concept of acid-base strength (Tümay, 2016); difficulty in understanding the definition of salt hydrolysis and characteristics of salt (Sesen & Tarhan, 2011); and difficulty in reasoning the concept of formulation and capacity of buffer solution (Maratusholihah et al., 2017; Sesen & Tarhan, 2011; Tarhan & Acar-Sesen, 2013). The various studies above can conclude the types of concepts that are misunderstood by students, however, generally there are no studies that are able to explain the relationship between these misconceptions and how these misconception patterns are understood at the item level and individual students. This information is crucial for teachers in making subsequent instructional decisions.

Studies on misconceptions commonly use raw scores as the reference. However, raw scores do not refer to final version of data. Therefore, they lack in-depth information to be used as reference in formulating conclusions (He et al., 2016; Sumintono & Widhiarso, 2015). Hence, research studies that use raw scores as reference to obtain conclusion are rather limited in presenting relevant information regarding reasoning difficulties and misconception characteristics of items and students. Psychometrically, this approach tends to have limitations in measuring accurately (Pentecost and Barbera, 2013), due to the difference of scales in the measurement characteristics (Linn & Slinde, 1977). To solve the limitation of conventional psychometric analysis method (Linacre, 2020; Perera et al., 2018; Sumintono, 2018), an approach of Rasch model analysis was applied. This analysis adopts an individual-centered statistical approach that employs probabilistic measurement that goes beyond raw score measurement (Boone & Staver, 2020; Liu, 2012; Wei et al., 2012).

Research studies on misconceptions in chemistry that use Rasch modelling were focusing on diagnosing the changes in students' understanding and learning progress (Hadenfeldt et al., 2013), measuring the content knowledge by pedagogical content knowledge (Davidowitz and Potgieter, 2016), measuring conceptual changes in hydrolysis (Laliyo et al., 2022), measuring scientific investigation competence (Arnold et al., 2018), investigating the item difficulty (Barbera, 2013) and (Park & Liu, 2019), and identifying misconceptions in electrolytes and non-electrolytes (Lu and Bi, 2016). In particular, research studies on misconceptions in chemistry by (Herrmann-Abell & DeBoer, 2016; Herrmann-Abell & DeBoer, 2011) were able to diagnose the misconception structures and detect problems on the items. Grounding from this, a study by Laliyo et al. (2020) was able to diagnose resistant misconceptions in concept of matter state change. In spite of this, research studies on misconceptions that evaluate reasoning difficulties and misconceptions are still relatively limited.

Concept reasoning difficulties and misconceptions often attach to a particular context, and thus are inseparable from the said context in which the content is understood (Davidowitz & Potgieter, 2016; Park & Liu, 2019). Students might be capable of developing an understanding that is different to the context if it involves a ground and scientific concept. However, misconceptions tend to be more sensitive and attached to the context (Nehm & Ha, 2011). The term 'context' in this study refers to a scientific content or topic (Cobb & Bowers, 1999; Grossman & Stodolsky, 1995; Park & Liu, 2019). The incorporation of context in research on misconceptions that apply Rasch model analysis opens up a challenging research area to be explored. This study intended to fill the literature gap by emphasizing the strength and the weakness of Rasch model in evaluating conceptual reasoning and estimating resistant item misconception patterns.

The reasoning difficulties of the concept of salt hydrolysis: $Na_5P_3O_{10}$, $NaOCl$, and $(NH_4)_2SO_4$ are analyzed by distractor-type multiple choices test. Each item contains one correct answer choice and three answer choices designed on a distractor basis. The answer choices of this distractor are answer choices that are generally understood by students but contain misconceptions. The design of this misconception test instrument is adapted from research

https://doi.org/10.33225/jbse/22.21.817

reported by Tümay (2016) regarding misconceptions in acid-base reaction, Seçken (2010) on misconceptions in salt hydrolysis, Damanhuri et al. (2016) regarding acid-base strength, and Orwat et al. (2017) on misconceptions in dissolving process and reaction of ionic compounds with water and chemical equations. According to Sadler (1999) and Herrmann-Abell and DeBoer (2011), distractor answer choices can minimize students giving answers by guessing; therefore, it increases the diagnostic power of the item. The distractor answer choice allows students to choose an answer according to their logical understanding of what they understand.

The problems on these items are detected by option probability curve, in which the item difficulty level is determined based on the size of item logit (Boone & Staver, 2020; Laliyo et al., 2022; Linacre, 2020). By dichotomous score, the curve that is appropriate with the probability of correct answer choice usually increases monotonously along with the increase in students' understanding; while the curve for the distractor sequence tends to decline monotonously as the students' understanding increases (Haladyna, 2004; Haladyna & Rodriguez, 2013; Herrmann-Abell & DeBoer, 2016). Items influenced by distractors will usually generate a curve that deviates from the monotonous behavior of traditional items (Herrmann-Abell & DeBoer, 2016; Herrmann-Abell & DeBoer, 2011; Sadler, 1998; Wind & Gale, 2015).

Problem Statement

Considering the previous explanation, this study was intended to answer the following questions. First, how is the validity and reliability of the measurement instrument employed in this study? This question is intended to explain the effectiveness of the measurement instrument and how valid the resulting data is, including explaining whether the measurement data is in accordance with the Rasch model. The test parameters used are the validity of the test constructs, summary of fit statistics, item fit analysis, and Wright maps.

Second, how does the item reasoning difficulties of salt hydrolysis of $Na_5P_3O_{10}Na_5P_3O_{10}$ and $NaOCl$ $NaOCl$ differ from each other? This question is to explain how the reasoning difficulties of students in different classes differ. Are there items that are responded to differently by the class of students seen from the same construct level? In addition, from the point of view of differences in item difficulty, it can be identified in strata, which construct the level of conceptual reasoning, which tends to be the most difficult for students to reason.

Third, based on changes in the misconception answer choice curve on an item, can it be diagnosed that the response pattern of students' items shows resistant misconceptions? This question is to detect a hierarchy of misconception answer choice curves on an item, which decreases as understanding increases along the spectrum of students' abilities. This hierarchy indicates that there is a dominant problem or difficulty experienced by students on the item in question; this can be proven by the response pattern of misconceptions on certain items, which are repeated on other similar items at the same construct level. If three similar items are found showing the same pattern of response choices for misconceptions, then this shows that there is a tendency for students' misconceptions to be resistant in the construct in question.

Research Methodology

Research Design

The study employed a non-experimental descriptive-quantitative approach, in which the measured variable was students' reasoning ability of concept of hydrolysis. The measured variable involved ten levels of constructs, where each construct has three typical items from different contexts of reasoning tasks. The measurement result was in the form of numbers, while each right answer on an item was given a score. The numbers represent the abstract concepts that are measured empirically (Chan et al., 2021; Neuman, 2014). No interventions in any way were made in the learning process and learning materials. In other words, no treatments were applied to students to ensure that they can answer all question items in the measurement instruments correctly. The scope of the construct comprised properties of salt-forming compounds, properties of salts in water, properties of salts based on their constituent compounds, types of salt hydrolysis reactions, calculation of pH, types of compounds forming buffer solutions, and properties of buffer solutions based on their constituent compounds. The research was conducted for six months, from January to June 2022. The research permit for this study was obtained from the government, the school administration staff, and the university board of leaders.

https://doi.org/10.33225/jbse/22.21.817

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Respondents

A total of 849 respondents were involved in this study. The respondents were 537 upper-secondary school students (A), 165 university students majoring Chemistry Education (B), and 147 Chemistry students (C). The reason for selecting respondents in strata was to estimate that the difficulty of reasoning on certain items may be experienced by respondents at all grade levels. The A group (16-17 age range) was selected from six leading schools in Gorontalo by random sampling technique. This technique allows the researchers to obtain the most representative sample from the entire population in focus. In Gorontalo, there were 62 public upper-secondary schools spread over six districts/cities. Each area was randomly assigned to one school, and the sample was randomly selected from every eleventh grade in those schools (Neuman, 2014). Meanwhile, students B and C (aged 19-21 years) were randomly selected from a population of 1200 students from the Faculty of Mathematics and Natural sciences, from one of the universities in Gorontalo, Indonesia. Prior to conducting this study, the respondents in A group were confirmed to have learned formally about acid-base, properties of hydrolyzed salts, hydrolysis reactions, pH calculations, and buffer solution reactions. For the B and C group, these concepts were re-learned in the Basic Chemistry and Physical Chemistry courses. With regard to research principles and ethics as stipulated by the Institutional Review Board (IRB), students who are voluntarily involved in this research were asked for their consent, and they were notified that their identities are kept confidential, and the information obtained is only intended for scientific development (Taber, 2014).

Development of Instruments

The research instrument involved 30 items that measure the students' reasoning ability on the concept of hydrolysis. The instrument was in the form of multiple-choice test that was adapted from the previous study (Laliyo et al., 2022; Suteno et al., 2021), and developed by referring to the recommendations from Wilson (2005). Table 1 shows the conceptual map of reasoning of salt hydrolysis that involves ten levels of constructs. A difference in level of reasoning construct represents the qualitative improvement of the measured construct (Wilson, 2009, 2012). These construct levels refer to the Curriculum Standard of Chemistry Subject in the Eleventh Grade in Indonesia, as per the Regulation of Ministry of Education and Culture of Republic of Indonesia No. 37/2018. Each level of construct has three typical items, for example, 1/Item1A, 6/Item1B, and 11/Item1C. These items measure the level 1 construct, i.e., determining the characteristics of forming compounds of $Na_5P_3O_{10}$, $NaOCl$, and $(NH_4)_2SO_4$. These three items are different from each other from the context of reasoning task of hydrolysis solution.

