PUBLICATION HISTORY

“Measuring Changes in Students’ Understanding Ability of the Hydrolysis Concept: Stacking and Racking Techniques in Rasch Measurement Analysis”

By: Lukman A. R. Laliyo

SUBMISSION

Gmail **Lukman Laliyo <lukman.laliyo019@gmail.com>**

Confirming submission to Heliyon

Heliyon <em@editorialmanager.com> Wed, Jul 7, 2021 at 4:54 PM
Reply-To: Heliyon <info@heliyon.com>
To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

CC: "Bambang Sumintono" bambang@um.edu.my, "Citra Panigoro" citrapanigoro@ung.ac.id

This is an automated message.

Measuring Changes in Students' Understanding Ability of the Hydrolysis Concept: Stacking and Racking Techniques in Rasch Measurement Analysis

Dear Dr. Laliyo,

We have received the above referenced manuscript you submitted to the Education section of Heliyon. It has been assigned the manuscript number HELIYON-D-21-05440. To track the status of your manuscript, please log in as an author at https://www.editorialmanager.com/heliyon/, and navigate to the "Submissions Being Processed" folder.

Thank you for submitting your work to Heliyon, an open access journal that is part of the Cell Press family.

Kind regards,
Heliyon

More information and support

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

---

Lukman Abdul Rauf Laliyo
Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo-Indonesia
Jenderal Sudirman Street No.6, Kota Tengah District, Gorontalo City 96128, Indonesia
lukman.laliyo019@gmail.com

Editor-in-Chief
Heliyon

July 7, 2021

Dear Editorial-in-Chief
I am pleased to submit an original research article entitled **"Measuring Changes in Students' Understanding Ability of the Hidrolysis Concept: Stacking and Racking Techniques in Rasch Measurement Analysis"** by Lukman Abdul Rauf Laliyo, Bambang Sumintono, Citra Panigoro for consideration for publication in Heliyon.

This manuscript aims to measure the changes in students' understanding ability of hydrolysis concept through a pretest-posttest quasi-experimental design and the stacking and racking techniques in the Rasch model. Further, we would like to highlights four major points of the findings in this research:

1. Raw scores have a bias in a conventional psychometric measurement.
2. Stacking and racking techniques measure the changes in students' ability and item difficulty level.
3. The learning process in the context of socio-scientific issues improves students' understanding.
4. The changes in students' hydrolysis understanding in the pre- and post-test with negative values result from misconceptions.

We believe this finding will contribute to the development of further research related to the teaching and learning activities.

We believe that this manuscript is appropriate for publication by Heliyon because it correlated with the prior publications in this journal.

This manuscript has not been published and is not under consideration for publication elsewhere. We also have no conflicts of interest to disclose.
Thank you for your consideration!

Sincerely,
Lukman Abdul Rauf Laliyo
Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo-Indonesia

Measuring Changes in Students' Understanding Ability of the Hydrolysis Concept: Stacking and Racking Techniques in Rasch Measurement Analysis

Lukman Abdul Rauf Laliyo[1,*], Bambang Sumintono[2], Citra Panigoro[3]

[1] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo, 96128, Indonesia
[2] Institute of Educational Leadership, University of Malaya, Kuala Lumpur, 50603, Malaysia
[3] Faculty of Fisheries and Marine Science, Universitas Negeri Gorontalo, 96128, Indonesia

*Corresponding Author

Email: lukman.laliyo019@gmail.com

Author's ORCID ID

Lukman Abdul Rauf Laliyo : https://orcid.org/0000-0003-3281-7202

Bambang Sumintono : https://orcid.org/0000-0002-5490-3665

Citra Panigoro : -

Highlights

- Raw scores have a bias in a conventional psychometric measurement
- Stacking and racking measure students' ability and item difficulty level changes
- The learning process in socio-scientific issues improves students' understanding
- Misconceptions influence the negative values of students' pre-and post test

Abstract

The present study aimed to measure the changes in students' understanding ability of hydrolysis concept through a pretest-posttest quasi-experimental design and the stacking and racking techniques in the Rasch model. Such a model was based on a person- and item-centered statistics to determine how students' ability changed during the treatment, as well as the changes in item difficulty level. Eleventh-grade students in one of the senior high schools in the eastern part of Indonesia were selected as the sample. The experimental class (N=57) was taught employing the scientific inquiry approach in the context of social-scientific issues; meanwhile, the control class (N=50) used the conventional method of learning. Moreover, this research utilized 15 three-tier multiple-choice items to evaluate students' understanding. The results of the *Mann-Whitney U* and *Wilcoxon* ($p < 0.05$) tests showed that the experimental class had better understanding ability of hydrolysis concept than the control class. Further analysis also found that the mean difference in the logit of the pre- and post-test items of the experimental class was better. In some cases, nevertheless, there were negative changes in students' ability due to the transition to a lower rating. It is concluded from the findings that learning innovations cannot solely change students' misconceptions. Besides, stacking and racking analyses are proven superior in estimating the changes in students' ability and item difficulty level. These research findings can also be a reference for researchers and practitioners of chemistry education to measure the changes and progress in students' learning ability.

Keywords: measuring, understanding changes, stacking, racking

Introduction

Studies of pedagogical innovation application in a pretest-posttest quasi-experimental design frequently examine an effect yet infrequently elaborate the changes brought forth by the effect in question. Pre- and post-test changes should be given in detail, both in terms of the students' ability and item difficulty level. However, this has not been the main focus in chemistry education research to date. Using raw scores that have a bias in conventional psychometric measurement is among the limitations [1]. Raw scores are not final data as they do not have the same measurement scale, nor do they have a great deal of information as the basis for drawing conclusions [2,3].

Around the 2000s, the Rasch model was introduced as a probabilistic-based psychometric measurement that went beyond the use of raw scores [4,5]. The model was formulated by George Rasch in the 1960s [3], and has been widely applied to analyze various types of data, e.g., dichotomous, polytomous, multi-rating, and multi-rater data. This measurement model is generally relied on to overcome the limitations of conventional psychometric measurement [1,6]. Its analyses, including item fit, PCA *(Principal Component Analysis)*, Wright map, are commonly used for international test analysis, namely TIMSS and PISA [7].

In chemistry education research, the Rasch model has been employed to evaluate learning understanding and progress [8], and to diagnose students' preconception [9,10] and misconceptions [11,12]. This model is also utilized to connect the measurement of content knowledge with pedagogical content knowledge [13], to diagnose the pattern of item difficulty [14], and to identify students' preconception that tends to be permanent [15]. Stacking technique is applied to measure the changes in students' ability; it was revealed that the post-test ability of children who learnt through games was better than their pre-test ability [16]. Thus far, the stacking and racking techniques in Rasch model to measure the changes in students' chemistry understanding have not been extensively used.

This research relied on the Rasch model's stacking and racking techniques to measure the changes in students' understanding of hydrolysis concepts through a pre-test and post-test. It intended to test the effectiveness of scientific inquiry approach in the context of social-scientific issue (SSI) in developing students' epistemological understanding and reasoning [17]. Experts believe that students already have their own understanding [18], yet it relatively comprises misconception [19]. Even some students tend to reject the scientific explanation that is contrary to what they have understood [20,21] and prefer to defend their misconceptions [9,15]. In terms of hydrolysis learning, such rejection often occurs when students are asked to describe the concept of acid and base strength [22, 23]. This study featured three specific questions, as follows: (1) is there any significant difference in hydrolysis understanding ability of students in the experimental and control classes? (2) in terms of students' ability, how are the changes in hydrolysis understanding during the pre- and post-test of students in the experimental and control classes? (3) in terms of items, how are the changes in the difficulty level of each item in the pre- and post-test of the experimental and control classes?

Literature review

Understanding of hydrolysis concept

The concept of hydrolysis is a learning topic in high school that is strongly related to SSI. Students with a good understanding of the hydrolysis concept will manage to clarify scientifically why detergents, bleaching agents (NaOCl), and fertilizers can pollute the environment. Despite this, linking this issue as the problem in learning is rarely carried out. The learning process puts more emphasis on mastering theoretical concepts [17]. Consequently, students find it challenging to use their comprehension to explain socio-scientific phenomena around them [24]. This difficulty is on account of their misconceptions regarding acid-base reaction [25], making them unable to elaborate the concept of salt hydrolysis [23], particularly to determine acid and base strength [19]. In addition, it is reported that students are struggling with correctly explaining the dissolving process and the reaction of ionic compounds with water, writing down chemical equations, and having different interpretations about the dissolving process mentioned earlier [21].

Socio-scientific issues (SSI) as the learning context

SSI as the learning context is performed by integrating socio-scientific issues in students' surroundings with certain topics, e.g., hydrolysis of salt. It trains students to develop scientific literacy skills [26, 27, 28].. The selected SSI is more likely to have a conceptual bond with science [24, 29], and its resolution requires many perspectives [28], including the dimension of moral and ethical evaluation of students [30]. For instance, the contexts of climate change, pollution, and global warming [17,31] are placed as socio-scientific phenomena that the students should elucidate based on insights that have been learned. On this ground, contexts function to encourage students to actively get involved in grasping problems [32], developing and utilizing their knowledge [33], improving their critical thinking [34], and being able to scientifically describe the discussed socio-scientific phenomena [17, 31, 35]. In the end, the integration of SSI contexts enables the learning process to be more significant in enhancing students' understanding [31,36], and students are skilled to negotiate about the social aspect of the studied phenomena [37,38].

Method of study

Research approach

This quantitative research relied on a control group pretest-posttest quasi-experimental design [39] by applying the racking and stacking techniques in the Rasch model [40] for data analysis. As standard techniques, racking and stacking are introduced by Benjamin Wright to measure the extent to which students and items change before and after treatment [41]. This kind of information is immensely helpful for teachers, especially in devising learning strategies that meet students' needs [16]. There may be some cases that learning implementation does not match students' characteristics or needs, so that the shortcomings of the applied pedagogical innovation can be evaluated.

Subject

Eleventh-grade students in one of the senior high schools in the eastern part of Indonesia participated as the subject. This study was conducted in the first semester of the 2019-2020 academic year, in which research permission was obtained from the government and school administrators. Students taken part in this study had volunteered to participate. On top of that, the purpose of the research was also informed to the students. The gained information will be confidential and only used for science development [42]. The 16- to 17-year-old students in the experimental class (N=57) and control class (N=50) were determined randomly out of ten classes.

Table 1. Conceptual Map of the Understanding of Hydrolysis Concept.

Item	Understanding Ability	Level
1	Balancing the reaction of sodium hypochlorite salt (NaOCl) hydrolysis in water	2
2	Stating the partial hydrolysis reaction: $\mathrm{NaOCl} \rightarrow \mathrm{Na^+} + \mathrm{OCl^-}$	2
3	Determining corrosive alkali of sodium hypochlorite salt (NaOCl)	1
4	Calculating the pH of hydrolysis of sodium hypochlorite salt (NaOCl) with NaOCl = 0.1 M ; Ka = 10^{-5})	3
5	Determining the properties of NaOCl in the reaction: $\mathrm{OCl^-} + \mathrm{H_2O} \rightarrow \mathrm{HOCl} + \mathrm{OH^-}$	2
6	Calculating the pH of sodium hypochlorite salt (NaOCl) that comes from a mixture of HOCl and NaOH (partially hydrolyzed), if the Ka HOCl = 10^{-5} and there is an increase in the pH of the solution mixture.	3
7	Determining aluminum salt ($Al_2(SO_4)_3$) properties in water	1
8	Determining aluminum salt ($Al_2(SO_4)_3$) properties in water that is partially	1

	hydrolyzed by the Al^{3+} ion	
9	Determining the properties of detergent solution causing eutrophication	1
10	Determining the properties of detergent solution (sodium tripolyphosphate salt) that is partially hydrolyzed	1
11	Determining the impact of the disposal of detergent waste on the environment	2
12	Determining the properties of ammonium sulfate salt $(NH4)_2SO_4$	1
13	Stating the equation of $(NH_4)_2SO_4$ reaction in water, partially hydrolyzed	2
14	Students' attitude towards the use of monosodium glutamate ($C_5H_8NO_4Na$)	2
15	Determining the properties of monosodium glutamate salt ($C_5H_8NO_4Na$)	1

Note: Level 1 = Determining the hydrolyzed salt properties,

Level 2 = Stating hydrolysis reactions of different salts in water,

Level 3 = Calculating the pH of the hydrolyzed salt solution.

Instrument

This present work utilized a three-tier multiple-choice test (henceforth, 3TMC) as the instrument to evaluate students' understanding of the hydrolysis concept. 3TMC followed the Standard of Chemistry 2013 Curriculum of Senior High School under the Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 37 of 2018. It was developed by following recommendations from [43, 44, 45]. 3TMC consisted of 15 items (Table 1) with different construction levels. Each item integrated diagnostic and summative measurements [46] and certainty of response index (CRI) [47,48]. The validity of the test construction by three experts was stated under Fleiss' kappa (K = .96), implying that the experts agreed that the validity of 3TMC items was considered good.

Each 3TMC item featured three questions (Q1, Q2, Q3), in which every response of the students to each item (Q1, Q2, Q3) was adjusted to the rubric (Table 2). For example, students' responses to items were as follows: Q1, Q2 "correct", and Q3 "certain" under the code CCC, category of Scientific Knowledge (SK). On the other hand, if the response patterns in Q1, Q2 are "incorrect" and Q3 "uncertain", the code will be IIU, category of Lack of Knowledge (LOK). The CCC and IIU understandings were rated six and one, respectively.

Table 2. All Possibilities of Responses.*)

(Q1)	(Q2)	(Q3)	Code	Conceptual Understanding Category	Rating
Correct	Correct	Certain	CCC	Scientific Knowledge (SK)	6
Correct	Incorrect	Certain	CIC	Misconception False Positive (MFP)	5
Incorrect	Correct	Certain	ICC	Misconception False Negative (MFN)	4
Incorrect	Incorrect	Certain	IIC	All-Misconception (ALM)	3
Correct	Correct	Uncertain	CCU	Lack of Confidence/Lucky Guess. (LG)	2
Correct	Incorrect	Uncertain	CIU	Lack of Knowledge (LOK)	1
Incorrect	Correct	Uncertain	ICU	Lack of Knowledge (LOK)	1
Incorrect	Incorrect	Uncertain	IIU	Lack of Knowledge (LOK)	1

*[47, 48, 49]

Data collection

Pre-test and post-test

Before the intervention, this research underwent a pre-test; whereas, a post-test was done after the intervention. Both tests were supervised by teachers in the school. The construction of pre- and post-test items was the same. Students wrote down their responses on the provided answer sheet. The students must work on all items according to the allocated time (45 minutes). The instrument was immediately collected and should have the same quantity as the research subject.

Learning implementation

Two chemistry teachers (teacher A and teacher B) were chosen to be the instructors of experimental and control classes. This implementation scenario was adapted from a study reported by Grooms [37]. At the beginning of the learning process, students were grouped into five—each consisted of six students. Next, in the observation stage, students were asked to watch a video for five minutes; the video represents socio-scientific issues around the students. In the questioning stage, they were stimulated to formulate questions regarding their observation, as well as formulating hypotheses. The students experimented by following the student worksheet and were accompanied by the teacher in the information collection stage. Thereupon, students had presentation and discussion sessions, during which they reported their experiment results and drew conclusions [50,51]. Meanwhile, in the control class, the learning process was performed conventionally and focused more on content mastery and problem-solving practice. The teacher also facilitated learning initiatives.

Data analysis

A non-parametric statistical test of Mann-Whitney U and a sign test of Wilcoxon were applied to determine students' ability difference in the pre- and post-test in inter-class and intra-class [52]. This research used WINSTEP 4.5.5 software as the Rasch model analysis [6]. The changes in students' ability were estimated by stacking and racking techniques. The stacking technique was to analyze the changes in students' ability, and the racking technique was to investigate the changes in item difficulty level [40].

The stacking technique put both pre- and post-test data vertically. Every student appeared twice in the data set, and each item emerged once in experimental and control classes [53]. This allowed the researchers to check out any changes of the students after the intervention [40]. All students had two rows of data, i.e., pre-test and post-test [54]. The students being examined should be based on the same item, making the changes in students' ability during the pre- and post-test could be measured [40]. Students' ability in each item of pre- and post-test could be compared as the data were analyzed in one single measurement, yet resulting in two item measures for every student and one measure for every item. Conversely, the racking analysis put both pre- and post-test data horizontally. Every item appeared twice in the data set, and each student emerged once; this enabled the researchers to check out any changes in item difficulty level before and after the intervention [40].

Ethics

The study is carried out within the guidelines of State University of Gorontalo for Research Data, including gathering informed consent from the from the local government and the school administrators on behalf of the students. The purpose of the research was also informed to the students. As the gained information will be confidential and only used for science development, the students' names are anonymized.

Results

Instrument effectiveness

Unidimensional and assessment scale analysis

The 3TMC instrument has a good unidimensionality (Appendix 1). Raw variant index arrives at above 40% (41.3%), meaning that the instrument can effectively measure students' understanding ability of hydrolysis concept [55]. Assessment scale analysis (Appendix 2) informs that the observation mean starts from logit -.34 (category 1 , LOK) to logit +1.11

(category 6, SK). This indicates that the category of students' understanding takes place consistently [6].

Person and item reliability

Based on Table 3, the person reliability index (.78) shows good response consistency of the students [56], implying that the scale is able to discriminate the category of students' understanding ability properly. This interpretation also applies to the item reliability index (.97); students' responses to items are excellent. The high value of item reliability also reveals that the item can define other variables very well [57]. Thus, the 3TMC instrument is reliable to be employed in the experimental and control classes.

Table 3. Reliability of Person and Item.

Parameter (N)	Mean Logit (SD)	Separation	Reliability	KR(20)
Person (214)	.69 (.59)	1.90	.78	.85
Item (15)	.00 (.41)	6.04	.97	

Cronbach alpha

According to Rasch model calculation, the coefficient of Cronbach Alpha (.85) reflects very high interaction between 214 students and 15 items with excellent category [3, 58]. The instrument of 3TMC has an excellent internal psychometric consistency and is considered very reliable.

Person and item separation index

Person separation index estimates how well 3TMC differentiates students' understanding ability of the hydrolysis concept. The higher the person separation index, the more likely students are to respond to items with correct understanding. The item separation index, on the contrary, shows how broad the item distribution in defining easy and difficult items. The broader the item distribution, the better and more appropriate it is [4]. In this study, the person separation index (1.90) and item separation index (6.04) reflect a moderate 3TMC distribution to students and items. Such criteria support 3TMC as the appropriate and reliable instrument to evaluate students' understanding of the hydrolysis concept.

Item validity

Table 4 illustrates the order of item suitability. An item is considered to experience a misfit if the measurement result is not in line with the following criteria: Outfit mean square residual (MNSQ): $.5 < y < 1.5$; Outfit standardized mean square residual (ZSTD): $-2 < Z < +2$; and point measure correlation (PTMEA CORR): $.4 < x < .8$ [4]. Item 15 does not meet the Outfit MNSQ criterion; Item 15, 6, 12, and 13 are not in accordance with Outfit (ZSTD) criterion. No items experience misfit or have negative values in terms of PTMEA CORR criterion; simply put, all items fulfill those criteria mentioned previously and are suitable and valid. This result also suggests that the 3TMC instrument has good measurement effectiveness.

Table 4. Item Statistics: Misfit Order

Item	Measure	Outfit MNSQ	Outfit ZSTD	PTMEA CORR.
1	-.55	1.18	.96	.52
15	.35	1.55	4.29	.52
14	-.35	1.14	.83	.54
6	.34	1.30	2.53	.44
11	-.52	.93	-.34	.56
2	.23	1.14	1.20	.55
10	.08	1.14	1.06	.59

9	-.61	.86	-.71	.55
8	.61	1.10	.96	.45
3	-.44	.81	-1.09	.49
5	-.30	.81	-1.19	.61
7	.19	.92	-.60	.58
4	.38	.90	-.95	.55
12	-.01	.71	-2.29	.63
13	.58	.71	-3.09	.64

The difference in students' understanding ability of hydrolysis concept

Based on person measure data of every student, we statistically test the difference in students' understanding of hydrolysis concept between pre- and post-test in the experimental class and control class, as well as the difference in pre- and post-test between the aforementioned classes using the non-parametric tests of Mann-Whitney U and Wilcoxon, presented in the following Table 5 and 6.

Table 5. The Result of the Mann-Whitney U Test based on students' pre-test and Post-Test Ability in Experimental and Control Classes ($p < 0.05$).

Test	Experimental Class (N=57)	Control Class (N=50)	*U*	*p*
Pre-test	0.65(-0.59-1.10)[a]	0.18(-1.33-0.85)[a]	667.500	0.000
Post-test	1.91(-0.03-3.62)[a]	1.24(-1.57-1.74)[a]	282.000	0.000

[a]*descriptive statistics given as median (Min-Max)*

Table 6. The Result of the Wilcoxon Test of Students' Pre-Test and Post-Test in Experimental and Control Classes ($p < 0.05$).

Class	Pre-test	Post-test	Z	p*
Experimental	0.65(-0.59-1.10)[a]	1.91(-0.03-3.62)[a]	-6.570	0.000
Control	0.18(-1.33-0.85)[a]	1.24(-1.57-1.74)[a]	-6.147	0.000

[a]*descriptive statistics given as median (Min-Max)*

The result of the Mann-Whitney test (Table 5) brings out the fact that statistically, there is a significant difference in the results of pre-test ($U = 667.500$, $p < 0.05$) and post-test (282.000, $p < 0.05$) among students in experimental and control classes. Further, the Wilcoxon test result (Table 6) shows that the results of pre-test ($z=-6.570$, $p < 0.05$) and post-test ($z = -6.147$, $p < 0.05$) of students in experimental and control classes are significantly different. This can be seen that students' ability after treatment (post-test) is higher than before treatment (pre-test), both in experimental and control classes. The ability of students in the experimental class, however, is better than those in the control class. Accordingly, such findings indicate that the learning process in the context of SSI is more effective than conventional learning.