Table 1
Conceptual Map of Reasoning of Salt Hydrolysis

Concept Reasoning Level	Serial Number/Item/Context Reasoning Task		
	A	B	C
Level 1. Determining the properties of forming compounds of salt	1/Item1A	6/Item1B	11/Item1C
Level 2. Explaining the properties of compounds that are completely and partially ionized in salt solutions	16/Item2A	21/Item2B	26/Item2C
Level 3. Determining the properties of salt in water	2/Item3A	I7tem3B	12/Item3C
Level 4. Explaining the properties of salt based on the forming compounds	17/Item4A	22/Item4B	27/Item4C
Level 5. Determining types of hydrolysis reaction of salt	3/Item5A	8/Item5B	13/Item5C
Level 6. Explaining result of salt hydrolysis reaction	18/Item6A	23/Item6B	28/Item6C
Level 7. Calculating pH level of salt solution	4/Item7A	9/Item7B	14/Item7C
Level 8. Explaining pH calculation result of salt solution	19/Item8A	24/Item8B	29/Item8C
Level 9. Determine types of forming compounds of buffer solution	5/Item9A	10/Item9B	15/Item9C
Level 10. Explaining the properties of buffer solution based on the forming compounds	20/Item10A	25/Item10B	30Item10C

Description: A = $Na_5P_3O_{10}$ salt solution, B = $NaOCl$ salt solution, C = $(NH_4)_2SO_4$ salt solution

https://doi.org/10.33225/jbse/22.21.817

Each item was designed with four answer choices, with one correct answer and three distractor answers. The distractor functions to prevent students from guessing the correct answer choice, as is often the case with traditional items, by providing answer choices that are considered reasonable, particularly for students who hold firmly to their misconceptions (Herrmann-Abell & DeBoer, 2016; Herrmann-Abell & DeBoer, 2011; Naah & Sanger, 2012; Sadler, 1998). A score of 1 is given for the correct answer, while 0 is given for the incorrect answers. The probability of guessing each correct answer choice is relatively small, only 0.20 (Lu and Bi, 2016). Students will only choose an answer that is according to their comprehension. If the distractor answer choices on each item work well, the correct answer choices on each item should not be easy to guess (Herrmann-Abell & DeBoer, 2016; Herrmann-Abell & DeBoer, 2011).

The congruence of the correlation between constructs and items, or the suitability of answer choices with the level of the item's reasoning construct, or congruence of content with the constructs measured by (Wilson, 2005, 2008) were confirmed through the validation of three independent experts, i.e., one professor in chemistry education and two doctors in chemistry. The three expert validators agreed to determine Fleiss measure, $K = .97$, $p < .0001$, or that the item validity arrived at 'good' category (Landis & Koch, 1977).

Data Collection

The data collection was conducted face-to-face, at school supervised by classroom teachers and on campus supervised by researchers. Each respondent was asked to give written response through the answer sheet provided. All students were asked to work on all items according to the allotted time (45 minutes). Instrument manuscripts were collected right after the respondents finished giving responses, and the number of instruments was confirmed to be equal to the number of participating students. The data obtained in the previous process were still in the form of ordinal data. The data were then converted into interval data that have the same logit scale using the WINSTEPS software version 4.5.5 (Bond & Fox, 2015; Linacre, 2020). The result is a data calibration of the students' ability and the level of difficulty of items in the same interval size.

Conducting Rasch Analysis

The Rasch model analysis is able to estimate students' abilities and stages of development in each item (Masters, 1982). This allows the researchers to combine different responses opportunities for different items (Bond & Fox, 2007). It combines algorithm of probabilistic expectation result of item 'i' and student 'n' as: $\mathrm{Pni}\ (Xni = 1/(\beta n, \delta i\ ) = \ (\ e\verb|^|((\beta n - \delta\_i\ )))/(1 + (\beta n - \delta\_i\ )))$. The statement $\mathrm{P\_ni}\ (Xni = 1/(\beta n, \delta i\ ))$ is the probability of student n in the item i to generate a correct answer (x = 1); with the students' ability, ßn, and item difficulty level of δi (Bond & Fox, 2015; Boone & Staver, 2020). If the algorithm function is applied into the previous equation, it will be $\log\ (\mathrm{Pni}\ (Xni = 1/(\beta n, \delta i\ )) = \ \beta n - \ \delta i,$; thus, the probability for a correct answer equals to the students' ability minus item difficulty level (Sumintono & Widhiarso, 2015).

The measures of students' ability (person) βn and the item difficulty level δi are stated on a similar interval and are independent to each other, which are measured in an algorithm unit called odds or log that can vary from -00 to +00 (Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015). The use of logit scale in Rasch model is the standard interval scale that shows the size of person and item. Boone et al. (2014) argue that ordinal data cannot be assumed as linear data, therefore cannot be treated as a measurement scale for parametric statistic. The ordinal data are still raw and do not represent the measurement result data (Sumintono, 2018). The size of data (logit) in Rasch model is linear, thus, can be used for parametric statistical test with better congruence level compared to the assumption of statistical test that refers to raw score (Park & Liu, 2019).

Research Results

Validity and Reliability of the Instruments

The first step is to ensure the validity of test constructs by measuring the fit validity (Banghaei, 2008; Chan et al., 2021). This serves to determine the extent to which the item fits to the model, and because it is in accordance with the concept of singular attribute (Boone et al., 2014; Boone & Noltemeyer, 2017; Boone & Staver, 2020). The mean square residual (MNSQ) shows the extent of impact of any misfit with two forms of Outfit MNSQ and Infit

https://doi.org/10.33225/jbse/22.21.817

MNSQ. Outfit is the chi-square that is sensitive to the outlier. Items with outliers are often guess answers that happen to be correct chosen by low-ability students, and/or wrong answers due to carelessness for high-ability students. The mean box of Infit is influenced by the response pattern with focus on the responses that approach the item difficulty or the students' ability. The expected value of MNSQ is 1.0, while the value of PTMEA Corr. is the correlation between item scores and person measures. This value is positive and does not approach zero (Bond & Fox, 2015; Boone & Staver, 2020; Lu & Bi, 2016).

Table 2 indicates that the average Outfit MNSQ of test item is 1.0 logit; this is in accordance with the ideal score range between 0.5-1.5 (Boone et al., 2014). This means that the item is categorized as productive for measurement and has a logical prediction. The reliability value of the Cronbach's Alpha (KR-20) raw score test is 0.81 logit, indicating the interaction between 849 students and the 30 KPIH test items is categorized as good. In other words, the instrument has excellent psychometric internal consistency and is considered a reliable instrument (Adams & Wieman, 2011; Boone & Staver, 2020; Sumintono & Widhiarso, 2015). The results of the unidimensionality measurement using Principal Component Analysis (PCA) of the residuals show that the raw data variance at 23.5%, meeting the minimum requirements of 20% (Boone & Staver, 2020; Sumintono & Widhiarso, 2014). This means that the instrument can measure the ability of students in reasoning hydrolysis items very well (Chan et al., 2021; Fisher, 2007; Linacre, 2020).

Table 2
Summary of Fit Statistics

	Student (*N=849*)	Item (*N=30*)
Measures (logit)		
$\bar{x}$	-.20	.00
SE (standard error)	.03	.14
SD (standard deviation)	0.99	0.75
Outfit mean square		
$\bar{x}$	1.00	1.00
SD	0.01	0.02
Separation	1.97	9.15
Reliability	.80	.99
Cronbach's Alpha (KR-20)	.81	

The results of testing the quality of the item response pattern as well as the interaction between person and item show a high score of the separation item index (9.15 logit) and high item reliability index (.99 logit); this is the evidence of the level of students' reasoning abilities and supports the construct validity of the instrument (Boone & Staver, 2020; Linacre, 2020). The higher the index (separation and reliability) of the items, the stronger the researcher's belief about replication of the placement of items in other students that are appropriate (Boone et al., 2014; Boone & Staver, 2020; Linacre, 2020). The results of the measurement of the person separation index (1.97 logit) and the person reliability index (.80 logit) indicate that there is a fairly good instrument sensitivity in distinguishing the level of reasoning abilities of high-ability and low-ability students. The average logit of students is -.20 logit, indicating that all students are considered to have the abilities below the average test item (.00 logit). The deviation standard is at .99 logit, displaying a fairly wide dispersion rate of item reasoning ability of hydrolysis in students (Boone et al., 2014; Boone & Staver, 2020; Linacre, 2020).

The second step is to ensure the item quality by statistic fit test (Boone & Staver, 2020; Linacre, 2020). An item is considered as misfit if the measurement result of the item does not meet the three criteria of: *Outfit mean square residual (MNSQ):* $.5 < y < 1.5$; *Outfit standardized mean square residual (ZSTD):* $-2 < Z < +2$; and *point measure correlation (PTMEA CORR):* $.4 < x < .8$. Outfit ZSTD value serves to determine that the item has reasonable predictability. Meanwhile, the Pt-Measure Corr value is intended to check whether all items function as expected. If a positive value is obtained, the item is considered acceptable; however, if a negative value is obtained, then the item is considered not functioning properly, or contains misconceptions (Bond & Fox, 2015; Boone et al., 2014; Sumintono & Widhiarso, 2015). Table 3 indicates that all items are in the Outfit MNSQ range, while 18 items are not in the Outfit

ZSTD range, and 13 items are not in the Pt-Measure Corr range, and there is no negative value for the Pt-Measure Corr criteria. There is no single item that does not meet all three criteria, so all items are retained. If only one or two criteria are not met, the item can still be used for measurement purposes.