The changes in students' understanding ability of hydrolysis concept

From Table 7, it is worth noting that although the logit mean of pre-test items in experimental and control classes is not relatively different, the logit mean of post-test items shows a relatively major difference. The mean of logit difference of the pre- and post-test items in the experimental class (.78) is greater than that of the control class (.50), meaning that the ability changes of students in the experimental class are better than those in the control class.

Table 7. Mean of the Pre- and Post-Test Items Logit Difference in Experimental and Control Classes.

Class	Total		Mean		
	Students	Items	Pre-test Item Logit	Post-test Item Logit	Item Logit Difference
Experimental	57	15	.52	1.30	.78
Control	50	15	.17	.67	.50

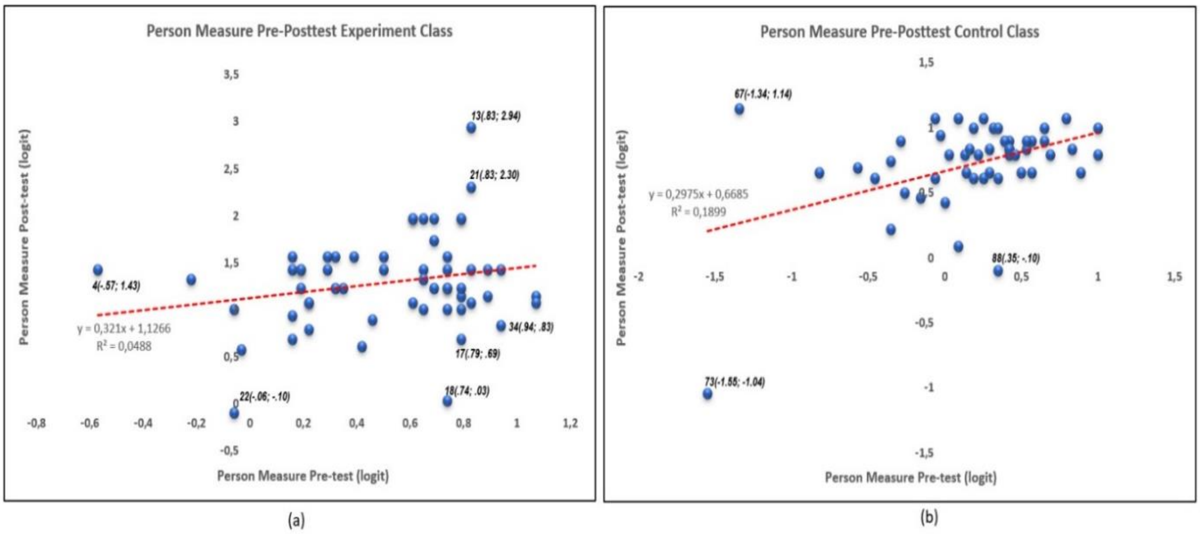

Fig. 1. Scatter Plots of Person Measures in Pre-Posttest of Experimental and Control Classes.

The above figure presents scatter plots of pre- and post-test in experimental and control classes as a line that intersects the horizontal axis. The gradient = 1 is plotted through the mean measurement of person in the pre- and post-test. As illustrated in Figure 1, the graph of person measure of the experimental class (a) looks more convincing than that of the control class (b) the range of logit scale of the vertical axis (post-test) of the experimental class (3.5 logit) is higher than that of the control class (1.5 logit). Besides, 28 students in the experimental class are above the curve, and 29 students are under the curve. For the control class, 22 students and 28 students are, in succession, above and under the curve. After treatment, students above the curve show a more significant improvement of ability than the mean of logit difference of pre- and post-test items in experimental (.78 logit) and control (.50 logit) classes.

Table 8. Student Scalogram (The Difference in Ability Measure of Pre- and Post-Test with Negative Values).

Student Serial Number	Class	Item (N-15) (Sorted by Difficulty Level)		DPP	CSE
			| 1 1 11 1 1		
			|911345207265438 (M) (MSE)		
			|---------------		
17	Experimental	Pre-test	+566564553566436 (.79) (.22)	-.10	.43
		Post-test	+655564563366553 (.69) (.21)		
18	Experimental	Pre-test	+666566636366333 (.74) (.21)	-.71	.39
		Post-test	+666661322521161 (.03) (.18)		
22	Experimental	Pre-test	+616664245121133 (-.06)(.18)	-.04	.36
		Post-test	+666662123521211 (-.10)(.18)		
34	Experimental	Pre-test	+666535666564653 (.94) (.24)	-.11	.47
		Post-	+666636466636453 (.83) (.23)		

		test			
72	Control	Pre-test	+666635656565653 (1.00)(.25)	-.21	.47
		Post-test	+666646655653533 (.79) (.22)		
81	Control	Pre-test	+566636563566356 (.89) (.23)	-.24	.43
		Post-test	+666664653653433 (.65) (.20)		
88	Control	Pre-test	+616665663261613 (.35) (.18)	-.45	.36
		Post-test	+612162566131613 (-.10)(.18)		

Note: M = measure, MSE= model standard error, DPP= difference in pre-post-test,

CSE = combined standard error. (1=LOK, 2=LG, 3=AM, 4=MFN, 5=MFP, 6=SK)

Red marks refer to post-test items that turn into a misconception

After learning treatment, the logit difference of pre- and post-test ability from 28 students (experimental class) and 22 students (control class) has successfully surpassed the mean of logit difference of pre- and post-test ability in each class. Next, the plotting of pre-test and post-test ability of four students in the experimental class (17, 18, 22, and 34) and three students in the control class (72, 81, and 88) is identified to be outside the modeled invariance curve. On top of that, the logit difference in pre-test and post-test ability of seven students mentioned earlier has a negative value; those students' ability is changed to lower understanding rating after treatment. The question is "on which item does it occur?" Provided in Table 8 is the result of the scalogram test of those seven students. However, this study only exemplifies one case, i.e., student 18 in the experimental class (hereinafter referred to as S18E).

S18E had the following ability measure: pre-test (.74 logit; SE .21 logit), post-test (.03 logit; SE .18 logit), and difference in pre- and post-test ability (-.71 logit; SE .39 logit); SE refers to standard error. Due to the fact that the logit of difference in pre- and post-test ability (-.71) was less than the combined standard error logit (.39), there was no significant difference in students' pre- and post-test ability. After a thorough investigation, in the pre-test ability (.74), S18E had a correct understanding on nine items (item 9, 1, 11, 14, 5, 12, 7, 6, and 15). In the post-test ability (.03), the student's understanding rating went lower on four items (item 5, 12, 6, and 15). It is assumed that such a change was in consequence of the student's misconception in stating the reaction of sodium hypochlorite salt (NaOCl) hydrolysis in the reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$ (item 5), determining the acid properties of ammonium sulfate salt $(NH_4)_2SO_4$ (item 12), calculating the pH of sodium hypochlorite salt (NaOCl) (item 6), and determining the base properties of monosodium glutamate salt (item 15).

Table 9. Data of Item Measures of Pre- and Post-Test of Experimental and Control Classes.

Item	Experimental Class (N=57)			Control Class (N=50)		
	Pre-test Item Logit	Post-test Item Logit	Item Logit Difference	Pre-test Item Logit	Post-test Item Logit	Item Logit Difference
1	.18	-3.25	-3.43	-.10	-1.19	-1.18
2	.99	.70	-.92	.60	-.66	-1.26

3	.37	-.89	-1.26	-.14	-1.3	-.89
4	.83	.50	-.33	.74	-.14	-.88
5	.22	-.71	-.93	.15	-.60	-.75
6	.48	.42	-.60	.52	.42	-.1
7	.91	-.35	-1.26	.44	-.50	-.49
8	.87	.78	-.90	.65	.54	-.11
9	.8	-1.5	-1.58	-.70	-1.46	-1.39
10	.59	-.19	-.78	.38	-.40	-.42
11	.18	-1.5	-1.68	-.60	-.60	.00
12	.61	-.24	-.85	.33	-.40	-.73
13	1.8	.36	-.72	.88	.36	-.52
14	.15	-1.14	-1.29	.19	-.72	-.91
15	.50	-.38	-.88	.59	.76	.17
Mean	.54	-.54	-1.70	.35	-.28	-.63

The difference in hydrolysis item difficulty level of students

Table 9 shows that the item difficulty level changes consistently in both experimental and control classes. The mean of item logit difference of the experimental class (-1.70) is less than that of the control class (-.63), signifying that after treatment, the item difficulty level of pre- and post-test in the experimental class turns to be easier than that of the control class. After examining the items, it is found that the difference in every pre- and post-test item logit in the experimental class is all negative in the value. On the other hand, the item logit difference in pre- and post-test in the control class has three items with negative values, one unchanging item (item 11), and one changing item being more difficult (item 15).

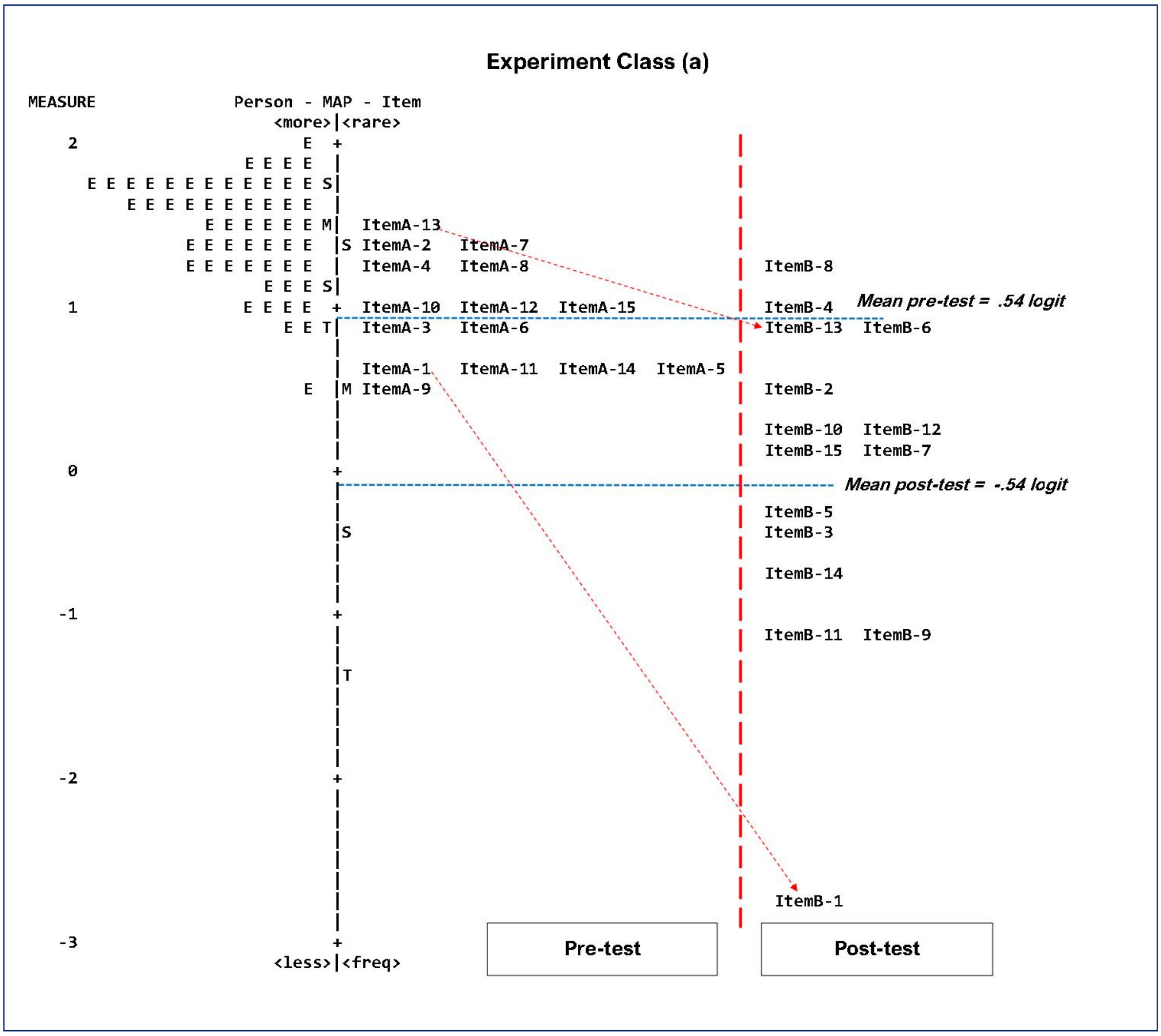

Fig. 2. Wright Map of Experimental Class Students.

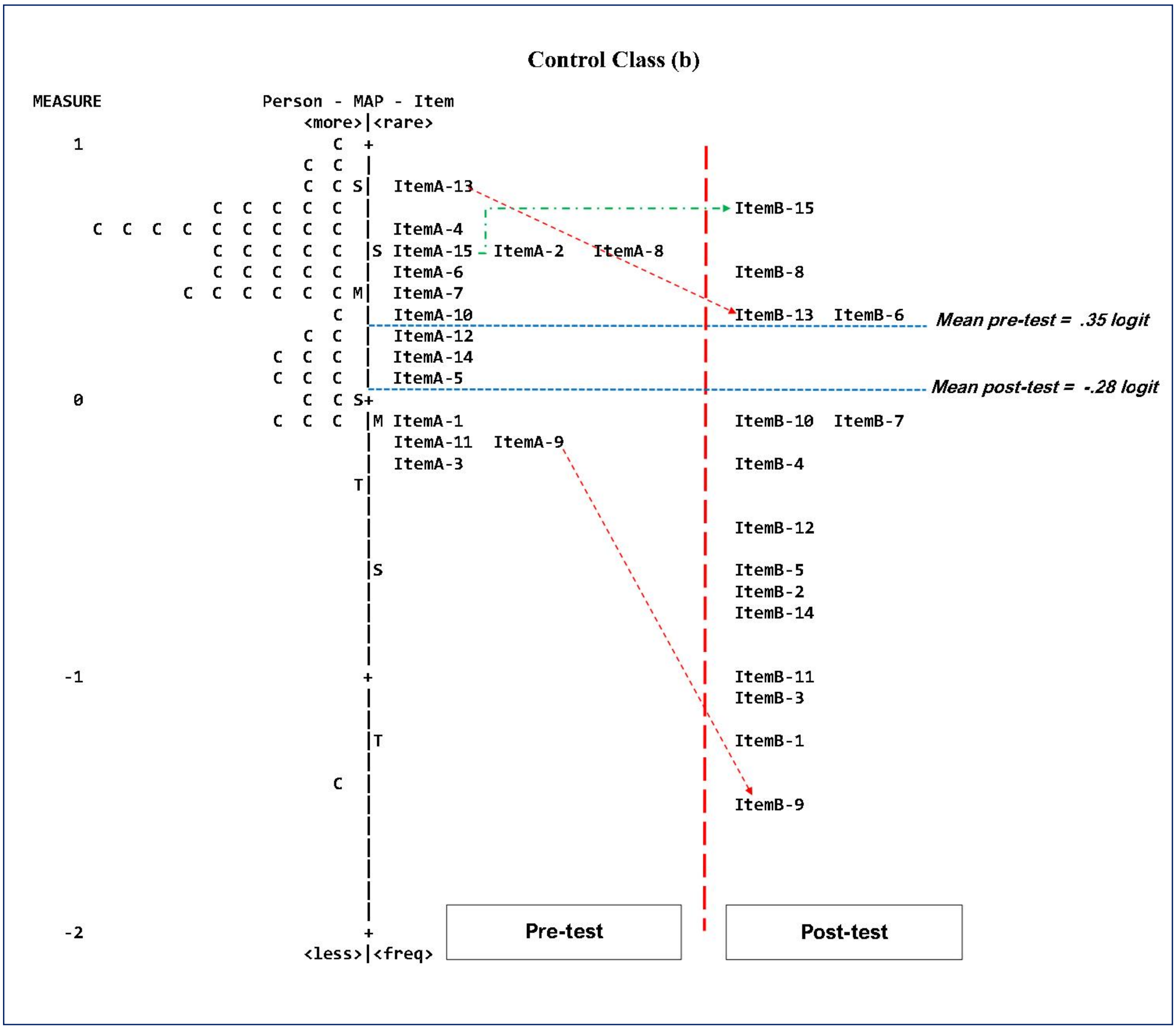

Fig. 3. Wright Map of Control Class Students.

Figure 2 and Figure 3 present the item Wright map to illustrate the changes in item difficulty level of pre- and post test in the experimental class and control class. Drawing from both figures in the pre-test, item 13 (measuring students' ability in stating the reaction of partial hydrolysis of $(NH_4)_2SO_4$ in water) has the highest difficulty level both in the experimental class (1.8 logit) and control class (.88 logit) Meanwhile, in the post-test, the measure of this item changes to .36 logit in the experimental class and .36 logit in the control class. The logit difference of pre-and post-test item 13 of the experimental class (-.72) is less than that of the control class (-.52), indicating that the change in difficulty level of item 13 in the experimental class is better than that of the control class.

Discussion

The changes in students' ability and item difficulty level

Producing information on the measurement of changes in students' ability and item difficulty level is among the advantages of stacking and racking techniques in the Rasch model. This allows researchers to detail the influence of the applied pedagogical innovation [13, 14, 59]. The stacking technique provides information regarding "who has changed"; in contrast, the racking technique offers information of "what has changed" [16,40]. Apart from positive changes, this study discovered that seven students experienced negative ability changes; S18E was among those students. His/her ability turned negative on items 5, 12, 6, and 15 after treatment. Why did it take place? How was the change? Figure 4 shows the result of the stacking analysis of S18E and racking analysis of item 5. Viewed from the measure of the post-test ability (.03) that is higher than the item 5 difficulty level of the post-test (-.71), the correct understanding of S18E on item 5 should have not changed. It is because items "being taught" mostly get way easier than the untaught ones [60]. Instead, S18E's understanding became negative, indicating that post-test item 5 was more difficult than the pre-test item 5. It is assumed that this issue was caused by the misconception of S18E in determining the hydrolysis reaction in water [19, 21, 61].

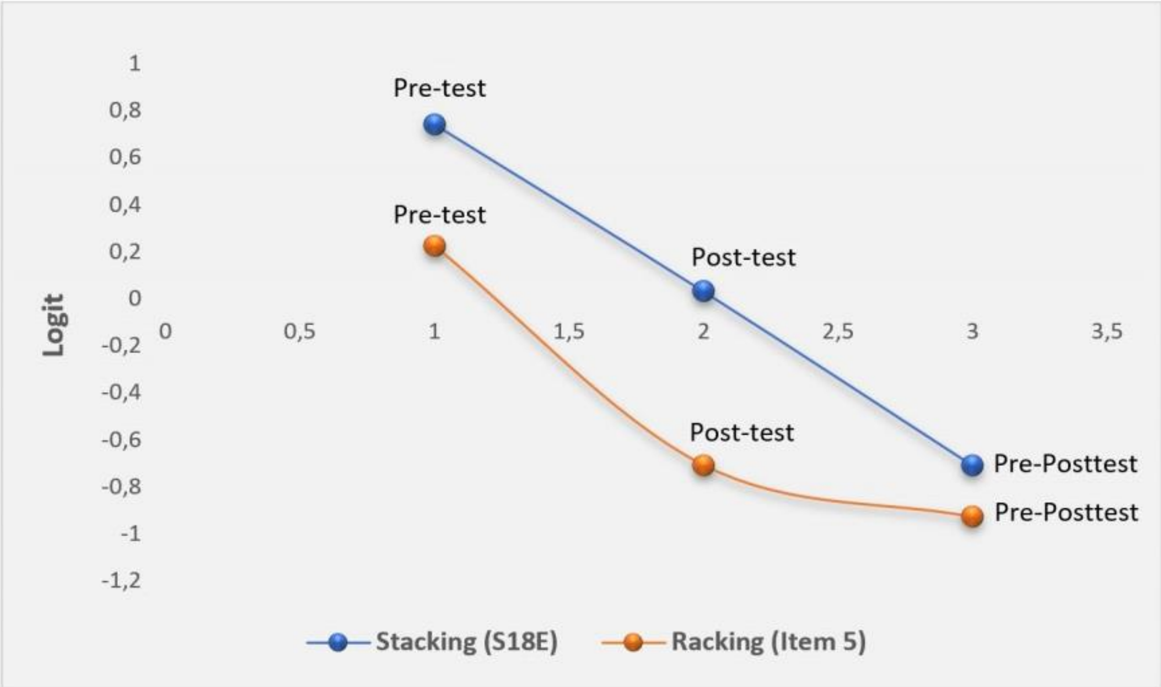

Fig. 4. Results of Stacking and Racking Analysis of Item 5.

The patterns of changes in students' ability and item difficulty level

The causes of these changes can be examined from the response patterns of S18E on item 5. This item has the student to pay attention to the reaction of NaOCl reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$ with the estimated pH = 7 and is alkaline. The question Q1 of this item is "is it correct that NaOCl is akaline?" S18E answered "correct" in the pre-test, yet responded to "incorrect" in the post-test. The next question Q2 is about the reason for students' answer in Q1, with four provided answers, as follows: (a) because NaOCl is formed from strong acids and weak bases; (b) because NaOCl is formed from weak acids and strong bases; (c) because NaOCl is formed from weak acids and weak bases; (d) because NaOCl is formed from strong acids and strong bases. In the pre-test, S18E chose the correct answer (b), yet s/he selected (b) in the post-test that contained misconception. Additionally, S18E chose the answer "certain" in question Q3 of the pre-test, and became "uncertain" in the post-test.