Table 3
Item Fit Analysis

Item	Measure	Infit MNSQ	Outfit MNSQ	Outfit ZSTD	Point Measure Correlation
Item1A	-1.21	,91	.82	-2.96*	.44
Item1B	-.55	.94	.95	-1.13	.44
Item1C	-1.13	.95	.91	-1.53	.40
Item2A	-.69	1.05	1.07	1.91	.32*
Item2B	.00	1.09	1.16	3.84*	.31*
Item2C	-.19	1.12	1.17	3.92*	.28*
Item3A	-.26	.89	.90	-2.41*	.49
Item3B	-.41	.87	.83	-4.31*	.52
Item3C	-.89	.95	.86	-2.71*	.43
Item4A	-.60	1.00	1.07	1.57	.36*
Item4B	-.59	.87	.84	-3.72*	.50
Item4C	-.80	.95	.89	-2.11*	.42
Item5A	-1.14	.98	.91	-1.45	.37*
Item5B	-.24	.96	.94	-1.55	.43
Item5C	-.87	.97	.89	-2.20*	.41
Item6A	.37	.99	1.03	.57	.41
Item6B	.42	.96	.97	-.65	.44
Item6C	.22	.93	.91	-2.20*	.48
Item7A	.50	.85	.83	-3.70*	.55
Item7B	.45	.83	.82	-3.98*	.56
Item7C	-.06	1.02	1.03	.64	.39*
Item8A	1.16	.89	.90	-1.35	.49
Item8B	1.58	1.11	1.22	2.20*	.27*
Item8C	.16	1.11	1.12	2.70*	.31*
Item9A	.49	1.16	1.40	7.40*	.25*
Item9B	.70	1.05	1.07	1.27	.36*
Item9C	.82	.99	1.06	1.06	.40
Item10A	.93	1.21	1.28	4.11*	.21*
Item10B	.84	1.18	1.27	4.13*	.23*
Item10C	.97	1.19	1.36	4.97*	.21*

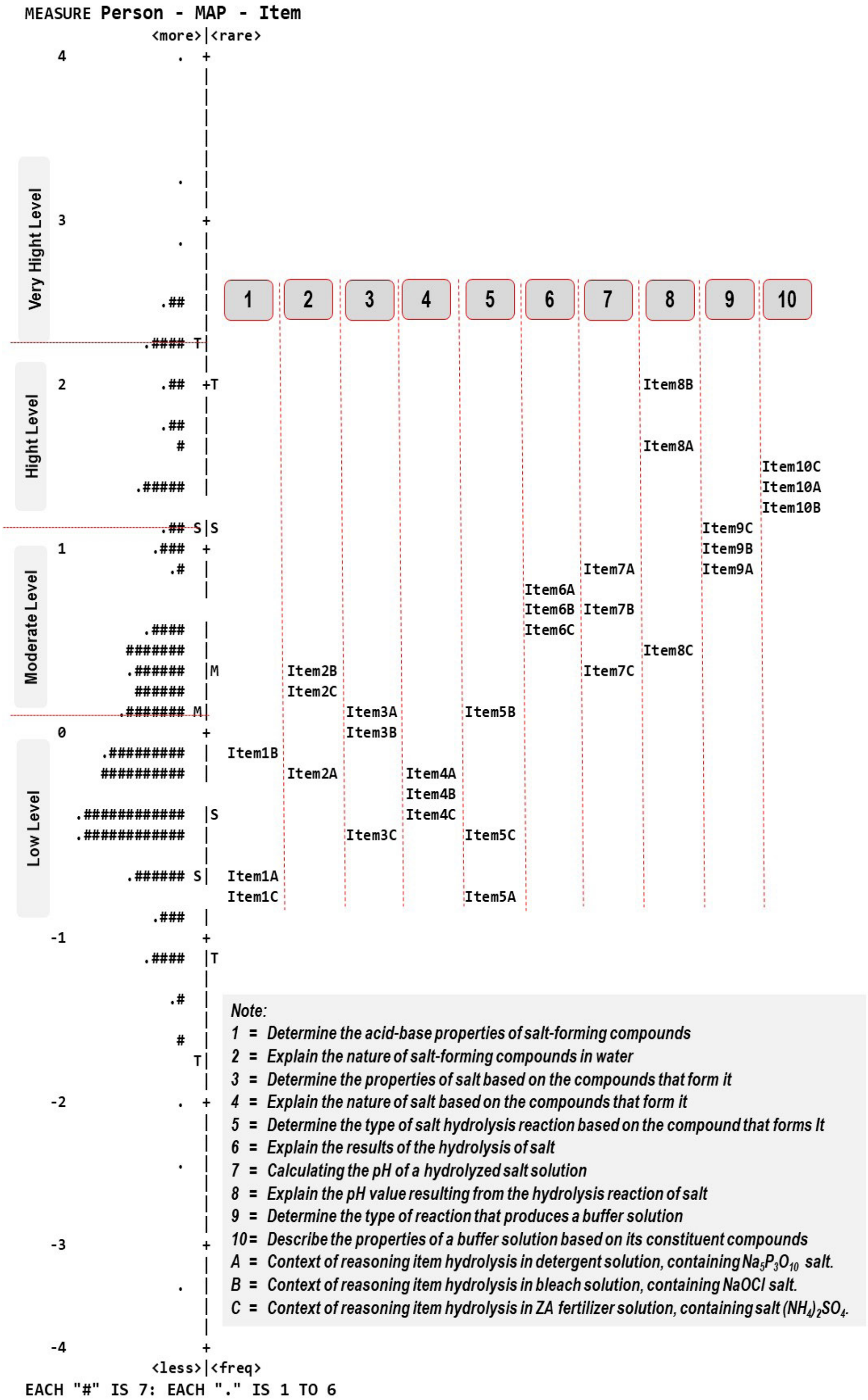
Description: (*) is the items not in the range of Outfit MNSQ and Point Measure Correlation

The third step is to measure the consistency between item difficulty level and students' ability level. Figure 1 below is a Wright map that shows the graphic representation of an increase in the students' ability and the item's difficulty levels within the same logit scale (Bond & Fox, 2015). The higher the logit scale, the higher the student's ability level and the item's difficulty level. On the other hand, the lower the logit scale, the lower the student's ability level and the item's difficulty level (Boone et al., 2014). Most of the items are at above average (.00 logit). Item8B (1.58 logit) is the most difficult item, while Item1A (-1.21 logit) is the easiest item. However, at the lower (<-1.21

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

logit) and higher (>1.58 logit) students' ability levels, there were no items equivalent to the intended ability level. Meanwhile, the distribution of students' abilities is in accordance with the logit size. The students with the highest ability reached 3.62 logit, while the students with the lowest ability obtained -3.61 logit.

Figure 1
Wright Map: Person-Map-Item



Note:
1 = Determine the acid-base properties of salt-forming compounds
2 = Explain the nature of salt-forming compounds in water
3 = Determine the properties of salt based on the compounds that form it
4 = Explain the nature of salt based on the compounds that form it
5 = Determine the type of salt hydrolysis reaction based on the compound that forms It
6 = Explain the results of the hydrolysis of salt
7 = Calculating the pH of a hydrolyzed salt solution
8 = Explain the pH value resulting from the hydrolysis reaction of salt
9 = Determine the type of reaction that produces a buffer solution
10= Describe the properties of a buffer solution based on its constituent compounds
A = Context of reasoning item hydrolysis in detergent solution, containing $Na_5P_3O_{10}$ salt.
B = Context of reasoning item hydrolysis in bleach solution, containing NaOCl salt.
C = Context of reasoning item hydrolysis in ZA fertilizer solution, containing salt $(NH_4)_2SO_4$.

Difference in Item Reasoning Difficulty of Salt Hydrolysis: $\mathbf{Na_5P_3O_{10}}$, $\mathbf{NaOCl}$, *and* $\mathbf{(NH_4)_2SO_4}$

Based on the size of logit value item (LVI), by dividing the distribution of measure of all logit items based on the average of item and deviation standard, the item reasoning difficulty level of salt hydrolysis of $Na_5P_3O_{10}$, $NaOCl$, and $(NH_4)_2SO_4$ is categorized into four categories: easiest items to reason (LVI ≤ -.75 logit), easy items to reason (-.75 ≥ LVI ≥ .00 logit), difficult items to reason (.00 ≥ LVI ≥ .75 logit), and most difficult items to reason (LVI > .75 logit). It is displayed in Table 4. From this table, two interesting points were discovered. First, there are no similar items with the same difficulty level. For example, Item2A (-.69) and Item2C (-.19) are easier for students to reason than Item2B (.00). Second, the sequence of item difficulty in saline solutions of $Na_5P_3O_{10}$, $NaOCl$, and $(NH_4)_2SO_4$ is different and does not match the conceptual map (Table 1). For example, Item 5A(-1.14), was found to be easier to reason than Item2A(-.69), Item4A(-.60) and Item3A (-.26). In contrast, Item8B (1.58) was the most difficult to reason than Item10B(.84), Item9B(.70). This finding explains that at the same construct level, the level of reasoning difficulty of three similar items turns out to be different.