Therefore, the response patterns of S18E in the pre-test were CCC (category of scientific knowledge - SK) and IIU (category of lack of knowledge - LOK) in the post-test. The patterns of changes in the pre- and post-test of S18E were merely on account of the student's inability to understand the reaction of NaOCl formed by weak acids and strong bases. On this ground, S18E does not understand the concept of acid and base, as well as acid-base reaction. These findings strengthen previous studies [19, 21, 25, 62].

It is also intriguing to claim that the same techniques can examine the changes other students experience. Such information can be obtained through stacking and racking techniques in the Rasch model [3, 16, 40]. Although the result of 3TMC does not give information regarding changes, students' ability changes can be measured to be more clearly and accurately after being analyzed through the Rasch model. On top of that, other variables, namely sex, learning style, and the like, can function as additional information in connection with the pedagogical or curricular effects on students' ability [59].

Conclusion

The stacking technique provides information regarding changes in students' ability, allowing researchers to identify which students experience positive or negative changes after the learning treatment. Despite the fact that students also have negative changes, it is found out that students taught with the SSI context relatively have a better understanding ability of the hydrolysis concept than those who learn conventionally. On the other hand, the racking technique allows researchers to determine items that are considered easiest or most difficult. Therefore, the integration of both analysis techniques in the Rasch model is able to give accurate details in evaluating the influence of pedagogical innovations and student learning outcomes.

Limitations and further studies

This study is subject to several limitations that should be borne in mind, e.g., relatively limited sample size and other aspects that have not been considered: learning style, motivation, and hydrolysis preconception of students. Further studies are expected to investigate the correlation between those aspects and the effectiveness of changes in students' understanding ability. Moreover, this research did not take into account the impact of item characteristics on the parameter of item difficulty level, i.e., whether different levels of item difficulty are resulted from different understanding or other causes, such as different contexts of problems presented on each item [63]. In essence, this present work did not examine the effect of problem presentation contexts as item characteristics on the item difficulty level; nonetheless, it followed the two-step processes suggested in the racking technique [14, 40]. Further studies are also expected to find techniques to integrate contexts and item characteristics in a measurement model. It is assumed that different contexts of problem presentation on every item will influence measurement results.

In addition to relying on quantitative measurement, further studies should be strengthened by the analysis of a structured interview to delve into the aspects that drive students to ability changes. Therefore, the linkages between the process during treatment and ability changes can be elaborated in detail and accurately, i.e., the part of the process that leads students to change their understanding of specific ideas taught to them. By the interview, the impact of the learning implementation in the SSI context can be analyzed to determine its effect on the changes in understanding ability and item difficulty level.

Acknowledgment

We would like to express our deep gratitude to the Head of Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo.

Conflicts of interest

The authors declare no conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material

All the data are presented in the manuscript. The data supporting the findings of the article is also available in the appendix section in the end part of the manuscript.

Author Contributions

Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing: Lukman Abdul Rauf Laliyo

Validation, supervision, writing—review and editing: Bambang Sumintono

Project administration, resources, funding acquisition: Citra Panigoro

References

1. Sumintono B. Rasch model measurements as tools in assesment for learning. In: Proceedings of the 1st International Conference on Education Innovation (ICEI 2017); 2017 October 14; Surabaya. Paris: Atlantis Press; 2018. Available from: https://doi.org/10.2991/icei-17.2018.11
2. He P, Liu X, Zheng C, Jia M. Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. Che Edu Res & Prac. 2016;17(2):381–393. doi: 10.1039/C6RP00004E
3. Sumintono B, Widhiarso W. Aplikasi pemodelan Rasch pada assessment pendidikan [Application of Rasch modeling in educational assessment]. Bandung: Penerbit Trim Komunikata; 2015. https://www.researchgate.net/publication/282673464%0AAplikasi

4. Boone WJ, Yale MS, Staver JR. Rasch analysis in the human sciences. Dordrecht: Springer; 2014. doi: 10.1007/978-94-007-6857-4

5. Liu X. Developing measurement instruments for science education research. In: Fraser B, Tobin KG, McRobbie CJ, editors. Second International Handbook of Science Education. Netherlands: Springer; 2012. Pp. 651-665.

6. Linacre JM. 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. US: Winsteps; 2020.

7. Sabah S, Hammouri H, Akour M. Validation of a scale of attitudes toward science across countries using Rasch model: Findings from TIMSS. J of Baltic Sci Edu. 2013;12(5):692–703.

8. Hadenfeldt JC, Bernholt S, Liu X, Neumann K, Parchmann I. Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter. J of Che Edu. 2013; 90(12):1602–1608. doi: 10.1021/ed3006192

9. Laliyo L, Botutihe, & Panigoro C. The development of two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. Inter J of Learn, Teach & Edu Res. 2019;18(9):216–237. doi: 10.26803/ijlter.18.9.12

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

11. Herrmann-Abell CF, DeBoer GE. Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. Che Edu Res & Pract. 2011;2(2):184–192. doi: 10.1039/c1rp90023d

12. Herrmann-Abell CF, Deboer GE. Using Rasch modeling and option probability curves

to diagnose students' misconceptions. In: the 2016 AERA Annual Meeting; 2016 April 8-12; Washington DC. USA: American Educational Research Assossiation; 2016. Available from: https://www.aaas.org/sites/default/files/Project2061_CHA-AERA%20energy%20paper%204-7-16.pdf

13. Davidowitz B, Potgieter M. Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. Inter J of Sci Edu. 2016;38(9):1483–1503. doi: 10.1080/09500693.2016.1196843

14. Park M, Liu X. An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. Research in Science Education. 2019. doi: 10.1007/s11165-019-9819-y

15. Laliyo LAR, Tangio JS, Sumintono B, Jahja M, Panigoro C. Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter. J of Baltic Sci Edu. 2020;19(5). doi: 10.33225/jbse/20.19.824

16. Ling M, Pang V, Ompok CC. Measuring change in early mathematics ability of children who learn using games: stacked analysis in rasch measurement. In: Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings, 2016 July 30-August 3; Singapore. Singapore: Springer; 2018. Available from: https://doi.org/10.1007/978-981-10-8138-5

17. Kinslow AT, Sadler TD, Nguyen HT. Socio-scientific reasoning and environmental literacy in a field-based ecology class. Environ Edu Res. 2018;4622:1–23. doi: 10.1080/13504622.2018.1442418

18. Chandrasegaran AL, Treagust DF, Mocerino M. An evaluation of a teaching intervention to promote students' ability to use multiple levels of representation when

describing and explaining chemical reactions. Res in Sci Edu. 2008;38(2):237–248. doi: 10.1007/s11165-007-9046-9

19. Damanhuri MIM., Treagust DF, Won M, Chandrasegaran AL. High school students' understanding of acid-base concepts: An ongoing challenge for teachers. Inter J of Environ & Sci Edu. 2016;11(1):9–27. https://doi.org/10.12973/ijese.2015.284a

20. McClary LM, Bretz SL. Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. International J of Sci Edu. 2012;34(15):2317–2341. https://doi.org/10.1080/09500693.2012.684433

21. Orwat K, Bernard P, Migdał-Mikuli A. Alternative conceptions of common salt hydrolysis among upper-secondaryschool students. J of Baltic Sci Edu. 2017;16(1): 64–76.

22. Lin JW, Chiu MH. Exploring the characteristics and diverse sources of students' mental models of acids and bases. Inter J of Sci Edu. 2007;29(6):771–803. doi: 10.1080/09500690600855559

23. Seçken N. Identifying student's misconceptions about SALT. Procedia - Soc & Behav Sci. 2010;2(2):234–245. doi: 10.1016/j.sbspro.2010.03.004

24. Owens DC, Sadler TD, Friedrichsen P. Teaching practices for enactment of socio-scientific issues instruction: An instrumental case study of an experienced biology teacher. Res in Sci Edu. 2019. doi: 10.1007/s11165-018-9799-3

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

26. DeBoer GE. Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. J of Res in Sci Teach. 2000;37(6):582–601. doi: 10.1002/1098-2736(200008)37:6<582::AID-

TEA5>3.0.CO;2-L

27. Sadler TD. Socio-scientific issues-based education: What we know about science education in the context of SSI. In: Sadler TD, editor. Socio-Scientific Issues in the Classroom: Teaching, Learning and Research. Dordrecht: Springer; 2011. pp. 355–369. doi: 10.1007/978-94-007-1159-4_16.

28. Zeidler DL. Socioscientific issues as a curriculum emphasis: theory, research and practice. In: Lederman NG, Abell SK editor. Handbook of research on science education. New York: Routledge; 2014. pp. 697–726.

29. Bruder R, Prescott A. Research evidence on the benefits of IBL. ZDM - International J on Math Edu. 2013;45(6):811–822. doi: 10.1007/s11858-013-0542-2

30. Espeja AG, Lagarón DC. Socio-scientific Issues (SSI) in initial training of primary school teachers: Pre-service teachers' conceptualization of SSI and appreciation of the value of teaching SSI. Procedia – Soc & Behav Scie. 2015;196:80–88. doi: 10.1016/j.sbspro.2015.07.015

31. Sadler TD. Moral and ethical dimensions of socioscientific decision-making as integral components of scientific literacy. Sci Edu. 2004;13(1): 39–48.

32. National Research Council. A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington DC: The National Academies Press; 2012. doi: 10.17226/13165

33. Sadler TD, Zeidler DL. Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. J of Res in Scie Teach. 2009;46(8):909–921. doi:10.1002/tea.20327

34. Lederman NG, Lederman JS, & Antink A. Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. International J of Edu in Math, Sci & and Tech. 2013;1(3):138–147. doi: 10.18404/ijemst.19784

35. Cooke AN, Fielding KS, Louis WR. Environmentally active people: The role of autonomy, relatedness, competence and self-determined motivation. Environ Edu Res. 2016;22(5):631–657. doi: 10.1080/13504622.2015.1054262
36. Gräber W, Nentwig P, Becker HJ, Sumfleth E, Pitton A, Wollweber K, Jorde D. Scientific literacy: From theory to practice. Res in Sci Edu - Past, Present, & Future. 2005;1996:61–70. doi: 10.1007/0-306-47639-8_6
37. Grooms J. A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction. J of Chem Edu. 2020. doi: 10.1021/acs.jchemed.0c00026
38. Presley ML, Sickel AJ, Muslu N, Merle D. A framework for socio-scientific issues based education. Sci Edu. 2013;22:26–32.
39. Creswell JW. Educational research: Planning, conducting, and evaluating quantitative and qualitative research. 4th ed. New York: Pearson Education, Inc; 2012.
40. Wright BD. Rack and stack: Time 1 vs. time 2 or pre-test vs. post-test. Rasch Mea Transac. 2003;17(1):905–906.
41. Anselmi P, Vidotto G, Bettinardi O, Bertolotti G. Measurement of change in health status with Rasch models. Heal & Qual of Life Outc. 2015;13(1):1–7. doi: 10.1186/s12955-014-0197-x
42. Taber KS. Ethical considerations of chemistry education research involving "human subjects." Che Edu Res & Prac. 2014;15(2):109–113. doi: 10.1039/c4rp90003k
43. Wilson M. Constructing measures: an item response modeling approach. USA: Lawrence Erlbaum Associates, Inc; 2005. doi: 10.4324/9781410611697
44. Wilson M. Cognitive diagnosis using item response models. Zeits Für Psych/J of Psych. 2008;216(2):74–88. doi: 10.1027/0044-3409.216.2.74
45. Wilson M. Measuring progressions: Assessment structures underlying a learning

progression. J of Res in Sci Teach. 2009;46(6):716–730. doi: 10.1002/tea.20318

46. Hoe KY, Subramaniam R. On the prevalence of alternative conceptions on acid-base chemistry among secondary students: Insights from cognitive and confidence measures. Che Edu Res & Prac. 2016;17(2):263–282. doi: 10.1039/c5rp00146c

47. Arslan HO, Cigdemoglu C, Moseley C. A Three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. Inter J of Sci Edu. 2012;34(11):1667–1686. doi: 10.1080/09500693.2012.680618

48. Hasan S, Bagayoko D, Kelley EL. Misconceptions and the certainty of response index (CRI). Phy Edu. 1999;34(5):294–299. doi: 10.1088/0031-9120/34/5/304

49. Habiddin H, Page EM. Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). Indo J of Chem. 2019;19(3):720–736. doi: 10.22146/ijc.39218

50. Mitarlis, Ibnu S, Rahayu S, Sutrisno. The effectiveness of new inquiry-based learning (NIBL) for improving multiple higher-order thinking skills (M-HOTS) of prospective chemistry teachers. Euro J of Edu Res. 2020;9(3):1309–1325. doi: 10.12973/eu-jer.9.3.1309

51. Pedaste M, Mäeots M, Siiman LA, de-Jong T, Van-Riesen SAN, Kamp ET, Manoli CC, Zacharia ZC, Tsourlidaki, E. Phases of inquiry-based learning: Definitions and the inquiry cycle. Edu Res Rev. 2015;14:47–61. doi: 10.1016/j.edurev.2015.02.003

52. Günter T, Alpat SK. What is the Effect of case-based learning on the academic achievement of students on the topic of "Biochemical oxygen demand?". Res in Sci Edu. 2019;49(6):1707–1733. doi: 10.1007/s11165-017-9672-9

53. Wright BD. Time 1 to Time 2 (Pre-test to Post-test) comparison: Racking and Stacking. Rasch Mea Transac. 1996;10(1):478.

54. Herrmann-abell CF, Flanagan JC, Roseman JE. Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: Using student measures to evaluate the promise of the intervention. In: Proceeding of the 2013 NARST Annual International Conference. 2013 April 6-9; Puerto Rico. Puerto Rico: NARST; 2013. Available from: https://www.aaas.org/sites/default/files/project2061/files/NARST2013-Flanagan-Paper5-Usingteachermeasurestoe.pdf
55. Fisher WP. Rating scale instrument quality criteria. Rasch Measure Transac. 2007;21(1): 1095. www.rasch.org/rmt/rmt211m.htm
56. Sumintono B, Widhiarso W. Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial [Application of Rasch Model for Social Science Studies]. In Trim B, editor.). Bandung: Trim Komunikata Publishing House; 2014. https://www.researchgate.net/publication/268688933%0AAplikasi
57. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences. 3rd ed. New York: Routledge Taylor & Francis Group; 2015. doi: 10.1088/1751-8113/44/8/085201
58. Adams WK, Wieman CE. Development and validation of instruments to measure learning of expert-like thinking. Inter J of Sci Edu. 2011;33(9):1289–1312. doi: 10.1080/09500693.2010.512369
59. Pentecost TC, Barbera J. Measuring learning gains in chemical education: A comparison of two methods. J of Chem Edu. 2013;90(7):839–845. doi: 10.1021/ed400018v
60. Combrinck C, Scherman V, Maree D. Evaluating anchor items and reframing assessment results through a practical application of the Rasch measurement model. South African J of Psy. 2017;47(3):316–329. doi: 10.1177/0081246316683569
61. Lu S, Bi H, Liu X. The effects of explanation-driven inquiry on students' conceptual

understanding of redox. Inter J of Sci Edu. 2018;40(15): 1857–1873. doi: 10.1080/09500693.2018.1513670

62. Ültay N, Çalik M. A comparison of different teaching designs of “acids and bases” subject. Eura Jof Math, Sci & Tech Edu. 2016;12(1):57–86. doi: 10.12973/eurasia.2016.1422a

63. Hartig J, Frey A, Nold G, Klieme E. An application of explanatory item response modeling for model-based proficiency scaling. Educational and Psychological Measurement. 2012;72(4):665–686. doi: 10.1177/0013164411430707

Appendix 1. Standardized Residual Variance in Eigen value Units.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	25.5726	100.0%		100.0%
Raw variance explained by measures	**=**	**10.5726**	**41.3%**		**44.4%**
Raw variance explained by persons	=	4.5141	17.7%		19.0%
Raw Variance explained by items	=	6.0586	23.7%		25.5%
Raw unexplained variance (total)	=	15.0000	58.7%	100.0%	55.6%
Unexplned variance in 1st contrast	=	2.1306	8.3%	14.2%	
Unexplned variance in 2nd contrast	=	1.5982	6.2%	10.7%	
Unexplned variance in 3rd contrast	=	1.4292	5.6%	9.5%	
Unexplned variance in 4th contrast	=	1.3280	5.2%	8.9%	
Unexplned variance in 5th contrast	=	1.2499	4.9%	8.3%	

Appendix 2. Summary of Category Structure.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	247	8	-.34	-.36	1.04	1.08	NONE	(-1.59)	1
2	2	79	2	-.06	-.03	.95	1.30	.95	-.72	2
3	3	527	16	.29	.23	1.06	.87	-1.80	-.21	3
4	4	193	6	.48	.48	.95	.83	1.36	.18	4
5	5	681	21	.62	.75	1.24	.86	-.64	.67	5
6	6	1482	46	1.11	1.07	1.05	1.18	.13	(1.65)	6
MISSING		1	0	-.44						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

REVIEW

Gmail

Lukman Laliyo <lukman.laliyo019@gmail.com>

Decision on submission HELIYON-D-21-05440 to Heliyon

Heliyon <em@editorialmanager.com> Wed, Aug 18, 2021 at 11:12 AM
Reply-To: Heliyon <info@heliyon.com>
To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Manuscript. Number.: HELIYON-D-21-05440
Title: Measuring Changes in Students' Understanding Ability of the Hydrolysis Concept: Stacking and Racking Techniques in Rasch Measurement Analysis
Journal: Heliyon

Dear Dr. Laliyo,

We have now received all of the reviewers' comments on your recent submission to Heliyon.

The reviewers have advised that your manuscript should become suitable for publication in our journal after appropriate revisions.

If you are able to address the reviewers' comments, which you can find below, I would like to invite you to revise and resubmit your manuscript. We ask that you respond to each reviewer comment by either outlining how the criticism was addressed in the revised manuscript or by providing a rebuttal to the criticism. This should be carried out in a point-by-point fashion as illustrated here: https://www.cell.com/heliyon/guide-for-authors#Revisions

To allow the editors and reviewers to easily assess your revised manuscript, we also ask that you upload a version of your manuscript highlighting any revisions made. You may wish to use Microsoft Word's Track Changes tool or, for LaTeX files, the latexdiff Perl script (https://ctan.org/pkg/latexdiff).

To submit your revised manuscript, please log in as an author at https://www.editorialmanager.com/heliyon/, and navigate to the "Submissions Needing Revision" folder under the Author Main Menu. Your revision due date is Sep 16, 2021.

We understand that the COVID-19 pandemic may well be causing disruption for you and your colleagues. If that is the case for you and it has an impact on your ability to make revisions to address the concerns that came up in the review process, please reach out to us.

I look forward to receiving your revised manuscript.

Kind regards,

Theresa J. Canada, EdD
Associate Editor - Education
Heliyon

Editor and Reviewer comments:

Reviewer #1: Methods: Are quite sound, however there is a strong case here for missing many context factors (it is raised in the limitations, but they are substantial in my opinion and undermine the whole paper). The level of data analysis is okay for the limited sample size, the scale is a good. HOWEVER, for an international publication I would expect - especially when it is a case study on a very specific situation - to control and work with context variables (they are raised as a point of concern two times, both in the discussion and in the limitations) and provide deeper, more theoretically focused arguments that take those considerations into account.

The methods themselves need a stronger foundation. They should take literature of a broader educational science and more generally speaking social science-based background into consideration and use it to contextualize the information available. Otherwise it is a more or less problematic paper in my opinion.

Results: Seem good and reasonable from a analysis perspective. However, see comment above. The results are presented in a good and understandable way, however the limitations raised lead to some questions marks.

Interpretation: The section concerning the discussions are more or less results sections and less of a discussion section. Statements like "strengthen" previous studies are of relevance, but are very limited. However, the authors themselves raise a point: Qualitative (structured) Interviews would be a good way to further contextualize the results

Other comments: To be honest the paper is interesting and covers a relevant issue for a specific part of the educational / instructor community, especially as we see the need a for a good and motivating STEM-education around the globe. However, the limitations of the study are quite strict, the paper is written very technical, does not embed the work in a broader context (=RELEVANCE) and makes it much more limited than it has to be.

I would recommend a revision of the paper, that can happen in two ways:
1) If there is no way to add substantial data on context variables, the narrative structure needs to be better and deal with the literature that explains potential effects better. Theory related work needs to be much stronger if you can't add the information that is missing.
2) If you have the chance: Do some of the proposed interviews, add some qualitative data that tackles the situation heads on!

Reviewer #2: Methods:

Results:

Interpretation:

Other comments:

Reviewer #3: This manuscript has developed a tool to measure Students' Understanding Ability of the Hydrolysis Concept and compared SSI with traditional methods in terms of intervention effects. From my point of view, the organization and logic of the manuscript are clear. Besides, utilizing Rasch Model as the method is common and appropriate, and it's of great use for chemistry teachers to develop the tools for measurement.

For publication, my comments and suggestions are as follows.

Methods:

Given the different educational backgrounds in different countries, it would be useful to provide an account of the participants' previous learning foundation of chemistry.

In "Learning implementation", more teaching details need to be explained, so that readers can tell the specific differences between SSI and traditional methods and how other variables are controlled. For instance, what SSI is included in the video watched by students? What experiment have the students done? And how long does each step take?

In "Learning implementation", why these two teachers are capable of conducting teaching intervention should be further stated.

Results:

None.

Interpretation:

The results of the study indicate that "the ability changes of students in the experimental class are better than those in the control class". Can the related theories be combined to further explain the reasons in "Discussion"?

Other comments:

In "Highlights", "pre-and post test" should be revised to "pre-and post-test"

It's also necessary to check whether the abbreviations of the journal titles are standard in the "References". As far as I'm

concerned, "of" hardly appears and "Education" is usually abbreviated as "Educ". The following are some examples from the manuscript:
Inter J of Sci Edu
J of Chem Edu
Eura J of Math, Sci & Tech Edu

*****

Data in Brief (optional):
We invite you to convert your supplementary data (or a part of it) into an additional journal publication in Data in Brief, a multi-disciplinary open access journal. Data in Brief articles are a fantastic way to describe supplementary data and associated metadata, or full raw datasets deposited in an external repository, which are otherwise unnoticed. A Data in Brief article (which will be reviewed, formatted, indexed, and given a DOI) will make your data easier to find, reproduce, and cite.

You can submit to Data in Brief when you upload your revised manuscript. To do so, complete the template and follow the co-submission instructions found here: www.elsevier.com/dib-template. If your manuscript is accepted, your Data in Brief submission will automatically be transferred to Data in Brief for editorial review and publication.

Please note: an open access Article Publication Charge (APC) is payable by the author or research funder to cover the costs associated with publication in Data in Brief and ensure your data article is immediately and permanently free to access by all. For the current APC see: www.elsevier.com/journals/data-in-brief/2352-3409/open-access-journal

Please contact the Data in Brief editorial office at dib-me@elsevier.com or visit the Data in Brief homepage (www.journals.elsevier.com/data-in-brief/) if you have questions or need further information.

More information and support

FAQ: How do I revise my submission in Editorial Manager?

https://service.elsevier.com/app/answers/detail/a_id/28463/supporthub/publishing/

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

---

REVISION

Lukman Abdul Rauf Laliyo
Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo-Indonesia
Jenderal Sudirman Street No.6, Kota Tengah District, Gorontalo City 96128, Indonesia
lukman.laliyo019@gmail.com

Editor-in-Chief
Heliyon

26 January, 2022

Dear Editorial-in-Chief
I am pleased to submit an original research article entitled **"Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model"** by Lukman Abdul Rauf Laliyo, Bambang Sumintono, Citra Panigoro for consideration for publication in Heliyon.

This research aimed to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the process-oriented guided inquiry learning (POGIL) model in the context of socio-scientific issues (SSI) with the pretest-posttest control group design. Further, we would like to highlights four major points of the findings in this research:

1. Raw scores have a bias in a conventional psychometric measurement.
2. Stacking and racking techniques measure the changes in students' ability and item difficulty level.
3. The learning process in the context of socio-scientific issues improves students' understanding.
4. The changes in students' hydrolysis understanding in the pre- and post-test with negative values result from misconceptions.

We believe this finding will contribute to the development of further research related to the teaching and learning activities.

We believe that this manuscript is appropriate for publication by Heliyon because it correlated with the prior publications in this journal.

This manuscript has not been published and is not under consideration for publication elsewhere.
We also have no conflicts of interest to disclose.
Thank you for your consideration!

Sincerely,
Lukman Abdul Rauf Laliyo
Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo-Indonesia

To
Editor-in-Chief
Heliyon

January 26, 2022

Dear editor-in-chief,

Thank you for your reply regarding our manuscript HELIYON-D-21-05440 entitled **"Measuring Changes in Students' Understanding Ability of the Hydrolysis Concept: Stacking and Racking Techniques in Rasch Measurement Analysis."**

We are grateful for the reviewer's comments, and the positive evaluation of our work. We have revised and modified the text and tables according to the referees' critiques. As a consequence we provide some changes and added many new and clarifying statements in all parts of the paper. These changes have improved the manuscript considerably and we hope that it can be published without delay.

Sincerely,
Corresponding Author on Behalf of All Authors,

Lukman A. R. Laliyo

Detailed response: we have addressed and responded to the comments by the reviewer as follows.

REVIEWER 1

1. **Methods:** Are quite sound, however there is a strong case here for missing many context factors (it is raised in the limitations, but they are substantial in my opinion and undermine the whole paper). The level of data analysis is okay for the limited sample size, the scale is a good. HOWEVER, for an international publication I would expect - especially when it is a case study on a very specific situation - to control and work with context variables (they are raised as a point of concern two times, both in the discussion and in the limitations) and provide deeper, more theoretically focused arguments that take those considerations into account. The methods themselves need a stronger foundation. They should take literature of a broader educational science and more generally speaking social science-based background into consideration and use it to contextualize the information available. Otherwise it is a more or less problematic paper in my opinion.

   Authors Responses:
   Thank you for your valuable comments. We have improved the methodology, especially related to the research approach, sampling, and some additional theories related to the intervention variables. Further, the context of the research and its urgency have also been improved, according to the issues discussed.

2. **Results**: Seem good and reasonable from a analysis perspective. However, see comment above. The results are presented in a good and understandable way, however the limitations raised lead to some questions marks.

   Authors Responses:
   We thank the reviewer for giving this important comment. The research limitations have been corrected according to reviewers' criticism.

3. **Interpretation**: The section concerning the discussions are more or less results sections and less of a discussion section. Statements like "strengthen" previous studies are of relevance, but are very limited. However, the authors themselves raise a point: Qualitative (structured) Interviews would be a good way to further contextualize the results.

   Authors Responses:
   We thank the reviewer for giving this important comment. We have improved the discussion section, by adding some important points related to research findings and how to interpret changes, in terms of students' abilities. We have also outlined a description of previous studies, which are corroborated by the findings of this study. Regarding the point of qualitative (structure) interviews, we also have revised this section.

4. **Other comments:** To be honest the paper is interesting and covers a relevant issue for a specific part of the educational / instructor community, especially as we see the need a for a good and motivating STEM-education around the globe. However, the limitations of the study are quite strict, the paper is written very technical, does not embed the work in a broader context (=RELEVANCE) and makes it much more limited than it has to be.

   Authors Responses:
   We thank the reviewer for giving this important comment. The limitations of the study

have been revised and the paper has been re-written in broader context as well.

5. I would recommend a revision of the paper, that can happen in two ways: 1) If there is no way to add substantial data on context variables, the narrative structure needs to be better and deal with the literature that explains potential effects better. Theory related work needs to be much stronger if you can't add the information that is missing. 2) If you have the chance: Do some of the proposed interviews, add some qualitative data that tackles the situation heads on!

   Authors Responses:
   We thank the reviewer for giving this important comment. We have made substantial improvements to the data on context variables, and revised the narrative structure and theory used. Apart from time and cost constraints, the proposed interview was not carried out due to the tendency of the research focus to more emphasis on quantitative analysis aspects.

REVIEWER 3

1. **Methods**: Given the different educational backgrounds in different countries, it would be useful to provide an account of the participants' previous learning foundation of chemistry.

   Authors Responses:
   We thank the reviewer for giving this important comment. The methodology section has been revised by adding some additional information regarding the reviewer's comment.

2. **Methods:** In "Learning implementation", more teaching details need to be explained, so that readers can tell the specific differences between SSI and traditional methods and how other variables are controlled. For instance, what SSI is included in the video watched by students? What experiment have the students done? And how long does each step take?

   Authors Responses:
   We thank the reviewer for giving this important comment. In terms of the implementation of learning, it has been revised by adding learning steps and activities carried out at each of the steps referred to.

3. **Methods:** In "Learning implementation", why these two teachers are capable of conducting teaching intervention should be further stated.

   Authors Responses:
   We thank the reviewer for giving this important comment. We have re-collected data. We only use one teacher so that the process can be controlled and not biased.

4. **Interpretation:** The results of the study indicate that "the ability changes of students in the experimental class are better than those in the control class". Can the related theories be combined to further explain the reasons in "Discussion"?

   Authors Responses:
   We thank the reviewer for giving this important comment. This part has been revised. We describe it in the discussion and conclusion section.

5. **Other comments:** In "Highlights", "pre-and post test" should be revised to "pre-and post-test"

   Authors Responses:
   We thank the reviewer for giving this important comment. This part has been revised.

6. **Other comments:** It's also necessary to check whether the abbreviations of the journal titles are standard in the "References". As far as I'm concerned, "of" hardly appears and "Education" is usually abbreviated as "Educ". The following are some examples from the manuscript:
   Inter J of Sci Edu
   J of Chem Edu
   Eura J of Math, Sci & Tech Edu

   Authors Responses:
   We thank the reviewer for giving this important comment. This part has been revised.

Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model

Lukman Abdul Rauf Laliyo[1,*], Bambang Sumintono[2], Citra Panigoro[3]

[1] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo, 96128, Indonesia
[2] Institute of Educational Leadership, University of Malaya, Kuala Lumpur, 50603, Malaysia
[3] Department of Aquatic Resource Management, Faculty of Fisheries and Marine Science, Universitas Negeri Gorontalo, 96128, Indonesia

*Corresponding Author

Email: lukman.laliyo019@gmail.com

Authors' ORCID ID

Lukman Abdul Rauf Laliyo : https://orcid.org/0000-0003-3281-7202

Bambang Sumintono : https://orcid.org/0000-0002-5490-3665

Citra Panigoro : -

Highlights

- Raw scores have a bias in a conventional psychometric measurement
- Stacking and racking measure students' ability and item difficulty level changes
- The learning process in socio-scientific issues improves students' understanding
- Misconceptions influence the negative values of students' pre-and post-test

Abstract

This research aimed to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the process-oriented guided inquiry learning (POGIL) model in the context of socio-scientific issues (SSI) with the pretest-posttest control group design. Such techniques were based on a person- and item-centered statistic to determine how students and items changed during interventions. Eleventh-grade students in one of the top-ranked senior high schools in the eastern part of Indonesia were involved as the participants. They provided written responses (pre- and post-test) to 15 three-tier multiple-choice items. Their responses were assessed through a rubric that combines diagnostic measurement and certainty of response index. Moreover, the data were analyzed following the Rasch Partial Credit Model, using the WINSTEPS 4.5.5 software. The results suggested that students in the experimental group taught by the POGIL approach in the SSI context had better positive conceptual changes than those in the control class learning with a conventional approach. Along with the intervention effect, in certain cases, it was found that positive conceptual changes were possibly due to student guessing, which happened to be correct (lucky guess), and cheating. In other cases, students who experienced negative conceptual changes may respond incorrectly due to carelessness, the boredom of problem-solving, or misconception. Such findings have also proven that some students tend to give specific responses after the intervention in certain items, indicating that not all students fit the intervention. Besides, stacking and racking analyses are highly significant in detailing every change in students' abilities, item difficulty levels, and learning progress.

Keywords: stacking, racking, Rasch model, hydrolysis conceptual changes, inquiry model

Introduction

Central to defining the quality of pedagogical innovation in science classes is conceptual changes. The changes refer to how ideas or conceptions the students understand according to their ways of thinking [1, 2] become scientifically accurate [3]. It is because such ideas generally comprise misconceptions [4, 5, 6, 7], are not in accordance with scientific concepts [8, 9], tend to be resistant [10], changeable and varied [11], so that they should be improved if the correct conceptual understanding is to be taught [12, 13].

Some studies have been conducted on learning innovation testing to form an accurate and scientific conceptual understanding of the students, e.g., inquiry-based learning. This model presents conceptual conflicts and participatory experiments to facilitate conceptual changes [14, 15, 16]. Conceptual understanding-based learning involves various strategies in identifying and analyzing students' comprehension so that the investigation process can be designed to lead them to a more accurate and scientific conception [16, 17]. This research relied on a quasi-experimental design that assessed students' pre-test and post-test, evaluated the changes in performances for testing significant differences. This type of testing informs the researcher about the presence of an effect, but does not provide detailed information on the level and trait of the changes [18]. What if the researcher is willing to compare the extent to which the pre- and post-test change (differences in learning outcomes) and interpret the changes (the reasoning why those changes occur) in terms of content? This is a core question regarding the changes in some latent traits or changes in traits measured after the intervention. In most studies, interpreting the changes in pre-test and post-test tends to be limited to identifying whether or not an effect prevails.

Pre- and post-test changes should be given in detail regarding the students' understanding ability and item difficulty levels. However, this has not been much revealed due

to the limitations of its measurement techniques and analyses and has not been the main focus in chemistry education research to date. One reason for this issue is the debate in the psychometric community regarding the ability to measure changes accurately [18]. This debate questions the use of raw scores in the conventional psychometric analysis, which largely contains measurement biases [19], as follows: 1) the difference in pre- and post-test scores will be negatively correlated with the pre-test score, especially for students with low pre-test scores [18, 20]; 2) the difference in pre- and post-test scores shows low test reliability [21]; 3) low measurement properties due to different scales [22].

Raw scores are not final data, so that they do not have a great deal of information for drawing conclusions [23, 24]. Around the 1950s, Dr. Georg Rasch, a mathematician from Denmark, introduced the formulation of the Rasch measurement model [24]. The model has been widely applied to analyze various types of data, e.g., dichotomous, polytomous, multi-rating, and multi-rater data. In the mid-2000s, the Rasch model was used as a probabilistic-based psychometric measurement that went beyond the use of raw scores [25, 26], and was used to overcome the limitations of conventional psychometric measurement [19, 27]. Its analyses, including item fit, PCA (Principal Component Analysis), and Wright map, are commonly used for international test analyses, namely TIMSS and PISA [28].

In chemistry education research, the Rasch model has been relied on to evaluate learning understanding and progress [29], to diagnose students' preconceptions [1], misconceptions [13, 30, 31, 32], link the measurement of content knowledge with pedagogical content knowledge [33], and investigate item difficulty patterns [13, 34]. Even so, studies on the Rasch model to reveal the chemistry conceptual changes in students' understanding and item difficulty levels are relatively hard to find as of today. The present study aims to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the POGIL approach in the context of SSI and

students who learn conventionally. Such techniques are based on a person- and item-centered statistic to estimate how students and items change during the intervention.

POGIL is a student-centered learning strategy that teaches content or process skills. The philosophical foundation of POGIL is the involvement of an interactive process of careful thinking, discussing ideas, perfecting understanding, practicing skills, reflecting progress, and evaluating performances [35]. POGIL is able to lead the process of designing a participatory experiment that presents a conceptual conflict as a strategy to encourage students to form an accurate concept [14]. Therefore, POGIL intervention is more likely to be potential in driving epistemological understanding and reasoning [36], making students have opportunities to change their conceptions to be more accurate and scientific [16]. Nevertheless, it is also worth noting that some students potentially have misconceptions resistant to changes [3].

SSI functions as a learning context through the intergration of social problems that students are familiar with. It also has a conceptual connection with salt hydrolysis [37, 38], and its resolution requires many perspectives [39], including the dimension of moral and ethical evaluation of students [40]. The SSI context is a socio-scientific phenomenon that the students should explain based on their conceptual viewpoints. It encourages them to actively get involved in grasping problems [41], developing and utilizing their knowledge [42], improving their critical thinking [43], and being able to scientifically describe the discussed socio-scientific phenomenon [36, 44, 45]. For such reasons, the integration of SSI can build up students' scientific literacy [39, 46, 47]. In the end, this integration enables the learning process to be more significant in enhancing students' understanding [45, 48]. Besides, they are skilled in negotiating the social aspect of the studied phenomenon [49, 50]. For instance, the issues of global warming, climate change, and pollution [36].

Salt hydrolysis is a learning topic in high school that is strongly related to SSI. Students with a good understanding of hydrolysis will manage to clarify scientifically why

detergents, bleaching agents (NaOCl), and fertilizers can pollute the environment. Despite this, the linkage of this issue as the problem in learning hydrolysis is rarely carried out. The learning process is more emphasized on mastering theoretical concepts [36]. As a consequence, students find it challenging to use their hydrolysis understanding to explain socio-scientific phenomena around them [37]. This challenge is on account of their misconceptions regarding acid-base reaction [51], making them unable to elaborate the concept of salt hydrolysis [52] and determine acid and base strength [53]. In addition, they are struggling with correctly explaining the dissolving process and the reaction of ionic compounds with water, writing down chemical equations, and having different interpretations of the dissolving process mentioned earlier [54]. On this ground, it is essential to reveal how the hydrolysis concept changes if intervened with the POGIL approach in the SSI context, through the following specific questions: (1) is there a significant hydrolysis conceptual change of the students after the learning process in experimental and control groups? (2) if compared, how is the hydrolysis conceptual change through the intervention of POGIL in the SSI context and conventional learning? (3) in addition to intervention, is there any other factor that also contributes to the students' hydrolysis conceptual changes?

Method of Study

This study relied on a quantitative approach with a quasi-experimental and pretest-posttest control group design [55] by comparing the extent to which the hydrolysis concept changes after the intervention. Researchers carried out the learning process for 12 meetings, gave tests, and collected data on the results of the intervention and measurement.

The changes of students and items were analyzed using the stacking and racking techniques in the Rasch model [56]. As standard techniques, racking and stacking were introduced by Benjamin Wright to measure the extent to which conceptual understanding of students and items change before and after interventions [57].

In regards to students' understanding, the measurement was to identify students who had specific hydrolysis conceptual changes in responding to the learning intervention. In terms of items, the measurement was done to identify which items had special characteristics and been understood by students differently during the learning intervention [57]. Thus, the scientific inquiry approach might not be suitable for some students, or some items might be too hard after the intervention. This insightful information is immensely helpful for researchers and education practitioners, especially in evaluating the weaknesses of pedagogical innovations being applied and devising learning strategies that meet students' needs in learning [58].

Participants

Eleventh-grade students aged 16-17 years in one of the senior high schools in the eastern part of Indonesia were involved as the sample. This top-ranked school gets an "A" accreditation (excellent) from the National Accreditation Board for High School. The sample was determined by convenience sampling in six randomly assigned classes. Three classes (N=97) were experimental groups that applied the POGIL model in the SSI context. The other three classes (N=93), as control groups, applied conventional learning without the SSI context. The same teacher taught these classes following the Curriculum 2013 of Chemistry Subject (revised in 2016). There was no special classroom for learning the concept of hydrolysis, i.e., taking up the regular learning process at school. Before learning the hydrolysis concept, the students had previously learned the concept of acid and base to understand the concept of salt hydrolysis way better. Research permission was obtained from the government and school administrators. In accordance with principles of research ethics, research purpose and procedures were informed to all the students being involved and that they were voluntarily participating. Additionally, their information is confidential and only used for science development [59].

Learning implementation

Students in the experimental group studied employing the process-oriented guided inquiry learning (POGIL) in the SSI context [35]. Meanwhile, in the control class, the learning process was performed conventionally; the teacher facilitated learning initiatives. The learning process focused more on content mastery and problem-solving practice. Applying the POGIL model in the SSI context highlights teacher assistance to guide the students to prepare their conceptual understanding based on epistemological reasoning they get from experiments, discussions, and collaborations [49, 60]. Researchers carried out the learning process for eight weeks to apply the intervention to the sample, gave tests, collected data on the results of the intervention and measurement. The first three weeks were the preparation stages when researchers and the teacher shared perceptions, and asked the teacher to perform a learning simulation under the scenario, including different assistance techniques in leading the students to conduct experiments, and to ask analytical questions. The pre-test was carried out in the third week. Further, the learning implementation was done for four weeks, and the post-test was executed in the eighth week.

The learning stages with POGIL in the SSI context consist of orientation, exploration, concept formation, application, and closing. During the orientation stage, the teacher presented familiar contextual phenomena related to the concept of hydrolysis. The teacher asked initial questions to provoke curiosity and arouse motivation and interest of the students. While watching the video, had the students responded and explained the relationship between the phenomena and acids and bases, hydrolysis, and buffers. In the exploration stage, the teacher developed analytical questions with data, images, and multiple video clips to give perspectives on learning objectives and to delve into the concept that had been and would be learned. Next, the teacher assisted the students in doing experiments guided by a worksheet, and at the same time, asked analytical questions to lead them and strengthen their conceptual understanding. In the concept formation stage, the teacher asked students to build their conceptual understanding based on the exploration results, accompanied by critical and fundamental questions to guide students in building a conceptual understanding of the salt hydrolysis and buffer solution.

Following the formation stage was the application stage when the teacher presented contextual problems in the SSI context, particularly those comprising social problems in society, that closely linked with the understanding of salt hydrolysis and buffer solution concepts. Such problems included 1) the use of bleaching agents (detergents), 2) the functions of alum ($Al_2(SO_4)_3$ for water purification, 3) the harmful effects of detergent waste, 4) the beneficial and harmful effects of artificial fertilizer $(NH_4)_2SO_4$ for soil fertility, and 5) the harmful effects of monosodium glutamate (MSG) for health. In this stage, the teacher guided the students through collaborative discussions and critical questions, intending to give them perspectives on SSI phenomena and encourage them to collect information and do experiments following student activity sheets. Thereupon, the students had presentation and discussion sessions, during which they reported their experiment results and drew conclusions [61, 62]. The teacher asked them to describe the possible problems and solutions from their understanding of the studied concepts. This enabled the students to form their conceptual understanding that is closely related to contexts; the learning process was from contextual to abstract [37, 63]**.** From such a condition, the teacher led the students to apply their knowledge in different contexts and situations and solve problems. The final stage was closing or teacher assistance in guiding the students to explain the conclusion and reflection on the learning process as the end of the learning activities.