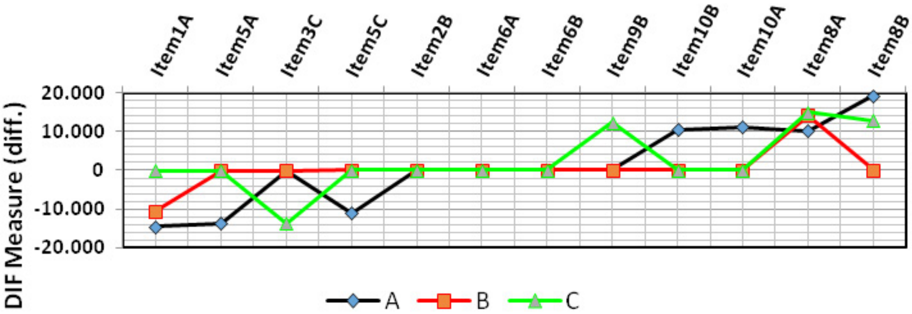
Table 4
Logit Value Item (LVI) Analysis (N=30)

Difficulty Level	Item Code (logit)		
	A	B	C
Very Difficult: (LVI > .75 logit).	Item8A(1.16) Item10A(.93)	Item8B(1.58) Item10B(.84)	Item10C(.97) Item9C(.82)
Difficult: (.00 ≥ LVI ≥ .75 logit)	Item7A(.50) Item9A(.49) Item6A(.37)	Item9B(.70) Item7B(.45) Item6B(.42) Item2B(.00)	Item6C(.22) Item8C(.12)
Easy: (-.75 ≥ LVI ≥ .00 logit)	Item3A(-.26) Item4A(-.60) Item2A(-69)	Item5B(-.24) Item3B(-.41) Item1B(-.55) Item4B(-.59)	Item7C(-.06) Item2C(-.19)
Very Easy: (LVI ≤ -.75 logit).	Item5A(-1.14) Item1A(-1.21)	--	Item4C(-.80) Item5C(-.87) Item3C(-.89) Item1C(-1.13)

Description: A = $Na_5P_3O_{10}$, saline solution, B = $NaOCl$ salt solution, C = $(NH_4)_2SO_4$ salt solution

The testing of difference of item reasoning difficulty level from the difference of students' grade level applied Differential Item Functioning (DIF) (Adams et al., 2021; Bond & Fox, 2007; Boone, 2016; Rouquette et al., 2019). An item is considered as DIF if the t value is less than -2.0 or more than 2.0, the DIF contrast value is less than 0.5 or more than 0.5, and the probability (*p*) value is less than .05 or more than .05 (Bond & Fox, 2015; Boone et al., 2014; Chan et al., 2021). A total of 12 items were identified to yield significantly different responses (Figure 2). There are five curves that approach the upper limit, i.e., items with high reasoning difficulty level: Item9B (.70), Item10B (.84), Item10A (.93), Item8A (1.16), and Item8B (1.58). Moreover, four curves that approach the lower limit are items with low reasoning difficulty level, i.e.: Item1A (-1.21), Item5A (-1.14), Item3C (-.89), and Item5C (-.87).

Figure 2
Person DIF plot based on Difference of Students' Grade Level



Note: A = Upper-Secondary School students, B = Chemistry Education university students, C = Chemistry university students

Based on Figure 2, an interesting case was identified, where for student A, Item8B was more difficult than Item8A; on the other hand, for students B and C, Item8A was more difficult than Item8B. In other words, the characteristics of item difficulty among A, B, and C groups are different. It is possible that students in group A with low abilities could guess the correct answer to Item8A, while students B and C with high abilities answered Item8A incorrectly because of carelessness. In addition, it was found that the difficulty level was Item8B (1.58) > Item10B (.84) > Item9B (.74). That is, the difficulty level of the items is different; this happens because of differences in student responses.

Analysis of Changes in Item Misconception Curve and Pattern

The option probability curve is applied to detect the response pattern of students' choice of answers on each item. This curve provides a visual image of the distribution of correct answer choices and distractor answer choices (containing misconceptions) across the spectrum of students' knowledge (starting from high school students, chemistry education students, and chemistry students). This allows the researchers to evaluate if the shape of the curve is fit for purpose, or if there is something unusual that indicates a structured problem with an item. The shape of the curve can show a hierarchy of misconceptions that disappears sequentially as students become more knowledgeable about a topic, either through out-of-school experiences or through formal learning. In this article, we present the sample of option probability curve for three items: Item8A, Item8B, and Item8C.

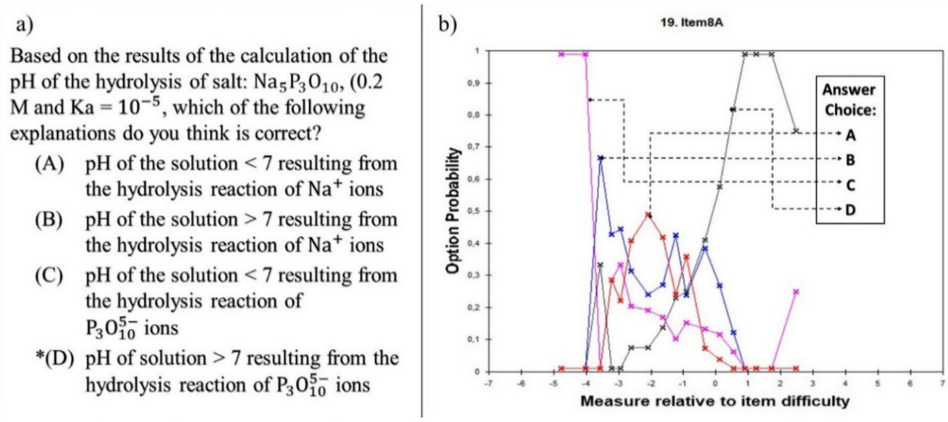
Sample 1

Figure 3 (a) displays Item8A (1.16 logit) that tests the students' reasoning on the pH calculation results of $Na_5P_3O_{10}$. The option probability curve of this item is shown in Figure 3 (b). Students whose reasoning ability is very low (between -5.0 and -1.0 logit on the overall ability scale) are more likely to choose answer A (pH level of the solution < 7 resulting from the hydrolysis reaction of ion Na^+). Students with abilities between -4.0 and +1.0 prefer the answer B (pH level of the solution > 7 resulting from the hydrolysis reaction of ion Na^+), and students with abilities between -5.0 and +3.0 are more likely to choose answer C (pH level of the solution < 7 resulting from the hydrolysis reaction of ion $P_3O_{10}^{5-}$). Meanwhile, students with abilities greater than -3.0 choose the correct answer D (pH level of the solution > 7 resulting from the hydrolysis reaction of ion $P_3O_{10}^{5-}$). The pattern of responses produced by students at this level of ability is understandable. At the lowest level, students do not understand the calculation of pH and ions resulting from the salt hydrolysis reaction (answer choice A). When their understanding of acids and bases develops, they choose the answer B. In this case, students can reason with the calculation of pH, but do not understand the hydrolysis reaction and the principle of reaction equilibrium. Conversely, students who pick the option C find difficulties in reasoning the calculation of pH but are able to correctly state the ions resulting from the hydrolysis reaction of $P_3O_{10}^{5-}$. The misconceptions in answer choice A are significant for low-ability students, but misconceptions in answer choices B and C are actually detected in high-ability students. The visualization of answer choices B and C curves appears with two peaks, reflecting an unusual or strange curve, then decreases and disappears as understanding increases.

https://doi.org/10.33225/jbse/22.21.817

Figure 3.

(a) Sample of Item8A (1.16 logit) Tests the Students' Reasoning on pH Calculation Result of $Na_5P_3O_{10}$,
(b) Option Probability Curve of the Said Item



a)

Based on the results of the calculation of the pH of the hydrolysis of salt: $Na_5P_3O_{10}$, (0.2 M and Ka = 10^{-5}, which of the following explanations do you think is correct?

(A) pH of the solution < 7 resulting from the hydrolysis reaction of Na^+ ions

(B) pH of the solution > 7 resulting from the hydrolysis reaction of Na^+ ions

(C) pH of the solution < 7 resulting from the hydrolysis reaction of $P_3O_{10}^{5-}$ ions

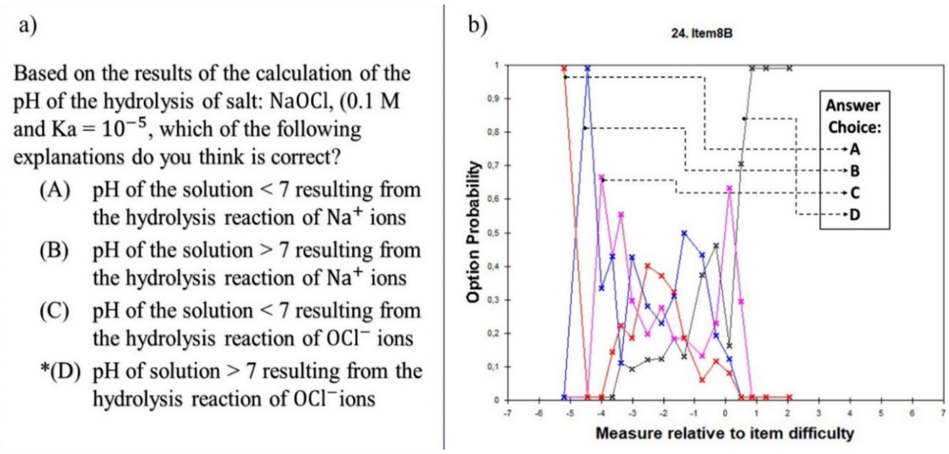
*(D) pH of solution > 7 resulting from the hydrolysis reaction of $P_3O_{10}^{5-}$ ions

Sample 2

Figure 4 (a) displays Item8B (1.58 logit) that tests the students' reasoning on the pH calculation results of $NaOCl$. The option probability curve of this item is shown in Figure 4 (b).