Instrument

Table 1 displays 15 items of diagnostic three-tier multiple choice test to measure students' hydrolysis conceptual understanding. The test was constructed following the Competence Standard of 2013 Chemistry Curriculum of Senior High School under Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 37 of 2018. The procedures of developing the instrument followed the recommendation by [64, 65, 66].

Table 1. Conceptual Map of Hydrolysis Concept Understanding.

Problem Context	Item	Conceptual Understanding	Ability Level	
Bleaching agents are formed of weak acid HOCl and strong base NaOH. Sodium hypochlorite salt (NaOCl) is reactive and dissolves the dye. In the water, the ion OCl^{-} will be hydrolyzed to HOCl and OH^{-} -	1	Balancing the salt (NaOCl) hydrolysis reaction in the water	2	Level 3: Students are able to calculate the pH of the hydrolyzed salt solution.
	2	Stating the partial hydrolysis reaction: $NaOCl \rightarrow Na^{+} + OCl^{-}$	2	
	3	Determining corrosive alkali of sodium hypochlorite salt (NaOCl)	1	
	4	Calculating the pH of hydrolysis of sodium hypochlorite salt (NaOCl) with NaOCl = 0.1 M; Ka = 10^{-5})	3	Level 2: Students are able to determine the hydrolysis reaction from
	5	Determining the property of NaOCl, in the reaction: $OCl^{-} + H_2O \rightarrow HOCl + OH^{-}$	2	
	6	Calculating the pH of sodium	3	

		hypochlorite salt (NaOCl) that comes from a mixture of HOCl and NaOH (partially hydrolyzed), if the Ka HOCl is 10^{-5} and there is an increase in the pH of the solution mixture.		different types of salt Level 1: Students are able to analyze the properties of the hydrolyzed salt
Water purification with alum $Al_2(SO_4)_3$ is the concept of salt hydrolysis, formed of H_2SO_4 and $Al(OH)_3$.	7	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water	1	
	8	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water that is partially hydrolyzed by the Al^{3+} ion	1	
The sodium tripolyphosphate (STPP) in detergents can pollute the environment, a eutrophication process.	9	Determining the properties of detergent solution causing eutrophication	1	
	10	Determining the properties of detergent solution (sodium tripolyphosphate salt) that is partially hydrolyzed	1	
	11	Determining the impact of the disposal of detergent waste on the environment	2	
ZA fertilizer	12	Determining the properties of	1	

$(NH4)_2SO_4$ is an acidic salt.		ammonium sulfate salt $(NH4)_2SO_4$		
	13	Stating the equation of $(NH_4)_2SO_4$ reaction in the water, partially hydrolyzed	2	
Monosodium glutamate ($C_5H_8NO_4Na$) is L-glutamic acid salt, adversely impactful on human health	14	Students' attitude towards the use of monosodium glutamate ($C_5H_8NO_4Na$)	2	
	15	Determining the properties of monosodium glutamate salt ($C_5H_8NO_4Na$)	1	

Each item was designed in three questions (Q1, Q2, Q3) that integrated diagnostic [67, 68] and summative measurements [10] and certainty of response index (CRI) [69, 70]. Students' responses to items (Q1, Q2, Q3) were evaluated based on the rubric (Table 2). For example, students' responses to items were as follows: Q1, Q2 "correct", and Q3 "very sure" under the code CCC. Such a code indicated that students' conceptual understanding was in level 6, category of Scientific Knowledge (SK). On the other hand, if the response patterns in Q1, Q2 "incorrect" and Q3 "not sure", the code would be IIU, implying that students' conceptual understanding was in the category of Lack of Knowledge (LOK), or level 1. This instrument had been validated from the aspects of item conformity with the construct variable and language. The validity results by three experts were stated under Fleiss' kappa (K = .96), meaning that the experts agreed that the item validity was categorized good.

Table 2 All Possibilities of Responses [69, 70, 71]

(Q1)	(Q2)	(Q3)	Code	Conceptual Understanding Category	Level
Correct	Correct	Certain	CCC	Scientific Knowledge (SK)	6
Correct	Incorrect	Certain	CIC	Misconception False Positive (MFP)	5
Incorrect	Correct	Certain	ICC	Misconception False Negative (MFN)	4
Incorrect	Incorrect	Certain	IIC	All-Misconception (ALM)	3
Correct	Correct	Uncertain	CCU	Lack of Confidence/Lucky Guess. (LG)	2
Correct	Incorrect	Uncertain	CIU	Lack of Knowledge (LOK)	1
Incorrect	Correct	Uncertain	ICU	Lack of Knowledge (LOK)	1
Incorrect	Incorrect	Uncertain	IIU	Lack of Knowledge (LOK)	1

Data collection and analysis

Before the intervention, this research underwent pre-test data collection; whereas, the post-test data collection was done after the intervention. The construction of pre- and post-test items was the same. Students wrote down their responses on the provided answer sheet. Both tests were supervised by teachers in the school. The students must work on all items according to the allocated time (45 minutes). The instrument was immediately collected and should have the same number as the total participants.

The pre- and post-test measurement data were still ordinal data. The Rasch Partial Credit Model with WINSTEPS 4.5.5 software [27, 72] was used to convert ordinal data into

interval data to have the same logit scale. The result was a data calibration of the levels of student's ability and item difficulty in the same interval.

The stacking analysis technique put pre-test and post-test data vertically [73]; meanwhile, the items appeared once in the experimental and control groups, allowing the researchers to check out any changes of the students after the intervention [56]. The examination was based on the same item, making the changes in students' ability during the pre- and post-test be measured [56]. Hence, each student created two measures of abilities, namely pre-test and post-test, and one measure for each item. The research hypothesis is that the students' conceptual understanding from pre-test to post-test changes, both in the experimental and control groups.

Conversely, the racking analysis technique put both pre- and post-test data horizontally, in which each item appeared twice in data collection, and students' ability only emerged once. This enabled the researchers to check out the effects of learning implementation on each student's ability from the tests, especially the changes in item difficulty levels before and after the intervention [56].

Results

Rasch analysis properties of instrument

The summary of changes in concepts and items analyzed by the Rasch model is presented in Table 1. Table 2 provides the item fit statistic. An item is considered to experience a misfit if the measurement result is not in line with the following three criteria: Outfit mean-square residual (MNSQ): $.5 < y < 1.5$; Outfit standardized mean-square residual (ZSTD): $-2 < Z < +2$; and point measure correlation (PTMEA CORR): $.4 < x < .8$ [25]. All items comply with the Outfit MNSQ criterion; item 15 does not meet the Outfit MNSQ criterion; five items (item 1, 6, 12, 13, and 15) are not in accordance with the Outfit (ZSTD)

criterion; all items meet the PTMEA CORR criterion. Simply put, all items fulfill those criteria mentioned previously (none having a misfit), and are fit and valid.

Table 3. Item Statistics: Misfit Order

Item	Difficult	Error	Outfit MNSQ	Outfit ZSTD	PTMEA CORR.
1	-.38	.05	1.36	2.87	.47
2	.20	.04	1.13	1.56	.49
3	-.36	.05	.91	-.79	.43
4	.33	.04	1.09	.77	.55
5	-.25	.05	.94	-.55	.56
6	.26	.04	1.20	2.44	.41
7	.15	.04	.91	-1.17	.54
8	.47	.04	.90	-1.45	.44
9	-.47	.05	1.19	1.49	.46
10	.08	.04	1.09	1.04	.55
11	-.34	.05	1.04	.42	.51
12	-.06	.04	.71	-3.50	.60
13	.46	.04	.74	-4.12	.55
14	-.36	.05	1.00	.77	.55
15	.26	.04	1.31	3.74	.47

This instrument has a good unidimensionality (Appendix 1). Raw variant index arrives at above the standard of 20% (33.9%), indicating that the instrument can effectively measure students' understanding of the hydrolysis concept [74]. The assessment scale analysis (Appendix 2) informs that the observation mean starts from logit -1.73 for category 1 (LOK)

to logit +1.76 (category 6, SK). This signifies that the category of students' understanding takes place consistently [27]. In addition, the high item separation index (logit 6.71) and the high item reliability (logit .98) (Table 3) indicate that the respondents (students) are sufficient to confirm the level of item difficulty, strengthening the instrument construct validity [27]. The higher the item separation and reliability index, the more confident the researchers are about replicating item placement in other suitable sample students [25, 27]. Person separation index and person reliability that reach logit 2.0 and logit .75 (Table 4), respectively, imply that the instrument is quite sensitive to differentiate the high and low abilities of the students [25, 27]. According to the Rasch model calculation, the coefficient of Cronbach Alpha of logit .81 (Table 4) reflects an interaction between 380 students and 15 items with an excellent category [24, 75]. In other words, the interaction between students and items is very significant. The instrument has an excellent internal psychometric consistency and is considered very reliable.

Table 4. Person Separation and Reliability Statistics

Parameter	Measure	SD	Separation	Reliability	INFIT		OUTFIT		KR-20
					MNSQ	ZSTD	MNSQ	ZSTD	
Person (N=380)	.67	.52	1.72	.75	1.00	.04	1.02	.10	.81
Item (N=15)	.00	.32	6.71	.98	1.07	.41	1.02	-.01	

The Difference in Students' Understanding Ability of Hydrolysis Concept

The result of the Mann-Whitney test (Table 5) brings out the fact that statistically, there is a significant difference in the results of pre-test (U=3459.000), $p<0.05$) and post-test (U=1723.000, $p<0.05$) among students in experimental and control groups. Further, the Wilcoxon test result (Table 6) shows that the results of pre-test and post-test of students in the experimental group (Z=-8.076) and the control group (Z=-6.690) at the significant level *(p)* < 0.05 are significantly different. This suggests that students' understanding of the hydrolysis concept after the intervention (post-test) is higher than before the intervention (pre-test), both in experimental and control groups. However, the abilities of students in the experimental group are better than those in the control group. Accordingly, the learning process with the POGIL in the SSI context is better than the conventional learning.

Table 5. The result of the Mann-Whitney U test based on Students' Pre-Test and Post-Test Abilities in Experimental and Control Groups (p<0.05)

Test	Experimental Group (N=97)	Control Group (N=93)	U	p
Pre-test	0.5026(-0.57-1.26)[a]	0.3029(-1.61-1.03)[a]	3459.000	0.005
Post-test	1.1722(-0.09-3.00)[a]	0.7052(-1.06-1.47)[a]	1723.000	0.000

Table 6. The result of the Wilcoxon test of Students' Pre-Test and Post-Test in Experimental and Control Groups (p<0.05)

Group	Pre-test	Post-test	Z	p*
Experimental	0.5026(-0.57-1.26)[a]	1.1722(-0.09-3.00)[a]	-8.076	0.000
Control	0.3029(-1.61-1.03)[a]	0.7052(-1.06-1.47)[a]	-6.690	0.000

The Changes in Students' Understanding Ability of the Hydrolysis Concept

From the different changes in pre- and post-test (Table 7), students in the experimental and control groups have improved their understanding of the hydrolysis concept. The experimental group's mean of pre-test and post-test is logit .51 (S.E = logit .21) and logit 1.50 (S.E = logit .32), respectively, with the mean difference of both tests is (logit .99). In contrast, the mean of pre-test and post-test of the control group gets logit .26 (S.E = logit .20) and logit .87 (S.E = logit .26), respectively, with the mean difference of pre- and post-test is logit .61. Such differences indicate different effects of interventions in the experimental and control group.

Table 7. Logit of Mean of Pre- and Post-Test Items of Experimental and Control Groups

Group	Student	Item	Mean/SE (logit)		
			Pre-test	Post-test	Pre- and Post-test Difference
Experimental	97	15	.51/(.21)	1.50/(.32)	.99
Control	93	15	.26/(.20)	.87/(.24)	.61

Description: SE = Standar Error.

If the pre-test and post-test results of the experimental group are plotted in pairs (Figure 1), so that the mean difference in the sample pre- and post-test (logit +.99) is displayed as an intercept on the horizontal axis with the plotted slope = 1, several facts obtained: First, two lines that form the upper and lower curves separate 66 students around the empirical plot line, in which the pre-test and post-test mean is not significantly different from the mean difference in the pre- and post-test in the experimental group. Second, above the curve, 23 students experience significant changes; the mean of pre- and post-test is greater than the mean difference in sample pre-test and post-test. Third, seven students do not change, and ten students have negative changes (under the curve), so that they are under the curve. Similarly, the results of pre- and post-test of the control group (Figure 2) show that 53

students are around the empirical plot line; the abilities of 25 students change significantly (greater than the mean of sample pre- and post-test (logit +.61); two students do not change; 13 students experience negative changes in abilities. The difference in the plotting of pre-test and post-test results signifies different effects of interventions in the experimental and control groups.

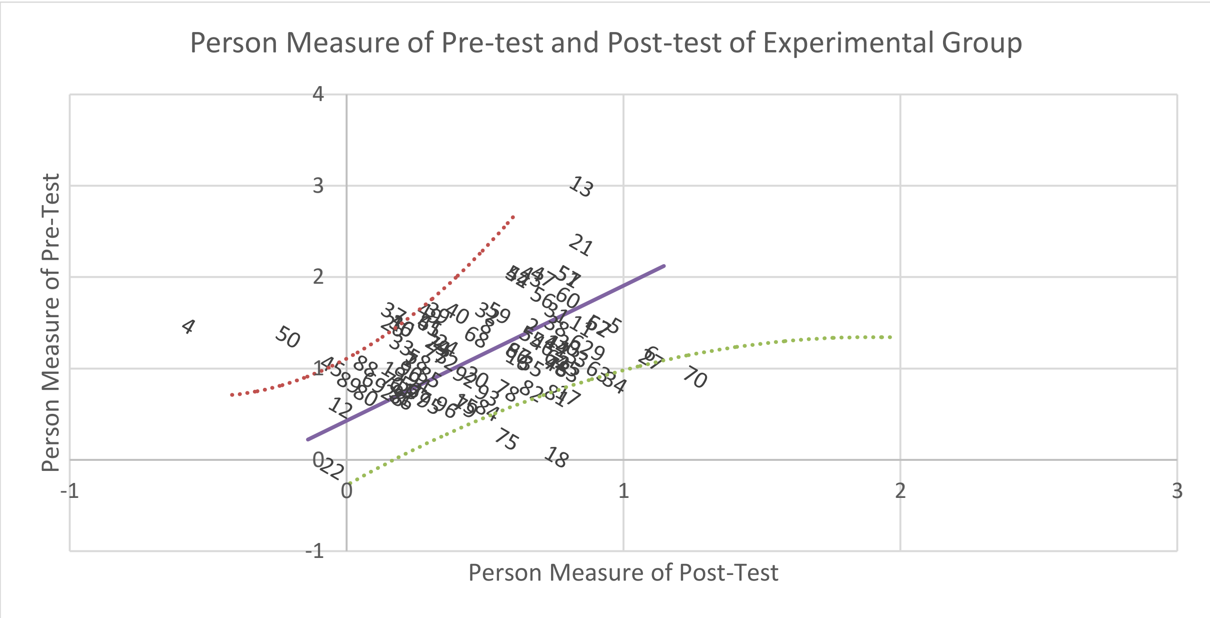

Figure 1. Scatter Plots of Person Measures in Pre- and Post-test of the Experimental Group

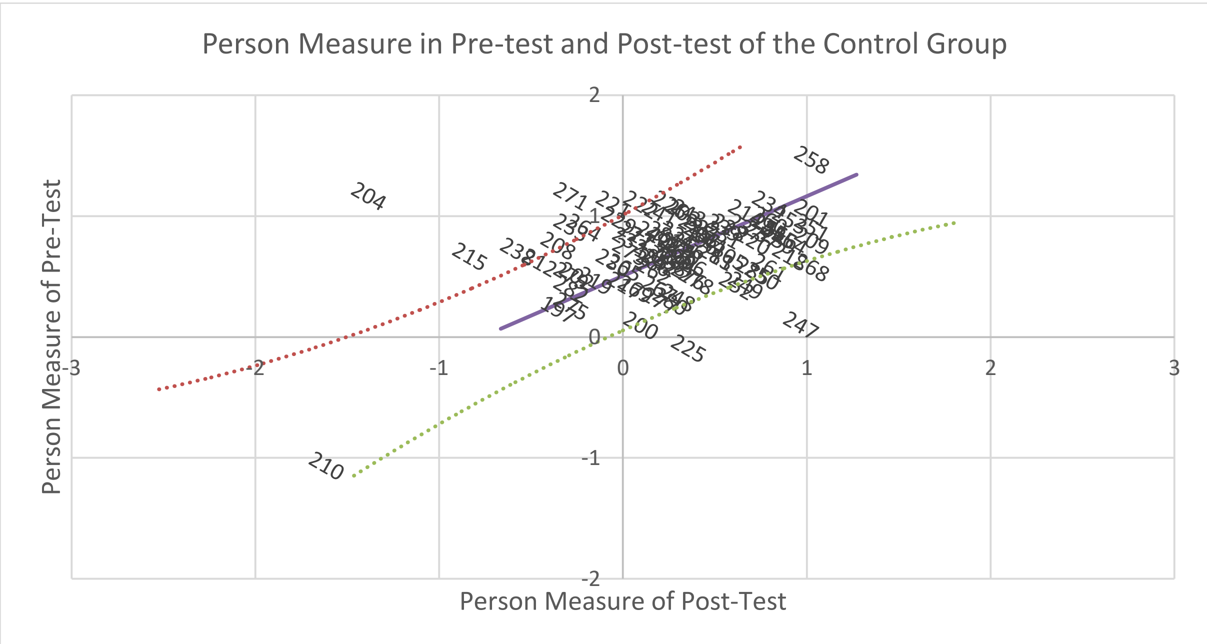

Figure 2. Scatter Plots of Person Measures in Pre- and Post-test of the Control Group

The Changes in Item Difficulty Level

Table 8 presents the results of the racking analysis in connection with the changes in item difficulty level in the pre- and post-test of experimental and control groups. It is shown that in terms of item difficulty level, the mean of pre-test of the experimental group is (logit .32), the mean of post-test is (logit -.34), and the mean difference of the pre- and post-test is (logit -.66). Moreover, the mean of pre-test of the control group is (logit .25), the mean of post-test is (logit -.25), and the mean difference of the pre- and post-test is (logit -.50). This research also finds out that seven items have significant changes in the item difficulty level in the experimental group, lower than the pre- and post-test mean difference of (logit -.66), namely item 1, 2, 5, 7, 9, and 11. Eight items with a difficulty level greater than the mean are item 3, 4, 6, 8, 12, 13, 14, and 15. Item 10 has the same difficulty level as the mean. In the control group, eight items change significantly or less than the pre- and post-test mean difference of (logit -.50), including item 2, 3, 4, 5, 9, 11, 12, and 14; five items (item 1, 6, 7, 8, 10, 13) are greater than the mean; one item (item 15) has negative changes or becomes more difficult. The most difficult item in the experimental group is item 1 (.80 logit) and the easiest one is item 14 (logit -.10). Meanwhile, the most difficult item in the control group is item 13 (logit .64), and item 3 (logit -.15) is the easiest one. These findings indicate differences in the item difficulty level changes between students taught by the POGIL in the SSI context and the conventional model.

Table 8. Data of item measures of pre- and post-test of experimental and control groups

Item	Experimental (Mean)			Control (Mean)		
	Pre-test	Post-test	Difference Pre- and Post-test	Pre-test	Post-test	Difference Pre- and Post-test

Item1	.16	-1.00	-1.16	-.06	-.76	-.7
Item2	.80	.01	-.79	.39	-.40	-.79
Item3	.20	-.63	-.43	-.15	-.83	-.68
Item4	.62	.25	-.37	.54	.02	-.52
Item5	.14	-.78	-.92	.10	-.49	-.59
Item6	.26	.22	-.04	.41	.30	-.11
Item7	.66	-.33	-.99	.33	-.06	-.39
Item8	.59	.45	-.14	.49	.47	-.02
Item9	-.04	-.85	-.81	-.08	-.93	-.85
Item10	.40	-.26	-.66	.32	-.01	-.33
Item11	.13	-.91	-1.04	.05	-.78	-.83
Item12	.33	-.23	-.56	.25	-.51	-.76
Item13	.77	.16	-.61	.64	.33	-.31
Item14	-.10	-.80	-.7	.15	-.83	-.98
Item15	.25	-.40	-.65	.39	.72	.33
Mean	.32	-.34	-.66	.25	-.25	-.50

Conceptual Changes in Students' Ability and Item Difficulty Levels

Apart from the effect of learning interventions, there are three other factors that tend to influence the changes in students' ability and item difficulty levels, as follows: 1) guessing which happened to be correct or (lucky guess), 2) cheating, 3) carelessness. These factors can be identified from the student's item response pattern using a scalogram. For instance, the response pattern of post-test item 7 for student 353, 375, and 170 (Table 9). These three students, in the seventh and eighth row from the left, cannot understand item 12 (logit -.06) and item 10 (logit .08). Meanwhile, they can correctly explain the more difficult item, i.e., item 7 (logit .15). This situation implies a lucky guess, which in fact, these students have higher post-test abilities than the item 7 logit. Next is a cheating indication in the response pattern of student 128, 129, 134, 137, and 146. Such an indication is initially detected from the same post-test mean (logit 1.61) and item response pattern. The last one is carelessness, e.g., student 110, 118, and 139 are considered to be careless as they cannot correctly explain the easy item 4 (logit .33), yet can accurately understand item 13 (logit .46), which is harder than item 4. Moreover, they get very high post-test abilities.