Figure 4

(a) Sample Item8B (1.58 logit) Testing the Students' Reasoning on pH Calculation Result of $NaOCl$
(b) the Option Probability Curve of the Said Item



a)

Based on the results of the calculation of the pH of the hydrolysis of salt: NaOCl, (0.1 M and Ka = 10^{-5}, which of the following explanations do you think is correct?

(A) pH of the solution < 7 resulting from the hydrolysis reaction of Na^+ ions

(B) pH of the solution > 7 resulting from the hydrolysis reaction of Na^+ ions

(C) pH of the solution < 7 resulting from the hydrolysis reaction of OCl^- ions

*(D) pH of solution > 7 resulting from the hydrolysis reaction of OCl^- ions

Students whose reasoning ability is very low (between -5.0 and -5.0 logit on the overall ability scale) are more likely to choose answer A (pH level of the solution < 7 resulting from the hydrolysis reaction of ion Na^+). The answer B (pH level of the solution > 7 resulting from the hydrolysis reaction of ion Na^+) and answer C (pH level of the solution < 7 resulting from the hydrolysis reaction of ion OCl^-) show two curve peaks in the probability of students' ability between -4.0 and +1.0 logit. Meanwhile, the answer D (pH level of the solution > 7 resulting from the hydrolysis reaction of ion OCl^-) increases along the improvement of students' ability, moving from -4.0 up to +3.0 logit. The response pattern expressed in the option probability curve for this item is interesting, because the answer choice curves A, B, and C further justify acid-base misconceptions and hydrolysis reactions, as happened in

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

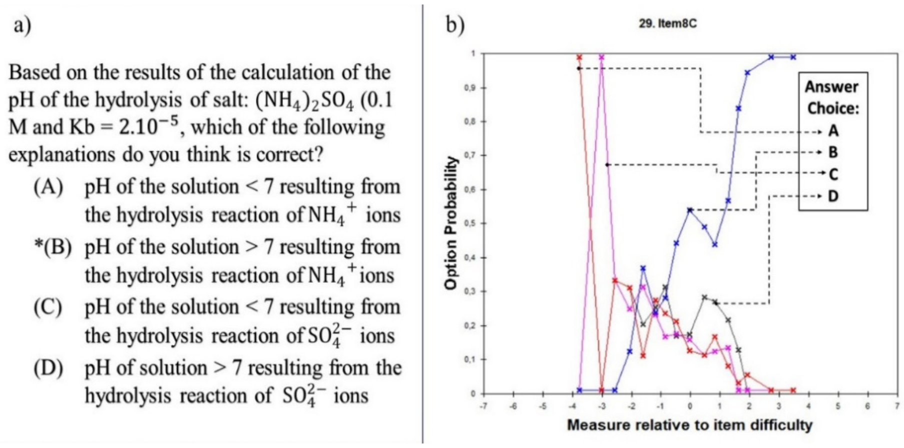
Item 8A. In addition, the visualization of answer choices B and C curves is seen with three peaks, reflecting unusual or odd curves, which decrease as understanding increases.

Sample 3

Figure 5 (a) displays Item8C (.12 logit) that tests the students' reasoning on the pH calculation results of $(NH_4)_2SO_4(NH_4)_2SO_4$. The option probability curve of this item is shown in Figure 3 (b).

Figure 5

(a) Sample of Item8C (.12 logit) Testing the Students' Reasoning on the pH Calculation Results of $(NH_4)_2SO_4$, (b) Option Probability Curve of the Said Item



a)

Based on the results of the calculation of the pH of the hydrolysis of salt: $(NH_4)_2SO_4$ (0.1 M and Kb = 2.10^{-5}, which of the following explanations do you think is correct?

(A) pH of the solution < 7 resulting from the hydrolysis reaction of NH_4^+ ions

*(B) pH of the solution > 7 resulting from the hydrolysis reaction of NH_4^+ ions

(C) pH of the solution < 7 resulting from the hydrolysis reaction of SO_4^{2-} ions

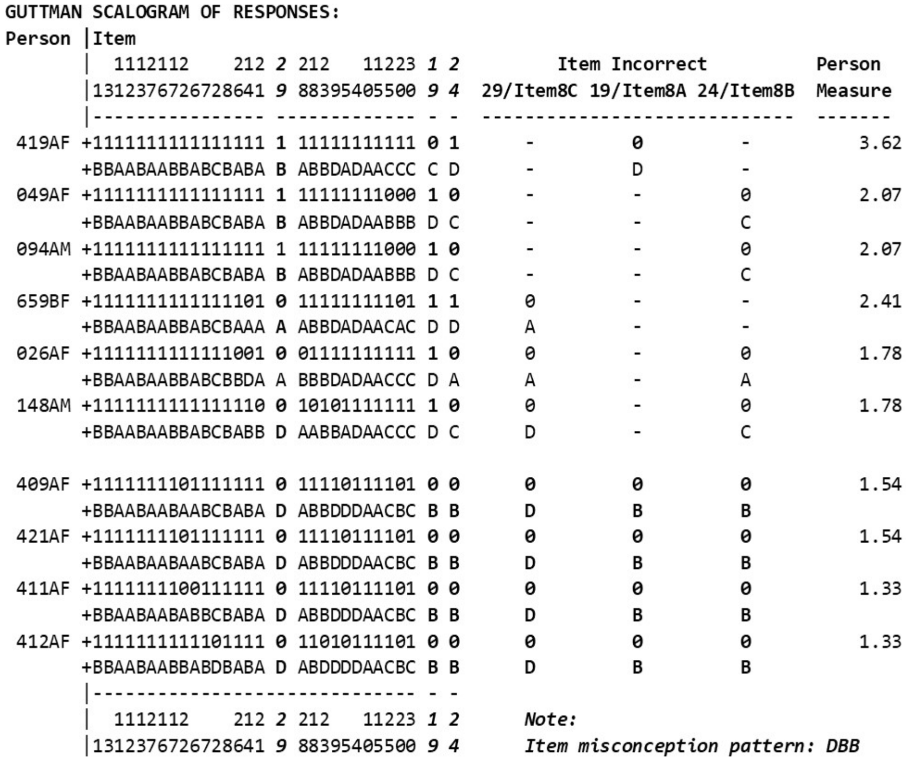
(D) pH of solution > 7 resulting from the hydrolysis reaction of SO_4^{2-} ions

The probability of answer A (pH level of the solution < 7 resulting from the hydrolysis reaction of ion NH_4^+) is the highest for students with lowest reasoning ability (between <-3.0 and 2.0 logit). The visualization of curve A shows three peaks, i.e., in the lowest capability range (<-3.0 logit), then in the capability range between -1.0 logit and 2.0 logit. The visualization of curve of answer C (pH level of the solution < 7 resulting from the hydrolysis reaction of ion SO_4^{2-}) also has three peaks, similar to the curve A; on the other hand, the curve of answer D (pH level of the solution > 7 resulting from the hydrolysis reaction of ion SO_4^{2-}) is at the ability range of high-ability students (<2.0 logit). The correct answer B (pH level of the solution > 7 resulting from the hydrolysis reaction of ion NH_4^+) at the ability range between -4.0 and 5.0 logit increases monotonously along with the decline in curve A, C, and D.

It is interesting to take a closer look at how the curves of the three items change using the Guttman Scalogram (Table 6). This table details several examples of student item response patterns, in two forms, namely the 0 and 1 dichotomy pattern, and the actual response pattern. This response pattern is ordered by the level of difficulty of the item (easiest at left to most difficult at right). The response patterns of 409AF (1.54), 421AF (1.54), 411AF (1.33) and 412AF (1.33), which were highly capable, chose the misconception answer D (for Item8C, fourteenth row from right), answer choice B (for Item8A, second row from right), and answer choice D (for Item8B first row from right). This is an example of a resistant item misconception pattern. Meanwhile, the response pattern of respondent 419AF (3.62) who chose the misconception answer C (for Item8A), 049AF (2.07) and 094AM (2.07) choosing the misconception answer C (for Item8B), and 659BF(2.41) choosing the misconception answer A (for Item8B). Item8C) is a different pattern of misconceptions.

https://doi.org/10.33225/jbse/22.21.817

Table 6
Scalogram Analysis



```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |  1112112    212 2 212   11223 1 2        Item Incorrect          Person
       |1312376726728641 9 88395405500 9 4  29/Item8C 19/Item8A 24/Item8B  Measure
       |---------------- ----------- - -  ---------------------------  -------
 419AF +1111111111111111 1 11111111111 0 1      -         0         -        3.62
       +BBAABAABBABCBABA B ABBDADAACCC C D      -         D         -
 049AF +1111111111111111 1 11111111000 1 0      -         -         0        2.07
       +BBAABAABBABCBABA B ABBDADAABBB D C      -         -         C
 094AM +1111111111111111 1 11111111000 1 0      -         -         0        2.07
       +BBAABAABBABCBABA B ABBDADAABBB D C      -         -         C
 659BF +1111111111111101 0 11111111101 1 1      0         -         -        2.41
       +BBAABAABBABCBAAA A ABBDADAACAC D D      A         -         -
 026AF +1111111111111001 0 01111111111 1 0      0         -         0        1.78
       +BBAABAABBABCBBDA A BBBDADAACCC D A      A         -         A
 148AM +1111111111111110 0 10101111111 1 0      0         -         0        1.78
       +BBAABAABBABCBABB D AABBADAACCC D C      D         -         C

 409AF +1111111101111111 0 11110111101 0 0      0         0         0        1.54
       +BBAABAABAABCBABA D ABBDDDAACBC B B      D         B         B
 421AF +1111111101111111 0 11110111101 0 0      0         0         0        1.54
       +BBAABAABAABCBABA D ABBDDDAACBC B B      D         B         B
 411AF +1111111100111111 0 11110111101 0 0      0         0         0        1.33
       +BBAABAABABBCBABA D ABBDDDAACBC B B      D         B         B
 412AF +1111111111101111 0 11010111101 0 0      0         0         0        1.33
       +BBAABAABBABDBABA D ABDDDDAACBC B B      D         B         B
       |----------------------------- - -
       |  1112112    212 2 212   11223 1 2      Note:
       |1312376726728641 9 88395405500 9 4      Item misconception pattern: DBB
```