Table 8. Scalogram

GUTTMAN SCALE OF RESPONSES:

Person	**Item** 11 11 1 1 913415207265438	ID Person	Pre- Mean	Post-test Mean	Pre Post Difference	Item Response Pattern
353	+666555536665554	353MFCB	.8	.97	.17	Lucky Guess
375	+166566516133664	375MMCB	-.28	.40	.68	Guessing answer
170	+664666446566556	170NFEB	.33	1.17	.84	accidentaly correct
128	+666666666646555	128DFEB	.76	1.61	.85	
129	+666666666646555	129DFEB	.51	1.61	1.10	Same response pattern
134	+666666666646555	134JFEB	.17	1.61	1.44	Cheating indication
137	+666666666646555	137MMEB	.04	1.61	1.21	
146	+666666666646555	146NFEB	.30	1.61	1.31	
110	+666666666666566	110NFEB	.85	3.00	2.15	Response pattern
118	+666666666666565	118RFEB	.85	2.36	1.51	"Careless"
139	+666666665666565	139MFEB	.62	2.01	1.39	

Negative Changes

Negative changes in conceptual understanding are detected from the changes in students' post-test logit less than the pre-test logit. For example, two students from the experimental group (E18 and E75) and the control group (C225 and C247) are taken; they have negative changes (Table 9). This means that these four students experience decreased abilities after the intervention. The pre-test item mean and the post-item mean of student E18 are (logit .76) and (logit .04), sequentially, with the mean difference of pre- and post-test arriving at (logit -.72). Moreover, the pre- and post-test item standard errors of student E18 are (logit .22) and (logit .18), respectively, with the combined standard error of logit .40. On account of the higher combined standard error than the pre- and post-test measures, the ability of student E18 in both tests is not significantly different. This also applied to student E75, C225, and C247.

Table 9. Scalogram results of student E18, E75, C225, and 247

<table>
<tr><th rowspan="2">ID Person</th><th rowspan="2">Test</th><th>Item Response Pattern</th><th colspan="4">Mean</th></tr>
<tr><th>| 11 11   1 1
|913415207265438
|---------------</th><th>Item Logit</th><th>S.E* Logit</th><th>Pre- test and post-test difference</th><th>Combined S.E</th></tr>
<tr><td rowspan="2">E18</td><td>Pre-test</td><td>+665666636366333</td><td>.76</td><td>.22</td><td rowspan="2">-.72</td><td rowspan="2">.40</td></tr>
<tr><td>Post-test</td><td>+666661322521161</td><td>.04</td><td>.18</td></tr>
<tr><td rowspan="2">E75</td><td>Pre-test</td><td>+562664552566426</td><td>.58</td><td>.20</td><td rowspan="2">-.35</td><td rowspan="2">.38</td></tr>
<tr><td>Post-test</td><td>+655664322323463</td><td>.23</td><td>.18</td></tr>
<tr><td rowspan="2">C225</td><td>Pre-test</td><td>+616665663261613</td><td>.36</td><td>.19</td><td rowspan="2">-.45</td><td rowspan="2">.37</td></tr>
<tr><td>Post-test</td><td>+611622566131613</td><td>-.09</td><td>.18</td></tr>
<tr><td rowspan="2">C247</td><td>Pre-test</td><td>+663636666666435</td><td>.97</td><td>.25</td><td rowspan="2">-.87</td><td rowspan="2">.43</td></tr>
<tr><td>Post-test</td><td>+563345555314133</td><td>.10</td><td>.18</td></tr>
</table>

Description: S.E = Standar Error

Discussion and Conclusion

The findings show changes in students' understanding abilities of the hydrolysis concept and items after the intervention. From the pre- and post-test mean difference, the

experimental group has better positive changes than the control group [58]. In addition to the effect of the intervention, there is another factor contributing to the positive conceptual changes mentioned above, in terms of students' ability and item difficulty levels [24, 58]. The factor refers to some students who "accidentally" give a correct response pattern (in the post-test). Even so, both groups have also experienced negative changes, implying that the intervention is specifically responded by students on account of the carelessness factor or a misconception-comprising response pattern [56, 58, 76]. Regarding this, not all learning objectives of the hydrolysis concept match the approach of POGIL in the SSI context. Negative changes of the students are because they are not epistemologically involved in the learning process, particularly in the observing, measuring, and calculating stages. These activities are interrelated up to group discussions as part of the stages of conceptual formation based on empirical facts [77]. Students are expected to explain and link the concepts they have learned following their epistemological reasoning [16, 78].

Furthermore, the interpretation of changes due to pedagogical interventions is exemplified by four students (Table 8) in item 5. In the pre-test, the ability of student E18 (logit .76), student E75 (logit .58), student C225 (logit .36), and student C247 (logit .96) is greater. They also respond to item 5 (-.25 logit) accurately. However, in the post-test item 5, the response of student E18, E75, C225, and C247 is incorrect due to their decreased post-test abilities. Therefore, the pre- and post-test mean difference is lower than item 5. Why do these changes occur? Such changes are exemplified by the response pattern of student E18 in item 5. This item measures students' ability in determining the reaction of NaOCl reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$, with the estimated pH = 7 and is alkaline. The question (Q1) of this item is, "is it correct that NaOCl is alkaline?". E18 answers "correct" in the pre-test, yet responds to "incorrect" in the post-test. The question (Q2) of this item is "what is your consideration for your answer in the Q1?". Four options are provided: (a) because NaOCl is

formed of strong acids and weak bases; (b) because NaOCl is formed of weak acids and strong bases; (c) because NaOCl is formed of weak acids and weak bases; (d) because NaOCl is formed of strong acids and strong bases. In the pre-test, E18 chooses the correct answer (b), yet selects the incorrect answer (a) in the post-test that comprises misconception. Next, in the Q3 of this item, E18 chooses "very sure" in the pre-test and "not sure" in the post-test. The item 5 response pattern of E18 becomes CCC (category of scientific knowledge - SK) in the pre-test and IIU (category of lack of knowledge - LOK) in the post-test. Accordingly, the response pattern changes from CCC to IIU. The pre- and post-test mean difference of E18 (logit -.72) lower than item 5 (-.25) signifies that the error of response pattern results from misconception. This also applies to the response pattern of E75 (logit -.35), C225 (logit -.45), and C247 (logit -.87).

The misconception refers to the inability to identify the NaOCl salt hydrolysis that is formed of weak acids and strong bases. In short, the four students tend to not understand the concept of acid and base and acid-base reaction. These findings strengthen several previous studies [51, 53,54, 79]. A study on the understanding of the acid-base concept of senior high school students in Malaysia concludes that some students have little understanding of the function of detergents as the cleaning agent, the difference between strong acids and strong bases, and the treatment for soil acidity using fertilizers [53]. In the same tune, such little understanding is because they do not conceptualize acid-base strength as a property that arises from the interaction of many reaction factors [51]. Additionally, research on an alternative conception of salt hydrolysis among senior high school students contends that the concept of hydrolysis is challenging for the students [54]. They are usually able to state the acidity of a salt solution correctly, yet writing a chemical equation to explain such a phenomenon is a great challenge. Most of the alternative conceptions are identifiable, rooted in the misunderstanding of equilibrium process, acid and base, material structure and other basic

problems, student tendency to use a wrong analogy, and the lack of laboratory practice.

This research findings and elaboration of negative changes (case E18) prove the advantages of the Rasch model, specifically its potential in linking the result of changes (pre- and post-test), the item difficulty level, and the content being measured [18]. Such information solely comes from the Rasch model-based stacking and racking analysis techniques. The stacking technique provides information regarding "who has changed"; in contrast, the racking technique offers information of "what has changed" [56, 58], allowing the researchers to spell out the effect of the applied pedagogical innovation [18, 33, 34]. Although the instrument measurement result of this work is not data-rich, the analysis strength of the Rasch model can describe in detail the conceptual changes, both in the students' ability and item difficulty levels.

Limitations and Further Studies

The primary limitation of this research is that it did not take into account the aspects of learning style, culture, and motivation that can change due to learning interventions. Future studies, therefore, can address these aspects. The present study can be continued by considering the context of a problem that closely connects with the parameter of item difficulty level. The analysis will be more interesting if it can prove that different item difficulty levels are influenced by problem contexts in each item [80]. Further studies are also expected to find an analysis technique that can integrate problem contexts, item characteristics, and item difficulty levels in a measurement model. It is assumed that different problem contexts in each item will be more likely to affect measurement results because problem contexts have conceptual linkage with items and student activities in doing experiments, measuring, interpreting data/graphs, and others. Thus, the linkages between the learning process during the intervention and conceptual changes in students' ability and item

difficulty levels can be explained in detail; which part of the process leads the students to change their understanding related to specific ideas taught to them.

Acknowledgment

We would like to express our deep gratitude to all parties for their assistance in data processing, especially the Head of Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo.

Conflicts of interest

The authors declare no conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material

All the data are presented in the manuscript. The data supporting the findings of the article is also available in the appendix section in the end part of the manuscript.

Author Contributions

Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing: Lukman Abdul Rauf Laliyo

Validation, supervision, writing—review and editing: Bambang Sumintono

Project administration, resources, funding acquisition: Citra Panigoro

References

1. Lu S, Bi H. Development of a measurement instrument to assess students' electrolyte conceptual understanding. Chem Education Research and Practice. 2016;17(4):1030–1040. doi: 10.1039/c6rp00137h
2. Yildirir HE, Demirkol H. (2018). Identifying mental models of students for physical and chemical change. J Balt Sci Educ. 2018;17(6): 986–1004. doi: 10.33225/jbse/18.17.986
3. Gette CR, Kryjevskaia M, Stetzer MR, Heron PRL. Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy. Physical Review Physics Education Research. 2018;14(1): 10113. doi: 10.1103/PhysRevPhysEducRes.14.010113
4. Johnstone AH. Chemical education research in Glasgow in perspective. Chemical Education Research and Practice. 2006;7(2): 49–63. doi: 10.1039/b5rp90021b
5. Johnstone AH. You can't get there from here. Journal of Chemical Education. 2010;87(1): 22–29. doi: 10.1021/ed800026d
6. Taber KS. Chemical misconceptions—Prevention, diagnosis, and cure. Journal Chemistry Education. 2003; 80(5): 491.
7. Taber KS. Challenging misconceptions in the chemistry classroom: Resources to support teachers. Educació Química EduQ. 2009;4:13–20. doi: 10.2346/20.2003.02.27
8. Alamina JI, Etokeren IS. Effectiveness of imagination stretch teaching strategy in correcting misconceptions of students about particulate nature of matter. Journal of Education, Society and Behavioural Science. 2018;27(1):1–11. doi: 10.9734/jesbs/2018/43063
9. Yaşar IZ, İnce E, Kırbaşlar FG. 7. Class science and technology course "structure of atom" subject readiness improvement test. 2014;152:662-667. Procedia - Social and

Behavioral Sciences. doi: 10.1016/j.sbspro.2014.09.259

10. Hoe KY, Subramaniam R. On the prevalence of alternative conceptions on acid-base chemistry among secondary students: Insights from cognitive and confidence measures. Che Edu Res & Prac. 2016;17(2):263–282. doi: 10.1039/c5rp00146c

11. Aktan DC. Investigation of students' intermediate conceptual understanding levels: The case of direct current electricity concepts. European Journal of Physics. 2013;34(1):33–43. doi: 10.1088/0143-0807/34/1/33

12. Allen M. Misconceptions in primary science. Maidenhead: Open University Press; 2014.

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

14. Almuntasheri S, Gillies RM, Wright T. The effectiveness of a guided inquiry-based, teachers' professional development programme on saudi students' understanding of density. Science Education International. 2016:27(1); 16–39.

15. Hashweh MZ. The complexity of teaching density in middle school. Research in Science and Technological Education. 2016;34(1): 1–24. doi: 10.1080/02635143.2015.1042854

16. Zvoch K, Holveck S, Porter L. Teaching for conceptual change in a density unit provided to seventh graders: A comparison of teacher- and student-centered approaches. Research in Science Education. 2019;51:1395–1421. doi: 10.1007/s11165-019-09907-8

17. Grob R, Holmeier M, Labudde P. Formative assessment to support students' competences in inquiry-based science education. Interdisciplinary Journal of Problem-Based Learning. 2017:11(2). doi: 10.7771/1541-5015.1673

18. Pentecost TC, Barbera J. Measuring learning gains in chemical education: A

comparison of two methods. J of Chem Edu. 2013;90(7):839–845. doi: 10.1021/ed400018v

19. Sumintono B. Rasch model measurements as tools in assesment for learning. In: Proceedings of the 1st International Conference on Education Innovation (ICEI 2017); 2017 October 14; Surabaya. Paris: Atlantis Press; 2018. Available from: doi: 10.2991/icei-17.2018.11

20. Micelli R, Settanni M, Vidotto G. Measuring change in training programs : An empirical illustration Measurement of change and Item Response Theory. Psychology Science Quarterly. 2008;50(3): 433–447.

21. Willoughby SD, Metz A. Exploring gender differences with different gain calculations in astronomy and biology. American Journal of Physics. 2009;77(7):651–657. doi: 10.1119/1.3133087

22. Linn RL, Slinde JA. The determination of the significance of change between pre- and posttesting periods. Review of Educational Research. 1977;47(1):121–150. doi: 10.3102/00346543047001121

23. He P, Liu X, Zheng C, Jia M. Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. Che Edu Res & Prac. 2016;17(2):381–393. doi: 10.1039/C6RP00004E

24. Sumintono B, Widhiarso W. Aplikasi pemodelan Rasch pada assessment pendidikan [Application of Rasch modeling in educational assessment]. Bandung: Penerbit Trim Komunikata; 2015. https://www.researchgate.net/publication/282673464%0AAplikasi

25. Boone WJ, Yale MS, Staver JR. Rasch analysis in the human sciences. Dordrecht: Springer; 2014. doi: 10.1007/978-94-007-6857-4

26. Liu X. Developing measurement instruments for science education research. In: Fraser B, Tobin KG, McRobbie CJ, editors. Second International Handbook of Science

Education. Netherlands: Springer; 2012. Pp. 651-665.

27. Linacre JM. 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. US: Winsteps; 2020.

28. Sabah S, Hammouri H, Akour M. Validation of a scale of attitudes toward science across countries using Rasch model: Findings from TIMSS. J of Baltic Sci Edu. 2013;12(5):692–703.

29. Hadenfeldt JC, Bernholt S, Liu X, Neumann K, Parchmann I. Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter. J of Che Edu. 2013; 90(12):1602–1608. doi: 10.1021/ed3006192

30. Herrmann-Abell CF, DeBoer GE. Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. Che Edu Res & Pract. 2011;2(2):184–192. doi: 10.1039/c1rp90023d

31. Herrmann-Abell CF, Deboer GE. Using Rasch modeling and option probability curves to diagnose students' misconceptions. In: the 2016 AERA Annual Meeting; 2016 April 8-12; Washington DC. USA: American Educational Research Assossiation; 2016. Available from: https://www.aaas.org/sites/default/files/Project2061_CHA-AERA%20energy%20paper%204-7-16.pdf

32. Laliyo LAR, Tangio JS, Sumintono B, Jahja M, Panigoro C. Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter. J of Baltic Sci Edu. 2020;19(5). doi: 10.33225/jbse/20.19.824

33. Davidowitz B, Potgieter M. Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. Inter J of Sci Educ. 2016;38(9):1483–1503. doi:

10.1080/09500693.2016.1196843

34. Park M, Liu X. An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. Research in Science Education. 2019. doi: 10.1007/s11165-019-9819-y

35. Treagust DF, Qureshi SS, Vishnumolakala VR, Ojeil J, Mocerino M, Southam DC. Process-oriented guided inquiry learning (POGIL) as a culturally relevant pedagogy (CRP) in Qatar: A perspective from grade 10 chemistry classes. Research in Science Education. 2018:1–19. doi: 10.1007/s11165-018-9712-0

36. Kinslow AT, Sadler TD, Nguyen HT. Socio-scientific reasoning and environmental literacy in a field-based ecology class. Environ Educ Res. 2018;4622:1–23. doi: 10.1080/13504622.2018.1442418

37. Owens DC, Sadler TD, Friedrichsen P. Teaching practices for enactment of socio-scientific issues instruction: An instrumental case study of an experienced biology teacher. Res in Sci Educ. 2019. doi: 10.1007/s11165-018-9799-3

38. Bruder R, Prescott A. Research evidence on the benefits of IBL. ZDM - International J on Math Educ. 2013;45(6):811–822. doi: 10.1007/s11858-013-0542-2

39. Zeidler DL. Socioscientific issues as a curriculum emphasis: theory, research and practice. In: Lederman NG, Abell SK editor. Handbook of research on science education. New York: Routledge; 2014. pp. 697–726.

40. Espeja AG, Lagarón DC. Socio-scientific Issues (SSI) in initial training of primary school teachers: Pre-service teachers' conceptualization of SSI and appreciation of the value of teaching SSI. Procedia – Soc & Behav Scie. 2015;196:80–88. doi: 10.1016/j.sbspro.2015.07.015

41. National Research Council. A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington DC: The National Academies Press;

2012. doi: 10.17226/13165

42. Sadler TD, Zeidler DL. Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. J of Res in Scie Teach. 2009;46(8):909–921. doi: 10.1002/tea.20327

43. Lederman NG, Lederman JS, & Antink A. Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. International J of Educ in Math, Sci & and Tech. 2013;1(3):138–147. doi: 10.18404/ijemst.19784

44. Cooke AN, Fielding KS, Louis WR. Environmentally active people: The role of autonomy, relatedness, competence and self-determined motivation. Environ Educ Res. 2016;22(5):631–657. doi: 10.1080/13504622.2015.1054262

45. Sadler TD. Moral and ethical dimensions of socioscientific decision-making as integral components of scientific literacy. Sci Educ. 2004;13(1): 39–48.

46. DeBoer GE. Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. J of Res in Sci Teach. 2000;37(6):582–601. doi: 10.1002/1098-2736(200008)37:6<582::AID-TEA5>3.0.CO;2-L

47. Sadler TD. Socio-scientific issues-based education: What we know about science education in the context of SSI. In: Sadler TD, editor. Socio-Scientific Issues in the Classroom: Teaching, Learning and Research. Dordrecht: Springer; 2011. pp. 355–369. doi: 10.1007/978-94-007-1159-4_16.

48. Gräber W, Nentwig P, Becker HJ, Sumfleth E, Pitton A, Wollweber K, Jorde D. Scientific literacy: From theory to practice. Res in Sci Educ - Past, Present, & Future. 2005;1996:61–70. doi: 10.1007/0-306-47639-8_6

49. Grooms J. A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction. J of Chem

Educ. 2020. doi: 10.1021/acs.jchemed.0c00026

50. Presley ML, Sickel AJ, Muslu N, Merle D. A framework for socio-scientific issues based education. Sci Educ. 2013;22:26–32.

51. Tümay H. Emergence, learning difficulties, and misconceptions in chemistry undergraduate students’ conceptualizations of acid strength. Sci & Educ. 2016;25(1–2): 21–46. doi: 10.1007/s11191-015-9799-x

52. Seçken N. Identifying student’s misconceptions about SALT. Procedia - Soc & Behav Sci. 2010;2(2):234–245. doi: 10.1016/j.sbspro.2010.03.004

53. Damanhuri MIM., Treagust DF, Won M, Chandrasegaran AL. High school students’ understanding of acid-base concepts: An ongoing challenge for teachers. Inter J of Environ & Sci Educ. 2016;11(1):9–27. doi: 10.12973/ijese.2015.284a

54. Orwat K, Bernard P, Migdał-Mikuli A. Alternative conceptions of common salt hydrolysis among upper-secondary school students. J of Baltic Sci Educ. 2017;16(1): 64–76.

55. Creswell JW. Educational research: Planning, conducting, and evaluating quantitative and qualitative research. 4th ed. New York: Pearson Education, Inc; 2012.