Discussion

The results of the study have shown empirical evidence regarding the validity and reliability of the measurement instruments at a very good level. This means that the used instrument is effective to evaluate the difficulty of students' conceptual reasoning. On top of that, it is also highlighted that: (1) the order of item reasoning difficulty level of salt hydrolysis of $Na_5P_3O_{10}$, $NaOCl$, and $(NH_4)_2SO_4$ is different (not matching the construct map), and there are no similar items with the same difficulty level despite being in the same construct level; (2) the difficulty level of similar items is different, it is possible that it occurs due to different student responses, where low-ability students can guess the correct answer, while high-ability students are wrong in answering items due to carelessness; (3) The visualization of changes in the answer choice curves and the pattern of item misconceptions shows the evidence that high-ability students tend to have a response pattern of item misconceptions that tend to be resistant, especially related to the construct of calculating the pH of the salt solution.

The results of the research above show that the difficulty level of the three salt hydrolysis compounds ($\mathrm{Na_5P_3O_{10}}$, NaOCl and $\mathrm{(NH_4)_2SO_4}$) tends to be different. This difference is relatively caused by the poor level of mastery of the content and, therefore, gives different reasoning responses in the context of the three salt hydrolysis compounds in question. This fact reinforces the findings of Davidowitz and Potgieter (2016) and Park and Liu (2019) that reasoning and misconceptions tend to be strongly influenced by students' content mastery. This fact has also been explained by Chu et al. (2009), that students showed the existence of context-dependent alternative conceptions or misconceptions in optics when items used different examples, despite evaluating students' understanding of the same concept. Research by Ozdemir and Clark (2009) supports the conclusion that students' reasoning is fragmented and tends to be inconsistent with items in different contexts. Likewise, diSessa et al. (2004) found that students' scientific explanations do not represent their overall understanding of their understanding of a particular item. However, Weston et al. (2015) proposed the opposite results, that students' responses to the four versions of the questions about photosynthesis are not significantly different. This is possible due to the fact that they do not focus on revealing students' misconceptions but rather focus on examining scientific ideas obtained from student responses.

To explain these problems, it is exemplified in the item misconception patterns of the students, for example: answer B (*pH level of the solution > 7 resulting from the hydrolysis reaction of ion* Na^{+}) for Item8A, answer B (*pH level of the solution > 7 resulting from the hydrolysis reaction of ion* Na^{+}) for Item8B, and answer D (pH level of the solution > 7 resulting from the hydrolysis reaction of ion $SO_4{}^{2-}$) for Item8C. It can be seen that all three show the same pattern of misconceptions, in terms of: (a) the pH value of the solution is > 7, and (b) the ions resulting from the hydrolysis reaction of the salt solution. This finding is interesting to observe further. This is because students do not master the concepts of strong acid and strong base accurately and scientifically; they also tend to find it difficult to reason about the hydrolysis reaction of salt solutions. For example, the hydrolysis reaction: $(NH_4)_2SO_4 \rightarrow 2NH_4{}^{+} + SO_4^{2-}$, where ion $NH_4{}^{+} + H_2O \leftrightarrow NH_4OH + H^{+}$. and excess of ion H^{+} cause pH level of the solution to be < 7 and acidic. In addition, the hydrolysis reaction of salt: $NaOCl \rightarrow OCl^{-} + Na^{+}$, where ion OCl^{-} that reacts with water becomes $OCl^{-} + H_2O \leftrightarrow HOCl + OH^{-}$., excess of ion OH^{-} causes pH level of the solution to be > 7 and the solution becomes basic. This is to say that students tend to lack adequate concept understanding on explaining the contribution of ions H^{+} and OH^{-} towards the pH change of saline solution. This finding supports Tümay's (2016) conclusion, that most of students are unable to conceptualize properties acid-base and strength of acid as the property that results from interaction between many factors. This finding is also supported by Nehm and Ha (2011), that the pattern of student responses is highly predictable regardless of the context, especially when the responses involve core scientific concepts. This means that students are more sensitive to their misconceptions than using correct conceptual reasoning in explaining the context of the item.

The results of this study have shown that although students are indeed able to state the acidity of a salt solution correctly, most of them have misconceptions in writing chemical equations. In addition, students tend to have difficulty explaining the nature of hydrolyzed salts, as a result of their inability to understand the acid-base properties of salt-forming compounds as well as to write down salt hydrolysis reaction equations that meet the principles of chemical equilibrium. Therefore, they experience difficulty calculating the pH of the saline solution. This supports the conclusions of Orwal et al. (2017) and Damanhuri et al. (2016), that students have difficulty in explaining the nature of acid-base, strong base and weak base, despite that more than 80% of them understand that ionized acids in water produce ion H^{+} and that the pH level of neutral solution equals to 7, as well as be able to write down the chemical equation for reaction between acid and base. The previous findings also strengthen the study by Solihah (2015), that students assume that the addition of a small amount of strong acid and strong base to a buffer solution does not affect the shift in equilibrium. However, the correct concept is that the addition of a small amount of strong acid and strong base affects the shift in equilibrium. Experts argue that difficulties in understanding the nature of acid-base tend to be influenced by the cultural background of students, and therefore their understanding becomes different and inconsistent (Chiu, 2007; Kala et al., 2013; Lin & Chiu, 2007).

Conclusions and Implications

Compared to the previous studies, the novelty of this study is that it can demonstrate the evidence and the measurement accuracy of reasoning difficulties as well as changes of item misconception curve and pattern on hydrolysis up to the individual scale of each item and each student. The Rasch model can estimate the character and nature of misconceptions, yielding valuable information for teachers in developing appropriate and measurable instructional strategies. The study shows how to combine the procedures of qualitative item development and quantitative data analysis that allow us to investigate deeper regarding the reasoning difficulties and misconceptions on hydrolysis. The example of using the option probability curve above can explain the prevalence of changes in students' misconception answer choices. The pattern of misconceptions was justified using the Guttman Scalogram map; thus, this study was able identify resistant item misconceptions that are commonly experienced by high-ability learners.

These research items are carefully developed and constantly aligned with key ideas about the concept of hydrolysis chemistry that have been learned by students in upper-secondary school. It is hoped that teachers, researchers, and curriculum material developers will be able to use quantitative items and methods similar to those discussed in this study to compare the effectiveness of various materials and approaches with greater precision and objectivity. While this study does not address questions about individual student performance or growth, it is hoped that the items will be useful in helping teachers diagnose individual learners' thinking so as to target learning more effectively.

This research contributes to the field of chemistry learning assessment by validating the reasoning ability test of the hydrolysis concept using psychometric analysis techniques based on the Rasch model of measurement. The validation of the reasoning ability test in this study is expected to fill the gaps in the literature that tend to be limited in conceptual reasoning in the field of hydrolysis chemistry. This is further expected to be one of the references in developing and integrating the Rasch model measurement in the school curriculum in the world, especially in Indonesia.

This research can also function as a guide for researchers in developing ways to assess students' conceptual reasoning abilities. This will provide valuable information regarding differences in ethnicity, gender, and grade level in assessing students' reasoning abilities. These findings will assist researchers in modifying the reasoning ability test developed in this study, into a new assessment that is more adaptive to the learning progress of students.

Research Limitation and Further Study

This study has not considered the differences in the context of the problem presentation and the characteristics of the item on the item difficulty level parameter. Therefore, it is difficult to distinguish the difficulty of items based on differences in students' understanding abilities or precisely because of differences in the context of the problem presented in each item. In addition, the reach of the student population has not yet reached other parts of the Indonesian territory. Future research is expected to be able to reach a wider population of students in Indonesia, taking into account the demographic aspects of students (such as ethnic, social, and cultural differences), and measuring their influence on the level of mastery of concepts and scientific reasoning in different content scopes.

Declaration of Interest

The authors declare no competing interest.

References

Adams, D., Chuah, K. M., Sumintono, B., & Mohamed, A. (2021). Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: a Rasch analysis. *Asian Education and Development Studies, 11*(2), 324-339. https://doi.org/10.1108/AEDS-05-2020-0100

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289–1312. https://doi.org/10.1080/09500693.2010.512369

Aktan, D. C. (2013). Investigation of students' intermediate conceptual understanding levels: The case of direct current electricity concepts. *European Journal of Physics*, *34*(1), 33–43. https://doi.org/10.1088/0143-0807/34/1/33

Alamina, J. I., & Etokeren, I. S. (2018). Effectiveness of imagination stretch teaching strategy in correcting misconceptions of students about particulate nature of matter. *Journal of Education, Society and Behavioural Science*, *27*(1), 1–11. https://doi.org/10.9734/jesbs/2018/43063

Allen, M. (2014). *Misconceptions in primary science* (2nd ed.). Maidenhead: Open University Press.