56. Wright BD. Rack and stack: Time 1 vs. time 2 or pre-test vs. post-test. Rasch Mea Transac. 2003;17(1):905–906.

57. Anselmi P, Vidotto G, Bettinardi O, Bertolotti G. Measurement of change in health status with Rasch models. Heal & Qual of Life Outc. 2015;13(1):1–7. doi: 10.1186/s12955-014-0197-x

58. Ling M, Pang V, Ompok CC. Measuring change in early mathematics ability of children who learn using games: stacked analysis in rasch measurement. In: Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings, 2016 July 30-August 3; Singapore. Singapore: Springer; 2018. Available from:

https://doi.org/10.1007/978-981-10-8138-5

59. Taber KS. Ethical considerations of chemistry education research involving "human subjects." Che Educ Res & Prac. 2014;15(2):109–113. doi: 10.1039/c4rp90003k

60. Aldresti F, Rahayu S, Fajaroh F. The influence of inquiry-based chemistry learning with the contex of socio-scientific issues on high school students' scientific explanation skills. Jurnal Pendidikan IPA. 2019;23(2). doi:10.18269/jpmipa.v23i2.XXX

61. Mitarlis, Ibnu S, Rahayu S, Sutrisno. The effectiveness of new inquiry-based learning (NIBL) for improving multiple higher-order thinking skills (M-HOTS) of prospective chemistry teachers. Euro J of Educ Res. 2020;9(3):1309–1325. doi: 10.12973/eu-jer.9.3.1309

62. Pedaste M, Mäeots M, Siiman LA, de-Jong T, Van-Riesen SAN, Kamp ET, Manoli CC, Zacharia ZC, Tsourlidaki, E. Phases of inquiry-based learning: Definitions and the inquiry cycle. Educ Res Rev. 2015;14:47–61. doi: 10.1016/j.edurev.2015.02.003

63. Romine WL, Sadler TD. Measuring changes in interest in science and technology at the college level in response to two instructional interventions. Res in Sci Educ. 2016;46(3):309–327. doi: 10.1007/s11165-014-9452-8

64. Wilson M. Constructing measures: an item response modeling approach. USA: Lawrence Erlbaum Associates, Inc; 2005. doi: doi: 10.4324/9781410611697

65. Wilson M. Cognitive diagnosis using item response models. Zeits Für Psych/J of Psych. 2008;216(2):74–88. doi: 10.1027/0044-3409.216.2.74

66. Wilson M. Measuring progressions: Assessment structures underlying a learning progression. J of Res in Sci Teach. 2009;46(6):716–730. doi: 10.1002/tea.20318

67. Chittleborough G, Treagust D. Correct interpretation of chemical diagrams requires transforming from one level of representation to another. Res in Sci Educ. 2008;38(4): 463–482. doi: 10.1007/s11165-007-9059-4

68. Treagust DF. Development and use of diagnostic tests to evaluate students' misconceptions in science. Inter J of Sci Educ. 1988;10(2):159–169. doi: 10.1080/0950069880100204

69. Arslan HO, Cigdemoglu C, Moseley C. A Three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. Inter J of Sci Educ. 2012;34(11):1667–1686. doi: 10.1080/09500693.2012.680618

70. Hasan S, Bagayoko D, Kelley EL. Misconceptions and the certainty of response index (CRI). Phy Educ. 1999;34(5):294–299. doi: 10.1088/0031-9120/34/5/304

71. Habiddin H, Page EM. Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). Indo J of Chem. 2019;19(3):720–736. doi: 10.22146/ijc.39218

72. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences. 3rd ed. New York: Routledge Taylor & Francis Group; 2015. doi: 10.1088/1751-8113/44/8/085201

73. Herrmann-abell CF, Flanagan JC, Roseman JE. Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: Using student measures to evaluate the promise of the intervention. In: Proceeding of the 2013 NARST Annual International Conference. 2013 April 6-9; Puerto Rico. Puerto Rico: NARST; 2013. Available from: https://www.aaas.org/sites/default/files/project2061/files/NARST2013-Flanagan-Paper5-Usingteachermeasurestoe.pdf

74. Fisher WP. Rating scale instrument quality criteria. Rasch Measure Transac. 2007;21(1): 1095. www.rasch.org/rmt/rmt211m.htm

75. Adams WK, Wieman CE. Development and validation of instruments to measure

learning of expert-like thinking. Inter J of Sci Educ. 2011;33(9):1289–1312. doi: 10.1080/09500693.2010.512369

76. Sumintono B, Widhiarso W. Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial [Application of Rasch Model for Social Science Studies]. In Trim B, editor.). Bandung: Trim Komunikata Publishing House; 2014. https://www.researchgate.net/publication/268688933%0AAplikasi

77. Rodriguez JG, Hunter KH, Scharlott LJ, Becker NM. (2020). A Review of Research on Process Oriented Guided Inquiry Learning: Implications for Research and Practice. J. Chem. Educ. 2020;97(10):3506–3520. doi: https://doi.org/10.1021/acs.jchemed.0c00355

78. Hancock TS, Friedrichsen PJ, Kinslow AT, Sadler TD. Selecting socio-scientific issues for teaching: A grounded theory study of how science teachers collaboratively design SSI-based curricula. Sci & Educ. 2019;28(6–7):639–667. doi: 10.1007/s11191-019-00065-x

79. Ültay N, Çalik M. A comparison of different teaching designs of "acids and bases" subject. Eura Jof Math, Sci & Tech Educ. 2016;12(1):57–86. doi: 10.12973/eurasia.2016.1422a

80. Hartig J, Frey A, Nold G, Klieme E. An application of explanatory item response modeling for model-based proficiency scaling. Educ & Psychol Measurem. 2012;72(4):665–686. doi: 10.1177/0013164411430707

Appendix 1. Standardized Residual Variance in Eigen value Units.

```
     Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item
information units
                                          Eigenvalue   Observed   Expected
Total raw variance in observations     =      22.7067 100.0%         100.0%
  Raw variance explained by measures   =       7.7067  33.9%          35.9%
    Raw variance explained by persons  =       2.7733  12.2%          12.9%
    Raw Variance explained by items    =       4.9334  21.7%          23.0%
  Raw unexplained variance (total)     =      15.0000  66.1% 100.0%   64.1%
    Unexplned variance in 1st contrast =       2.0698   9.1%  13.8%
```

```
    Unexplned variance in 2nd contrast =       1.5312   6.7%  10.2%
    Unexplned variance in 3rd contrast =       1.3696   6.0%   9.1%
    Unexplned variance in 4th contrast =       1.3124   5.8%   8.7%
    Unexplned variance in 5th contrast =       1.1945   5.3%   8.0%
```

Appendix 2. Summary of Category Structure.

```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
-------------------------------------------------------------------
|CATEGORY     OBSERVED|OBSVD SAMPLE|INFIT OUTFIT|| ANDRICH |CATEGORY|
|LABEL   SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
|-------------------+------------+------------++---------+--------|
|    1   1      317   6|  -.18  -.20|  1.06  1.08||  NONE   |( -1.73)| 1
|    2   2      190   3|   .10   .09|  1.03  1.18||      .46 |    -.77 | 2
|    3   3      963  17|   .33   .31|  1.02   .93||   -1.43 |    -.22 | 3
|    4   4      542  10|   .56   .52|  1.02   .97||      .98 |     .21 | 4
|    5   5     1262  22|   .62   .73|  1.27   .99||    -.22 |     .74 | 5
|    6   6     2425  43|  1.02   .98|   .97  1.04||      .20 |(  1.76)| 6
|-------------------+------------+------------++---------+--------|
| MISSING          1   0|  -.30      |            ||         |        |
-------------------------------------------------------------------
OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.
```

REVIEW 2

Gmail

Lukman Laliyo <lukman.laliyo019@gmail.com>

Decision on submission HELIYON-D-21-05440R1 to Heliyon

Heliyon <em@editorialmanager.com> Sat, Feb 5, 2022 at 9:41 PM
Reply-To: Heliyon <info@heliyon.com>
To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Manuscript. Number.: HELIYON-D-21-05440R1
Title: Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model
Journal: Heliyon

Dear Dr. Laliyo,

We have now received all of the reviewers' comments on your recent submission to Heliyon.

The reviewers have advised that your manuscript should become suitable for publication in our journal after appropriate revisions.

If you are able to address the reviewers' comments, which you can find below, I would like to invite you to revise and resubmit your manuscript. We ask that you respond to each reviewer comment by either outlining how the criticism was addressed in the revised manuscript or by providing a rebuttal to the criticism. This should be carried out in a point-by-point fashion as illustrated here: https://www.cell.com/heliyon/guide-for-authors#Revisions

To allow the editors and reviewers to easily assess your revised manuscript, we also ask that you upload a version of your manuscript highlighting any revisions made. You may wish to use Microsoft Word's Track Changes tool or, for LaTeX files, the latexdiff Perl script (https://ctan.org/pkg/latexdiff).

To submit your revised manuscript, please log in as an author at https://www.editorialmanager.com/heliyon/, and navigate to the "Submissions Needing Revision" folder under the Author Main Menu. Your revision due date is Mar 07, 2022.

We understand that the COVID-19 pandemic may well be causing disruption for you and your colleagues. If that is the case for you and it has an impact on your ability to make revisions to address the concerns that came up in the review process, please reach out to us.

I look forward to receiving your revised manuscript.

Kind regards,

David Gonzalez-Gomez
Section Editor
Heliyon

Editor and Reviewer comments:

Reviewer 1: Methods: - See first review: Good, improvments are there

Results: - Improved very much

Interpretation: - Improved and

Other comments: It is okay to publish the paper with some very minor revisions that address the fact that this is in general a case study . There are still severe limitations, but those are adressed in a sensible way.

Reviewer 3: Methods:
Already revised in line with previous comments.

Results:
Already revised in line with previous comments.

Interpretation:
Already revised in line with previous comments.

Other comments:
Congratulations to the authors for completing a good revision. In my opinion, the manuscript is largely ready for publication. There are only a few minor errors that need further refinement.

The chemical formula for alum should be $KAl(SO_4)_2 \cdot 12H_2O$.

Some journal titles are misspelled, for example ref 4 "Chemical [CHEMISTRY] Education Research and Practice" and ref 6 "Journal [OF] Chemistry [CHEMICAL] Education".

*****

Data in Brief (optional):
We invite you to convert your supplementary data (or a part of it) into an additional journal publication in Data in Brief, a multi-disciplinary open access journal. Data in Brief articles are a fantastic way to describe supplementary data and associated metadata, or full raw datasets deposited in an external repository, which are otherwise unnoticed. A Data in Brief article (which will be reviewed, formatted, indexed, and given a DOI) will make your data easier to find, reproduce, and cite.

You can submit to Data in Brief when you upload your revised manuscript. To do so, complete the template and follow the co-submission instructions found here: www.elsevier.com/dib-template. If your manuscript is accepted, your Data in Brief submission will automatically be transferred to Data in Brief for editorial review and publication.

Please note: an open access Article Publication Charge (APC) is payable by the author or research funder to cover the costs associated with publication in Data in Brief and ensure your data article is immediately and permanently free to access by all. For the current APC see: www.elsevier.com/journals/data-in-brief/2352-3409/open-access-journal

Please contact the Data in Brief editorial office at dib-me@elsevier.com or visit the Data in Brief homepage (www.journals.elsevier.com/data-in-brief/) if you have questions or need further information.

More information and support

FAQ: How do I revise my submission in Editorial Manager?

https://service.elsevier.com/app/answers/detail/a_id/28463/supporthub/publishing/

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

time. (Use the following URL: https://www.editorialmanager.com/heliyon/login.asp?a=r). Please contact the publication office if you have any questions.

REVISION 2

To
Editor-in-Chief
Heliyon

February 14, 2022

Dear editor-in-chief,

Thank you for your reply regarding our manuscript HELIYON-D-21-05440 entitled **"**Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model.**"**

We are grateful for the reviewer's comments, and the positive evaluation of our work. We have revised and modified the text and tables according to the referees' critiques. As a consequence we provide some changes and added many new and clarifying statements in all parts of the paper. These changes have improved the manuscript considerably and we hope that it can be published without delay.

Sincerely,
Corresponding Author on Behalf of All Authors,

Lukman A. R. Laliyo

Detailed response: we have addressed and responded to the comments by the reviewer as follows.

REVIEWER 1

1. **Methods:** Methods: - See first review: Good, improvements are there

   Authors Responses:
   We are deeply grateful towards this comment.

2. **Results**: Improved very much

   Authors Responses:
   We are deeply grateful towards this comment.

3. **Interpretation**: Improved

   Authors Responses:
   We are deeply grateful towards this comment.

4. **Other comments:** It is okay to publish the paper with some very minor revisions that address the fact that this is in general a case study. There are still severe limitations, but those are adressed in a sensible way.

   Authors Responses:
   We thank the reviewer for giving this important comment. The sentences regarding this fact have been added in the methodology section. (Please kindly see the highlighted sentences)

REVIEWER 3

1. **Methods**: Already revised in line with previous comments.

   Authors Responses:
   We are deeply grateful towards this comment.

2. **Methods:** Already revised in line with previous comments.

   Authors Responses:
   We are deeply grateful towards this comment.

3. **Interpretation:** Already revised in line with previous comments.

   Authors Responses:
   We are deeply grateful towards this comment.

4. **Other comments:** Congratulations to the authors for completing a good revision. In my opinion, the manuscript is largely ready for publication. There are only a few minor errors that need further refinement.
   (1) The chemical formula for alum should be $KAl(SO_4)_2 \cdot 12H_2O$. (2) Some journal titles are misspelled, for example ref 4 "Chemical [CHEMISTRY] Education Research and

Practice" and ref 6 "Journal [OF] Chemistry [CHEMICAL] Education"

Authors Responses:
We thank the reviewer for giving this important comment. This formula has been revised and the references list has been revised according to the journal guidelines.

Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model

Lukman Abdul Rauf Laliyo[1,*], Bambang Sumintono[2], Citra Panigoro[3]

[1] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo, 96128, Indonesia
[2] Institute of Educational Leadership, University of Malaya, Kuala Lumpur, 50603, Malaysia
[3] Department of Aquatic Resource Management, Faculty of Fisheries and Marine Science, Universitas Negeri Gorontalo, 96128, Indonesia

*Corresponding Author

Email: lukman.laliyo019@gmail.com

Authors' ORCID ID

Lukman Abdul Rauf Laliyo : https://orcid.org/0000-0003-3281-7202

Bambang Sumintono : https://orcid.org/0000-0002-5490-3665

Citra Panigoro : -

Highlights

- Raw scores have a bias in a conventional psychometric measurement
- Stacking and racking measure students' ability and item difficulty level changes
- The learning process in socio-scientific issues improves students' understanding
- Misconceptions influence the negative values of students' pre-and post-test

Abstract

This research aimed to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the process-oriented guided inquiry learning (POGIL) model in the context of socio-scientific issues (SSI) with the pretest-posttest control group design. Such techniques were based on a person- and item-centered statistic to determine how students and items changed during interventions. Eleventh-grade students in one of the top-ranked senior high schools in the eastern part of Indonesia were involved as the participants. They provided written responses (pre- and post-test) to 15 three-tier multiple-choice items. Their responses were assessed through a rubric that combines diagnostic measurement and certainty of response index. Moreover, the data were analyzed following the Rasch Partial Credit Model, using the WINSTEPS 4.5.5 software. The results suggested that students in the experimental group taught by the POGIL approach in the SSI context had better positive conceptual changes than those in the control class learning with a conventional approach. Along with the intervention effect, in certain cases, it was found that positive conceptual changes were possibly due to student guessing, which happened to be correct (lucky guess), and cheating. In other cases, students who experienced negative conceptual changes may respond incorrectly due to carelessness, the boredom of problem-solving, or misconception. Such findings have also proven that some students tend to give specific responses after the intervention in certain items, indicating that not all students fit the intervention. Besides, stacking and racking analyses are highly significant in detailing every change in students' abilities, item difficulty levels, and learning progress.

Keywords: stacking, racking, Rasch model, hydrolysis conceptual changes, inquiry model

Introduction

Central to defining the quality of pedagogical innovation in science classes is conceptual changes. The changes refer to how ideas or conceptions the students understand according to their ways of thinking [1, 2] become scientifically accurate [3]. It is because such ideas generally comprise misconceptions [4, 5, 6, 7], are not in accordance with scientific concepts [8, 9], tend to be resistant [10], changeable and varied [11], so that they should be improved if the correct conceptual understanding is to be taught [12, 13].

Some studies have been conducted on learning innovation testing to form an accurate and scientific conceptual understanding of the students, e.g., inquiry-based learning. This model presents conceptual conflicts and participatory experiments to facilitate conceptual changes [14, 15, 16]. Conceptual understanding-based learning involves various strategies in identifying and analyzing students' comprehension so that the investigation process can be designed to lead them to a more accurate and scientific conception [16, 17]. This research relied on a quasi-experimental design that assessed students' pre-test and post-test, evaluated the changes in performances for testing significant differences. This type of testing informs the researcher about the presence of an effect, but does not provide detailed information on the level and trait of the changes [18]. What if the researcher is willing to compare the extent to which the pre- and post-test change (differences in learning outcomes) and interpret the changes (the reasoning why those changes occur) in terms of content? This is a core question regarding the changes in some latent traits or changes in traits measured after the intervention. In most studies, interpreting the changes in pre-test and post-test tends to be limited to identifying whether or not an effect prevails.

Pre- and post-test changes should be given in detail regarding the students' understanding ability and item difficulty levels. However, this has not been much revealed due

to the limitations of its measurement techniques and analyses and has not been the main focus in chemistry education research to date. One reason for this issue is the debate in the psychometric community regarding the ability to measure changes accurately [18]. This debate questions the use of raw scores in the conventional psychometric analysis, which largely contains measurement biases [19], as follows: 1) the difference in pre- and post-test scores will be negatively correlated with the pre-test score, especially for students with low pre-test scores [18, 20]; 2) the difference in pre- and post-test scores shows low test reliability [21]; 3) low measurement properties due to different scales [22].

Raw scores are not final data, so that they do not have a great deal of information for drawing conclusions [23, 24]. Around the 1950s, Dr. Georg Rasch, a mathematician from Denmark, introduced the formulation of the Rasch measurement model [24]. The model has been widely applied to analyze various types of data, e.g., dichotomous, polytomous, multi-rating, and multi-rater data. In the mid-2000s, the Rasch model was used as a probabilistic-based psychometric measurement that went beyond the use of raw scores [25, 26], and was used to overcome the limitations of conventional psychometric measurement [19, 27]. Its analyses, including item fit, PCA (Principal Component Analysis), and Wright map, are commonly used for international test analyses, namely TIMSS and PISA [28].

In chemistry education research, the Rasch model has been relied on to evaluate learning understanding and progress [29], to diagnose students' preconceptions [1], misconceptions [13, 30, 31, 32], link the measurement of content knowledge with pedagogical content knowledge [33], and investigate item difficulty patterns [13, 34]. Even so, studies on the Rasch model to reveal the chemistry conceptual changes in students' understanding and item difficulty levels are relatively hard to find as of today. The present study aims to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the POGIL approach in the context of SSI and

students who learn conventionally. Such techniques are based on a person- and item-centered statistic to estimate how students and items change during the intervention.

POGIL is a student-centered learning strategy that teaches content or process skills. The philosophical foundation of POGIL is the involvement of an interactive process of careful thinking, discussing ideas, perfecting understanding, practicing skills, reflecting progress, and evaluating performances [35]. POGIL is able to lead the process of designing a participatory experiment that presents a conceptual conflict as a strategy to encourage students to form an accurate concept [14]. Therefore, POGIL intervention is more likely to be potential in driving epistemological understanding and reasoning [36], making students have opportunities to change their conceptions to be more accurate and scientific [16]. Nevertheless, it is also worth noting that some students potentially have misconceptions resistant to changes [3].

SSI functions as a learning context through the intergration of social problems that students are familiar with. It also has a conceptual connection with salt hydrolysis [37, 38], and its resolution requires many perspectives [39], including the dimension of moral and ethical evaluation of students [40]. The SSI context is a socio-scientific phenomenon that the students should explain based on their conceptual viewpoints. It encourages them to actively get involved in grasping problems [41], developing and utilizing their knowledge [42], improving their critical thinking [43], and being able to scientifically describe the discussed socio-scientific phenomenon [36, 44, 45]. For such reasons, the integration of SSI can build up students' scientific literacy [39, 46, 47]. In the end, this integration enables the learning process to be more significant in enhancing students' understanding [45, 48]. Besides, they are skilled in negotiating the social aspect of the studied phenomenon [49, 50]. For instance, the issues of global warming, climate change, and pollution [36].

Salt hydrolysis is a learning topic in high school that is strongly related to SSI. Students with a good understanding of hydrolysis will manage to clarify scientifically why

detergents, bleaching agents (NaOCl), and fertilizers can pollute the environment. Despite this, the linkage of this issue as the problem in learning hydrolysis is rarely carried out. The learning process is more emphasized on mastering theoretical concepts [36]. As a consequence, students find it challenging to use their hydrolysis understanding to explain socio-scientific phenomena around them [37]. This challenge is on account of their misconceptions regarding acid-base reaction [51], making them unable to elaborate the concept of salt hydrolysis [52] and determine acid and base strength [53]. In addition, they are struggling with correctly explaining the dissolving process and the reaction of ionic compounds with water, writing down chemical equations, and having different interpretations of the dissolving process mentioned earlier [54]. On this ground, it is essential to reveal how the hydrolysis concept changes if intervened with the POGIL approach in the SSI context, through the following specific questions: (1) is there a significant hydrolysis conceptual change of the students after the learning process in experimental and control groups? (2) if compared, how is the hydrolysis conceptual change through the intervention of POGIL in the SSI context and conventional learning? (3) in addition to intervention, is there any other factor that also contributes to the students' hydrolysis conceptual changes?

Method of Study

This study relied on a quantitative approach with a quasi-experimental and pretest-posttest control group design [55] by comparing the extent to which the hydrolysis concept changes after the intervention. Researchers carried out the learning process for 12 meetings, gave tests, and collected data on the results of the intervention and measurement.

The changes of students and items were analyzed using the stacking and racking techniques in the Rasch model [56]. As standard techniques, racking and stacking were introduced by Benjamin Wright to measure the extent to which conceptual understanding of students and items change before and after interventions [57]. The referred changes are cases (item and student levels) caused by the learning intervention and can be diagnosed based on the estimated changes.

In regards to students' understanding, the measurement was to identify students who had specific hydrolysis conceptual changes in responding to the learning intervention. In terms of items, the measurement was done to identify which items had special characteristics and been understood by students differently during the learning intervention [57]. Thus, the scientific inquiry approach might not be suitable for some students, or some items might be too hard after the intervention. This insightful information is immensely helpful for researchers and education practitioners, especially in evaluating the weaknesses of pedagogical innovations being applied and devising learning strategies that meet students' needs in learning [58].