Arnold, J. C., Boone, W. J., Kremer, K., & Mayer, J. (2018). Assessment of competencies in scientific inquiry through the application of Rasch measurement techniques. *Education Sciences*, *8*(4), 1-20. https://doi.org/10.3390/educsci8040184

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view.* Holt, Rinehart and Winston.

Banghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurement Transaction*, *22*(1), 1145–1162.

Barbera, J. (2013). A psychometric analysis of the chemical concepts inventory. *Journal of Chemical Education*, *90*(5), 546–553. https://doi.org/10.1021/ed3004353

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Routledge Taylor & Francis Group.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. https://doi.org/10.1088/1751-8113/44/8/085201

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, *15*(4), 1–7. https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), 1-13. https://doi.org/10.1080/2331186X.2017.1416898

Boone, W. J., & Staver, J. R. (2020). Correction to: Advances in Rasch analyses in the human sciences. In Boone, W. J., & Staver, J. R. (Eds.), *Advances in Rasch analyses in the human sciences* (pp. C1-C2). Springer. https://doi.org/10.1007/978-3-030-43420-5_21

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Bradley, J. D., & Mosimege, M. D. (1998). Misconceptions in acids and bases: A comparative study of student teachers with different chemistry backgrounds. *South African Journal of Chemistry*, *51*(3), 137–143.

Bruder, R., & Prescott, A. (2013). Research evidence on the benefits of IBL. *ZDM - International Journal on Mathematics Education*, *45*(6), 811–822. https://doi.org/10.1007/s11858-013-0542-2

Chan, S. W., Looi, C. K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: a Rasch model measurement analysis. *Journal of Computers in Education*, *8*(2), 213–236. https://doi.org/10.1007/s40692-020-00177-2

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2008). An evaluation of a teaching intervention to promote students' ability to use multiple levels of representation when describing and explaining chemical reactions. *Research in Science Education*, *38*(2), 237–248. https://doi.org/10.1007/s11165-007-9046-9

Chiu, M. H. (2007). A national survey of student's conceptions of chemistry in Taiwan. *International Journal of Science Education*, *29*(4), 421–452. https://doi.org/10.1080/09500690601072964

Cobb, P., & Bowers, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher*, *28*(2), 4-15. https://doi.org/10.2307/1177185

Chu, H. E., Treagust, D. F., & Chandrasegaran, A. L. (2009). A stratified study of students' understanding of basic optics concepts in different contexts using two-tier multiple-choice items. *Research in Science and Technological Education*, *27*(3), 253–265. https://doi.org/10.1080/02635140903162553

Damanhuri, M. I. M., Treagust, D. F., Won, M., & Chandrasegaran, A. L. (2016). High school students' understanding of acid-base concepts: An ongoing challenge for teachers. *International Journal of Environmental and Science Education*, *11*(1), 9–27. https://doi.org/10.12973/ijese.2015.284a

Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. *International Journal of Science Education*, *38*(9), 1483–1503. https://doi.org/10.1080/09500693.2016.1196843

Demircio-lu, G., Ayas, A., & Demircio-lu, H. (2005). Conceptual change achieved through a new teaching program on acids and bases. *Chemistry Education Research and Practice*, *6*(1), 36–51. https://doi.org/10.1039/B4RP90003K

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, *28*(6), 843–900. https://doi.org/10.1016/j.cogsci.2004.05.003

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095. www.rasch.org/rmt/rmt211m.htm

Gabel, D. (1999). Improving teaching and learning through chemistry education research: A look to the future. *Journal of Chemical Education*, *76*(4), 548-554. https://doi.org/10.1021/ed076p548

Gette, C. R., Kryjevskaia, M., Stetzer, M. R., & Heron, P. R. L. (2018). Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy. *Physical Review Physics Education Research*, *14*(1), 1-16. https://doi.org/10.1103/PhysRevPhysEducRes.14.010113

Grossman, P. L., & Stodolsky, S. S. (1995). Content as context: The role of school subjects in secondary school teaching. *Educational Researcher*, *24*(8), 5–23. https://doi.org/10.3102/0013189X024008005

Hadenfeldt, J. C., Bernholt, S., Liu, X., Neumann, K., & Parchmann, I. (2013). Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter. *Journal of Chemical Education*, *90*(12), 1602–1608. https://doi.org/10.1021/ed3006192

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Taylor & Francis. https://doi.org/10.4324/9780203825945

Haladyna, T. M., & Rodriguez, M. (2013). *Developing and validating test items*. Routledge. https://doi.org/10.4324/9780203850381

He, P., Liu, X., Zheng, C., & Jia, M. (2016). Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. *Chemistry Education Research and Practice*, *17*(2), 381–393. https://doi.org/10.1039/C6RP00004E

Herrmann-Abell, C. F., & DeBoer, G. E. (2016). Using Rasch modeling and option probability curves to diagnose students' misconceptions. *Paper Presented at the 2016 American Educational Research Association Annual Meeting Washington, DC April 8-12, 2016*, 1–12. https://files.eric.ed.gov/fulltext/ED572821.pdf

Herrmann-Abell, C. F., & DeBoer, G. E. (2011). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chemistry Education Research and Practice*, *12*(2), 184–192. https://doi.org/10.1039/c1rp90023d

Hoe, K. Y., & Subramaniam, R. (2016). On the prevalence of alternative conceptions on acid-base chemistry among secondary students: Insights from cognitive and confidence measures. *Chemistry Education Research and Practice*, *17*(2), 263–282. https://doi.org/10.1039/c5rp00146c

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, *7*, 75–83.

Johnstone, A. H. (2006). Chemical education research in Glasgow in perspective. *Chemical Education Research and Practice*, *7*(2), 49–63. https://doi.org/10.1039/b5rp90021b

Johnstone, A. H. (2010). You Can't Get There from Here. *Journal of Chemical Education*, *87*(1), 22–29. https://doi.org/https://doi.org/10.1021/ed800026d

Jonassen, D. H. (2010). Research Issues in Problem Solving. *The 11th International Conference on Education Research New Education Paradigm for Learning and Instruction*, 1–15.

Kala, N., Yaman, F., & Ayas, A. (2013). The effectiveness of Predict-Observe-Explain technique in probing students' understanding about acid-base chemistry: A case for the concepts of pH, pOH and strength. *International Journal of Science and Mathematics Education*, *11*(1), 555–574.

Kinslow, A. T., Sadler, T. D., & Nguyen, H. T. (2018). Socio-scientific reasoning and environmental literacy in a field-based ecology class. *Environmental Education Research*, *4622*, 1–23. https://doi.org/10.1080/13504622.2018.1442418

Kolomuç, A., & Çalik, M. (2012). A comparison of chemistry teachers' and grade 11 students' alternative conceptions of "rate of reaction." *Journal of Baltic Science Education*, *11*(4), 333–346. https://doi.org/10.33225/jbse/12.11.333

Laliyo, L. A. R., Sumintono, B., & Panigoro, C. (2022). Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model. *Heliyon*, *8*(March), e09126. https://doi.org/10.1016/j.heliyon.2022.e09126

Laliyo, L. A. R., Tangio, J. S., Sumintono, B., Jahja, M., Panigoro, C. (2020). Analytic Approach of Response Pattern of Diagnostic Test Items in Evaluating Students' Conceptual Understanding of Characteristics of Particle of Matter. *Journal of Baltic Science Education*, *19*(5), 824-841.

Landis, J. R., & Koch, G. G. (1977). Landis and Koch1977_agreement of categorical data. *Biometrics*, *33*(1), 159–174.

Lin, J. W., & Chiu, M. H. (2007). Exploring the characteristics and diverse sources of students' mental models of acids and bases. *International Journal of Science Education*, *29*(6), 771–803. https://doi.org/10.1080/09500690600855559

Linacre, J. M. (2020). *A User's Guide to W I N S T E P S ® M I N I S T E P Rasch-Model Computer Programs Program Manual 4.5.1*. www.winsteps.com.

Linn, R. L., & Slinde, J. A. (1977). The Determination of the Significance of Change Between Pre- and Posttesting Periods. *Review of Educational Research*, *47*(1), 121–150. https://doi.org/10.3102/00346543047001121

Liu, X. (2012). Developing measurement instruments for science education research. In B. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 651–665). Springer Netherlands.

Lu, S., & Bi, H. (2016). Development of a measurement instrument to assess students' electrolyte conceptual understanding. *Chemistry Education Research and Practice*, *17*(4), 1030–1040. https://doi.org/10.1039/c6rp00137h

Maratusholihah, N. F., Sri, R., & Fajaroh, F. (2017). Analisis miskonsepsi siswa sma pada materi hidrolisis garam dan larutan penyangga [Analysis of high school students' misconceptions on the material of salt hydrolysis and buffer solution]. *Jurnal Pendidikan*, *2*(7), 919-926. https://doi.org/10.17977/jptpp.v2i7.9645

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

Naah, B. M., & Sanger, M. J. (2012). Student misconceptions in writing balanced equations for dissolving ionic compounds in water. *Chemistry Education Research and Practice*, *13*(3), 186–194. https://doi.org/10.1039/C2RP00015F

Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, *48*(3), 237–256. https://doi.org/10.1002/tea.20400

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Pearson Education Ltd.

Orgill, M. K., & Sutherland, A. (2008). Undergraduate chemistry students' perceptions of and misconceptions about buffers and buffer problems. *Chemistry Education Research and Practice*, *9*(2), 131–143. https://doi.org/10.1039/b806229n

Orwat, K., Bernard, P., & Migdał-Mikuli, A. (2017). Alternative conceptions of common salt hydrolysis among upper-secondary school students. *Journal of Baltic Science Education*, *16*(1), 64–76.

Owens, D. C., Sadler, T. D., & Friedrichsen, P. (2019). Teaching practices for enactment of socio-scientific issues instruction: an instrumental case study of an experienced biology teacher. *Research in Science Education, 51*(April)*,* 375–398. https://doi.org/10.1007/s11165-018-9799-3

Ozdemir, G., & Clark, D. (2009). Knowledge structure coherence in Turkish students' understanding of force. *Journal of Research in Science Teaching*, *46*(5), 570–596. https://doi.org/10.1002/tea.20290

Park, M., & Liu, X. (2019). An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. *Research in Science Education, 51*(Suppl 1), 43–60. https://doi.org/10.1007/s11165-019-9819-y

Pentecost, T. C., & Barbera, J. (2013). Measuring learning gains in chemical education: A comparison of two methods. *Journal of Chemical Education*, *90*(7), 839–845. https://doi.org/10.1021/ed400018v

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Perera, C. J., Sumintono, B., & Jiang, N. (2018). The psychometric validation of the principal practice questionnaire based on Item Response Theory. *International Online Journal of Educational Leadership*, *2*(1), 21–38. https://doi.org/10.22452/iojel.vol2no1.3

Rouquette, A., Hardouin, J. B., Vanhaesebrouck, A., Sébille, V., & Coste, J. (2019). Differential Item Functioning (DIF) in composite health measurement scale: Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. *PLoS ONE*, *14*(4), 1–16. https://doi.org/10.1371/journal.pone.0215073

Sadler, P. M. (1998). Psychometric models for student-conceptions in science: Reconciling qualitative studies and distractor-driver assessment instruments. *Journal of Research in Science Teaching*, *35*(3), 265–296.

Sadler, P. M. (1999). The relevance of multiple-choice testing in assessing science understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 251–274). Elsevier Academic Press.

Seçken, N. (2010). Identifying student's misconceptions about SALT. *Procedia - Social and Behavioral Sciences*, *2*(2), 234–245. https://doi.org/10.1016/j.sbspro.2010.03.004

Sesen, B. A., & Tarhan, L. (2011). Active-learning versus teacher-centered instruction for learning acids and bases. *Research in Science and Technological Education*, *29*(2), 205–226. https://doi.org/10.1080/02635143.2011.581630

Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649

Soeharto, S., & Csapó, B. (2021). Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. *Heliyon*, *7*(11). https://doi.org/10.1016/j.heliyon.2021.e08352

Solihah, M. (2015). *Penggunaan instrumen diagnostik two-tier untuk mengidentifikasi pemahaman konsep siswa kelas XI SMA Negeri se-Kota malang pada materi larutann penyangga* [The use of two-tier diagnostic instruments to identify the understanding of the concept of class XI SMA Negeri Malang City on the material buffer solution]. Fakultas MIPA UM.

Sumintono, B. (2018). Rasch model measurements as tools in assessment for learning. *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017), October 2017*. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial* [Rasch model application for social science research]. Penerbit Trim Komunikata.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan* [Rasch modeling application in educational assessment]. Penerbit Trim Komunikata.

Sunyono, S., Tania, L., & Saputra, A. (2016). A learning exercise using simple and real-time visualization tool to counter misconceptions about orbitals and quantum numbers. *Journal of Baltic Science Education*, *15*(4), 452–463. https://doi.org/10.33225/jbse/16.15.452

Suteno, I. K., Laliyo, L. A. R., Iyabu, H., & Abdullah, R. (2021). Mengevaluasi level pemahaman konsep hidrolisis garam peserta didik menggunakan tes diagnostik pilihan ganda empat tingkat [Evaluating the level of understanding of the salt hydrolysis concept of students using a four-level multiple choice diagnostic test]. *Jurnal Pendidikan Sains Indonesia*, *9*(3), 482–497. https://doi.org/10.24815/jpsi.v9i3.20543

Taber, K. S. (2002). *Chemical misconceptions—Prevention, diagnosis, and cure*. Royal Society of Chemistry.

Taber, K. S. (2009). Challenging Misconceptions in the Chemistry Classroom: Resources to Support Teachers. *Educació Química EduQ*, *4*, 13–20. https://doi.org/10.2346/20.2003.02.27

Taber, K. S. (2014). Ethical considerations of chemistry education research involving "human subjects." *Chemistry Education Research and Practice*, *15*(2), 109–113. https://doi.org/10.1039/c4rp90003k

Taber, K. S. (2019). *The nature of the chemical concept re-constructing chemical knowledge in teaching and learning*. Royal Society of Chemistry. https://doi.org/10.1039/9781788013611-FP001

Tarhan, L., & Acar-Sesen, B. (2013). Problem based learning in acids and bases: Learning achievements and students' beliefs. *Journal of Baltic Science Education*, *12*(5), 565. https://doi.org/1648-1652

Tümay, H. (2016). Emergence, learning difficulties, and misconceptions in chemistry undergraduate students' conceptualizations of acid strength. *Science and Education*, *25*(1–2), 21–46. https://doi.org/10.1007/s11191-015-9799-x

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

Weston, M., Houdek, K. C., Prevost, L., Urban-Lurain, M., & Merrill, J. (2015). Examining the impact of question surface features on students' answers to constructed-response questions on photosynthesis. *CBELife Science Education*, *14*, 1–12.

Wilson, M. (2005). *Constructing measures: an item response modeling approach*. Lawrence Erlbaum Associates, Inc. https://doi.org/10.4324/9781410611697

Wilson, M. (2008). Cognitive diagnosis using item response models. *Zeitschrift Für Psychologie / Journal of Psychology*, *216*(2), 74–88. https://doi.org/10.1027/0044-3409.216.2.74

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*(6), 716–730. https://doi.org/10.1002/tea.20318

Wilson, M. (2012). Responding to a challenge that learning progressions pose to measurement practice. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progression in science* (pp. 317–344). Sense Publishers. https://doi.org/10.1007/978-94-6091-824-7

Wind, S. A., & Gale, J. D. (2015). Diagnostic opportunities using Rasch measurement in the context of a misconceptions-based physical science assessment. *Science Education*, *99*(4), 721–741. https://doi.org/10.1002/sce.21172

Yaşar, I. Z., İnce, E., & Kırbaşlar, F. G. (2014). 7. Class science and technology course "structure of atom" subject readiness improvement test. *Procedia - Social and Behavioral Sciences, 152*(October), 662-667. https://doi.org/10.1016/j.sbspro.2014.09.259

Yildirir, H. E., & Demirkol, H. (2018). Identifying mental models of students for physical and chemical change. *Journal of Baltic Science Education*, *17*(6), 986–1004. https://doi.org/10.33225/jbse/18.17.986

Received: *July 06, 2022* Revised: *September 05, 2022* Accepted: *October 06, 2022*

Cite as: Laliyo, L. A. R., La Kilo, A., Paputungan, M., Kunusa, W. R., Dama, L., & Panigoro, C. (2022). RASCH modelling to evaluate reasoning difficulties, changes of responses, and item misconception pattern of hydrolysis. *Journal of Baltic Science Education*, *21*(5), 817-835. https://doi.org/10.33225/jbse/22.21.817

Lukman A. R. Laliyo *(Corresponding author)*	PhD, Associate Professor, Gorontalo State University, Jl. Jend. Sudirman No. 6, Dulalowo Tim., Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128, Indonesia. E-mail: lukman.laliyo019@gmail.com Website: https://scholar.google.co.id/citations?user=owQHcqkAAAAJ&hl=id ORCID: https://orcid.org/0000-0003-3281-7202
Akram La Kilo	PhD, Assistant Professor, Gorontalo State University, Jl. Jend. Sudirman No. 6, Dulalowo Tim., Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128, Indonesia. E-mail: akram@ung.ac.id Website: https://scholar.google.co.id/citations?user=YWz4wUMAAAAJ&hl=id ORCID: https://orcid.org/0000-0002-4885-1838
Mardjan Paputungan	MSc, Associate Professor, Gorontalo State University, Jl. Jend. Sudirman No. 6, Dulalowo Tim., Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128, Indonesia. E-mail: marpa@ung.ac.id Website: https://www.ung.ac.id/formasi/people/196002151988031001 ORCID: https://orcid.org/0000-0003-3015-6965
Wiwin Rewini Kunusa	MSc, Assistant Professor, Gorontalo State University, Jl. Jend. Sudirman No. 6, Dulalowo Tim., Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128, Indonesia. E-mail: rewinikunusa2014@gmail.com Website: https://scholar.google.com/citations?user=5dukwJYAAAAJ&hl=en ORCID: https://orcid.org/0000-0001-5606-9195
Lilan Dama	PhD, Assistant Professor, Gorontalo State University, Jl. Jend. Sudirman No. 6, Dulalowo Tim., Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128, Indonesia. E-mail: lilandama@ung.ac.id Website: https://scholar.google.co.id/citations?user=cpg6CX8AAAAJ&hl=id ORCID: https://orcid.org/0000-0003-3759-6082
Citra Panigoro	MSc, Assistant Professor, Gorontalo State University, Jl. Jend. Sudirman No. 6, Dulalowo Tim., Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128, Indonesia. E-mail: citrapanigoro@ung.ac.id ORCID: https://orcid.org/0000-0003-3565-6694