Participants

Eleventh-grade students aged 16-17 years in one of the senior high schools in the eastern part of Indonesia were involved as the sample. This top-ranked school gets an "A" accreditation (excellent) from the National Accreditation Board for High School. The sample was determined by convenience sampling in six randomly assigned classes. Three classes (N=97) were experimental groups that applied the POGIL model in the SSI context. The other three classes (N=93), as control groups, applied conventional learning without the SSI context. The same teacher taught these classes following the Curriculum 2013 of Chemistry Subject (revised in 2016). There was no special classroom for learning the concept of hydrolysis, i.e., taking up the regular learning process at school. Before learning the hydrolysis concept, the students had previously learned the concept of acid and base to understand the concept of salt hydrolysis way better. Research permission was obtained from the government and school administrators. In accordance with principles of research ethics, research purpose and procedures were informed to all the students being involved and that they were voluntarily participating. Additionally, their information is confidential and only used for science development [59].

Learning implementation

Students in the experimental group studied employing the process-oriented guided inquiry learning (POGIL) in the SSI context [35]. Meanwhile, in the control class, the learning process was performed conventionally; the teacher facilitated learning initiatives. The learning process focused more on content mastery and problem-solving practice. Applying the POGIL model in the SSI context highlights teacher assistance to guide the students to prepare their conceptual understanding based on epistemological reasoning they get from experiments, discussions, and collaborations [49, 60]. Researchers carried out the learning process for eight weeks to apply the intervention to the sample, gave tests, collected data on the results of the intervention and measurement. The first three weeks were the preparation stages when researchers and the teacher shared perceptions, and asked the teacher to perform a learning simulation under the scenario, including different assistance techniques in leading the students to conduct experiments, and to ask analytical questions. The pre-test was carried out in the third week. Further, the learning implementation was done for four weeks, and the post-test was executed in the eighth week.

The learning stages with POGIL in the SSI context consist of orientation, exploration, concept formation, application, and closing. During the orientation stage, the teacher presented familiar contextual phenomena related to the concept of hydrolysis. The teacher asked initial questions to provoke curiosity and arouse motivation and interest of the students. While watching the video, had the students responded and explained the relationship between the phenomena and acids and bases, hydrolysis, and buffers. In the exploration stage, the teacher developed analytical questions with data, images, and multiple video clips to give perspectives on learning objectives and to delve into the concept that had been and would be learned. Next, the teacher assisted the students in doing experiments guided by a worksheet, and at the same time, asked analytical questions to lead them and strengthen their conceptual understanding. In the concept formation stage, the teacher asked students to build their conceptual understanding based on the exploration results, accompanied by critical and fundamental questions to guide students in building a conceptual understanding of the salt hydrolysis and buffer solution.

Following the formation stage was the application stage when the teacher presented contextual problems in the SSI context, particularly those comprising social problems in society, that closely linked with the understanding of salt hydrolysis and buffer solution concepts. Such problems included 1) the use of bleaching agents (detergents), 2) the functions of alum $KAl(SO_4)_2 \cdot 12H_2O$ for water purification, 3) the harmful effects of detergent waste, 4) the beneficial and harmful effects of artificial fertilizer $(NH_4)_2SO_4$ for soil fertility, and 5) the harmful effects of monosodium glutamate (MSG) for health. In this stage, the teacher guided the students through collaborative discussions and critical questions, intending to give them perspectives on SSI phenomena and encourage them to collect information and do experiments following student activity sheets. Thereupon, the students had presentation and discussion sessions, during which they reported their experiment results and drew conclusions [61, 62]. The teacher asked them to describe the possible problems and solutions from their understanding of the studied concepts. This enabled the students to form their conceptual understanding that is closely related to contexts; the learning process was from contextual to abstract [37, 63]**.** From such a condition, the teacher led the students to apply their knowledge in different contexts and situations and solve problems. The final stage was closing or teacher assistance in guiding the students to explain the conclusion and reflection on the learning process as the end of the learning activities.

Instrument

Table 1 displays 15 items of diagnostic three-tier multiple choice test to measure students' hydrolysis conceptual understanding. The test was constructed following the Competence Standard of 2013 Chemistry Curriculum of Senior High School under Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 37 of 2018. The procedures of developing the instrument followed the recommendation by [64, 65, 66].

Table 1. Conceptual Map of Hydrolysis Concept Understanding.

Problem Context	Item	Conceptual Understanding	Ability Level	
Bleaching agents are formed of weak acid HOCl and strong base NaOH. Sodium hypochlorite salt (NaOCl) is reactive and dissolves the dye. In the water, the ion OCl^{-} will be hydrolyzed to HOCl and OH^{-} -	1	Balancing the salt (NaOCl) hydrolysis reaction in the water	2	Level 3: Students are able to calculate the pH of the hydrolyzed salt solution.
	2	Stating the partial hydrolysis reaction: $NaOCl \rightarrow Na^{+} + OCl^{-}$	2	
	3	Determining corrosive alkali of sodium hypochlorite salt (NaOCl)	1	
	4	Calculating the pH of hydrolysis of sodium hypochlorite salt (NaOCl) with NaOCl = 0.1 M; Ka = 10^{-5})	3	Level 2: Students are able to determine the hydrolysis
	5	Determining the property of NaOCl, in the reaction: $OCl^{-} + H_2O \rightarrow HOCl + OH^{-}$	2	

	6	Calculating the pH of sodium hypochlorite salt (NaOCl) that comes from a mixture of HOCl and NaOH (partially hydrolyzed), if the Ka HOCl is 10^{-5} and there is an increase in the pH of the solution mixture.	3	reaction from different types of salt Level 1: Students are able to analyze the properties of the hydrolyzed salt
Water purification with alum $KAl(SO_4)_2 \cdot 12H_2O$ is the concept of salt hydrolysis, formed of H_2SO_4 and $Al(OH)_3$.	7	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water	1	
	8	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water that is partially hydrolyzed by the Al^{3+} ion	1	
The sodium tripolyphosphate (STPP) in detergents can pollute the environment, a eutrophication process.	9	Determining the properties of detergent solution causing eutrophication	1	
	10	Determining the properties of detergent solution (sodium tripolyphosphate salt) that is partially hydrolyzed	1	
	11	Determining the impact of the disposal of detergent waste on the environment	2	

ZA fertilizer $(NH4)_2SO_4$ is an acidic salt.	12	Determining the properties of ammonium sulfate salt $(NH4)_2SO_4$	1	
	13	Stating the equation of $(NH_4)_2SO_4$ reaction in the water, partially hydrolyzed	2	
Monosodium glutamate ($C_5H_8NO_4Na$) is L-glutamic acid salt, adversely impactful on human health	14	Students' attitude towards the use of monosodium glutamate ($C_5H_8NO_4Na$)	2	
	15	Determining the properties of monosodium glutamate salt ($C_5H_8NO_4Na$)	1	

Each item was designed in three questions (Q1, Q2, Q3) that integrated diagnostic [67, 68] and summative measurements [10] and certainty of response index (CRI) [69, 70]. Students' responses to items (Q1, Q2, Q3) were evaluated based on the rubric (Table 2). For example, students' responses to items were as follows: Q1, Q2 "correct", and Q3 "very sure" under the code CCC. Such a code indicated that students' conceptual understanding was in level 6, category of Scientific Knowledge (SK). On the other hand, if the response patterns in Q1, Q2 "incorrect" and Q3 "not sure", the code would be IIU, implying that students' conceptual understanding was in the category of Lack of Knowledge (LOK), or level 1. This instrument had been validated from the aspects of item conformity with the construct variable and language. The validity results by three experts were stated under Fleiss' kappa (K = .96), meaning that the experts agreed that the item validity was categorized good.

Table 2 All Possibilities of Responses [69, 70, 71]

(Q1)	(Q2)	(Q3)	Code	Conceptual Understanding Category	Level
Correct	Correct	Certain	CCC	Scientific Knowledge (SK)	6
Correct	Incorrect	Certain	CIC	Misconception False Positive (MFP)	5
Incorrect	Correct	Certain	ICC	Misconception False Negative (MFN)	4
Incorrect	Incorrect	Certain	IIC	All-Misconception (ALM)	3
Correct	Correct	Uncertain	CCU	Lack of Confidence/Lucky Guess. (LG)	2
Correct	Incorrect	Uncertain	CIU	Lack of Knowledge (LOK)	1
Incorrect	Correct	Uncertain	ICU	Lack of Knowledge (LOK)	1
Incorrect	Incorrect	Uncertain	IIU	Lack of Knowledge (LOK)	1

Data collection and analysis

Before the intervention, this research underwent pre-test data collection; whereas, the post-test data collection was done after the intervention. The construction of pre- and post-test items was the same. Students wrote down their responses on the provided answer sheet. Both tests were supervised by teachers in the school. The students must work on all items according to the allocated time (45 minutes). The instrument was immediately collected and should have the same number as the total participants.

The pre- and post-test measurement data were still ordinal data. The Rasch Partial Credit Model with WINSTEPS 4.5.5 software [27, 72] was used to convert ordinal data into

interval data to have the same logit scale. The result was a data calibration of the levels of student's ability and item difficulty in the same interval.

The stacking analysis technique put pre-test and post-test data vertically [73]; meanwhile, the items appeared once in the experimental and control groups, allowing the researchers to check out any changes of the students after the intervention [56]. The examination was based on the same item, making the changes in students’ ability during the pre- and post-test be measured [56]. Hence, each student created two measures of abilities, namely pre-test and post-test, and one measure for each item. The research hypothesis is that the students' conceptual understanding from pre-test to post-test changes, both in the experimental and control groups.

Conversely, the racking analysis technique put both pre- and post-test data horizontally, in which each item appeared twice in data collection, and students’ ability only emerged once. This enabled the researchers to check out the effects of learning implementation on each student’s ability from the tests, especially the changes in item difficulty levels before and after the intervention [56].

Results

Rasch analysis properties of instrument

The summary of changes in concepts and items analyzed by the Rasch model is presented in Table 1. Table 2 provides the item fit statistic. An item is considered to experience a misfit if the measurement result is not in line with the following three criteria: Outfit mean-square residual (MNSQ): $.5 < y < 1.5$; Outfit standardized mean-square residual (ZSTD): $-2 < Z < +2$; and point measure correlation (PTMEA CORR): $.4 < x < .8$ [25]. All items comply with the Outfit MNSQ criterion; item 15 does not meet the Outfit MNSQ criterion; five items (item 1, 6, 12, 13, and 15) are not in accordance with the Outfit (ZSTD)

criterion; all items meet the PTMEA CORR criterion. Simply put, all items fulfill those criteria mentioned previously (none having a misfit), and are fit and valid.

Table 3. Item Statistics: Misfit Order

Item	Difficult	Error	Outfit MNSQ	Outfit ZSTD	PTMEA CORR.
1	-.38	.05	1.36	2.87	.47
2	.20	.04	1.13	1.56	.49
3	-.36	.05	.91	-.79	.43
4	.33	.04	1.09	.77	.55
5	-.25	.05	.94	-.55	.56
6	.26	.04	1.20	2.44	.41
7	.15	.04	.91	-1.17	.54
8	.47	.04	.90	-1.45	.44
9	-.47	.05	1.19	1.49	.46
10	.08	.04	1.09	1.04	.55
11	-.34	.05	1.04	.42	.51
12	-.06	.04	.71	-3.50	.60
13	.46	.04	.74	-4.12	.55
14	-.36	.05	1.00	.77	.55
15	.26	.04	1.31	3.74	.47

This instrument has a good unidimensionality (Appendix 1). Raw variant index arrives at above the standard of 20% (33.9%), indicating that the instrument can effectively measure students' understanding of the hydrolysis concept [74]. The assessment scale analysis (Appendix 2) informs that the observation mean starts from logit -1.73 for category 1 (LOK)

to logit +1.76 (category 6, SK). This signifies that the category of students' understanding takes place consistently [27]. In addition, the high item separation index (logit 6.71) and the high item reliability (logit .98) (Table 3) indicate that the respondents (students) are sufficient to confirm the level of item difficulty, strengthening the instrument construct validity [27]. The higher the item separation and reliability index, the more confident the researchers are about replicating item placement in other suitable sample students [25, 27]. Person separation index and person reliability that reach logit 2.0 and logit .75 (Table 4), respectively, imply that the instrument is quite sensitive to differentiate the high and low abilities of the students [25, 27]. According to the Rasch model calculation, the coefficient of Cronbach Alpha of logit .81 (Table 4) reflects an interaction between 380 students and 15 items with an excellent category [24, 75]. In other words, the interaction between students and items is very significant. The instrument has an excellent internal psychometric consistency and is considered very reliable.

Table 4. Person Separation and Reliability Statistics

Parameter	Measure	SD	Separation	Reliability	INFIT		OUTFIT		KR-20
					MNSQ	ZSTD	MNSQ	ZSTD	
Person (N=380)	.67	.52	1.72	.75	1.00	.04	1.02	.10	.81
Item (N=15)	.00	.32	6.71	.98	1.07	.41	1.02	-.01	

The Difference in Students' Understanding Ability of Hydrolysis Concept

The result of the Mann-Whitney test (Table 5) brings out the fact that statistically, there is a significant difference in the results of pre-test (U=3459.000), $p<0.05$) and post-test (U=1723.000, $p<0.05$) among students in experimental and control groups. Further, the Wilcoxon test result (Table 6) shows that the results of pre-test and post-test of students in the experimental group (Z=-8.076) and the control group (Z=-6.690) at the significant level *(p)* < 0.05 are significantly different. This suggests that students' understanding of the hydrolysis concept after the intervention (post-test) is higher than before the intervention (pre-test), both in experimental and control groups. However, the abilities of students in the experimental group are better than those in the control group. Accordingly, the learning process with the POGIL in the SSI context is better than the conventional learning.

Table 5. The result of the Mann-Whitney U test based on Students' Pre-Test and Post-Test Abilities in Experimental and Control Groups (p<0.05)

Test	Experimental Group (N=97)	Control Group (N=93)	U	p
Pre-test	0.5026(-0.57-1.26)[a]	0.3029(-1.61-1.03)[a]	3459.000	0.005
Post-test	1.1722(-0.09-3.00)[a]	0.7052(-1.06-1.47)[a]	1723.000	0.000

Table 6. The result of the Wilcoxon test of Students' Pre-Test and Post-Test in Experimental and Control Groups (p<0.05)

Group	Pre-test	Post-test	Z	p*
Experimental	0.5026(-0.57-1.26)[a]	1.1722(-0.09-3.00)[a]	-8.076	0.000
Control	0.3029(-1.61-1.03)[a]	0.7052(-1.06-1.47)[a]	-6.690	0.000

The Changes in Students' Understanding Ability of the Hydrolysis Concept

From the different changes in pre- and post-test (Table 7), students in the experimental and control groups have improved their understanding of the hydrolysis concept. The experimental group's mean of pre-test and post-test is logit .51 (S.E = logit .21) and logit 1.50 (S.E = logit .32), respectively, with the mean difference of both tests is (logit .99). In contrast, the mean of pre-test and post-test of the control group gets logit .26 (S.E = logit .20) and logit .87 (S.E = logit .26), respectively, with the mean difference of pre- and post-test is logit .61. Such differences indicate different effects of interventions in the experimental and control group.

Table 7. Logit of Mean of Pre- and Post-Test Items of Experimental and Control Groups

Group	Student	Item	Mean/SE (logit)		
			Pre-test	Post-test	Pre- and Post-test Difference
Experimental	97	15	.51/(.21)	1.50/(.32)	.99
Control	93	15	.26/(.20)	.87/(.24)	.61

Description: SE = Standar Error.

If the pre-test and post-test results of the experimental group are plotted in pairs (Figure 1), so that the mean difference in the sample pre- and post-test (logit +.99) is displayed as an intercept on the horizontal axis with the plotted slope = 1, several facts obtained: First, two lines that form the upper and lower curves separate 66 students around the empirical plot line, in which the pre-test and post-test mean is not significantly different from the mean difference in the pre- and post-test in the experimental group. Second, above the curve, 23 students experience significant changes; the mean of pre- and post-test is greater than the mean difference in sample pre-test and post-test. Third, seven students do not change, and ten students have negative changes (under the curve), so that they are under the curve. Similarly, the results of pre- and post-test of the control group (Figure 2) show that 53

students are around the empirical plot line; the abilities of 25 students change significantly (greater than the mean of sample pre- and post-test (logit +.61); two students do not change; 13 students experience negative changes in abilities. The difference in the plotting of pre-test and post-test results signifies different effects of interventions in the experimental and control groups.

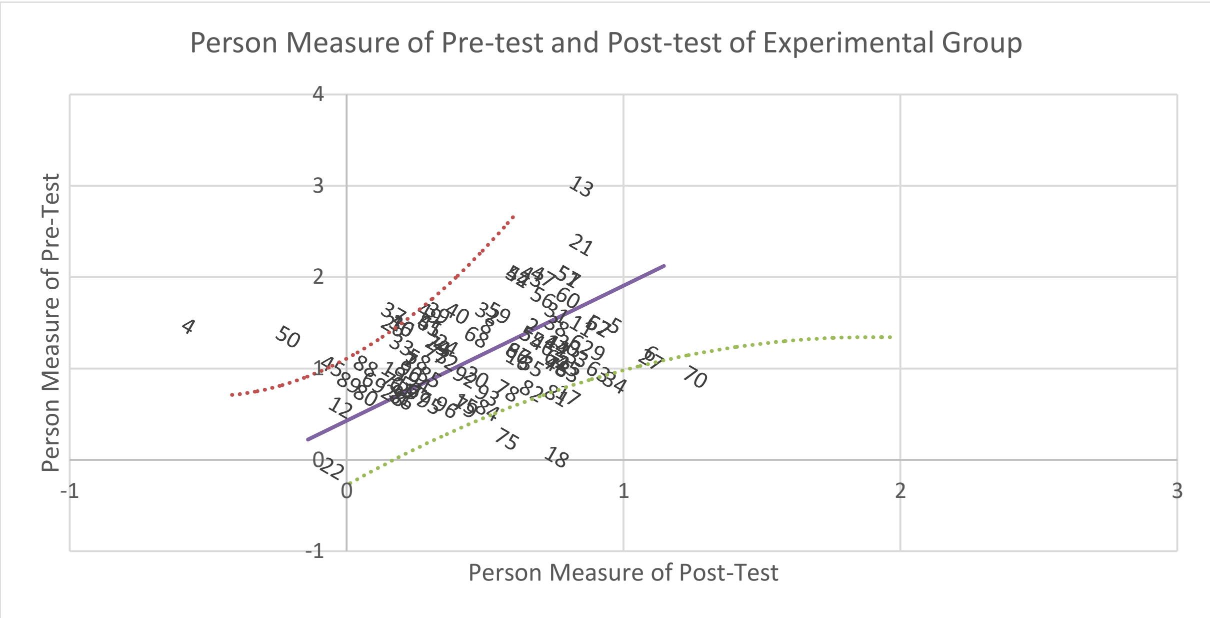

Figure 1. Scatter Plots of Person Measures in Pre- and Post-test of the Experimental Group

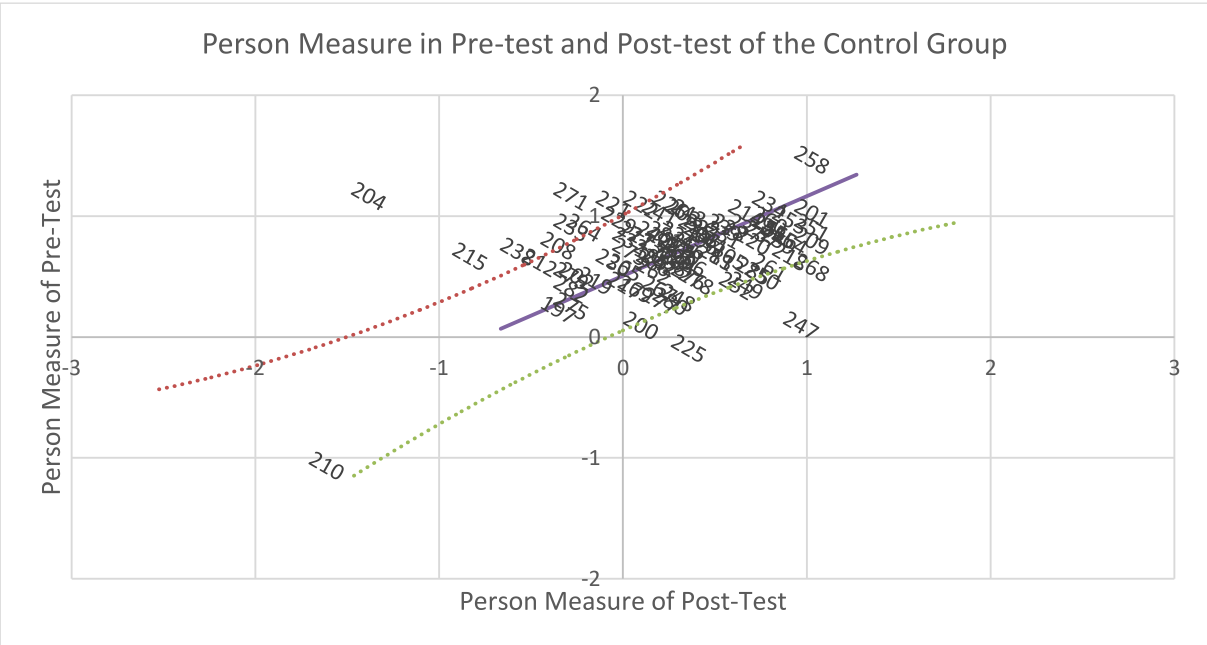

Figure 2. Scatter Plots of Person Measures in Pre- and Post-test of the Control Group

The Changes in Item Difficulty Level

Table 8 presents the results of the racking analysis in connection with the changes in item difficulty level in the pre- and post-test of experimental and control groups. It is shown that in terms of item difficulty level, the mean of pre-test of the experimental group is (logit .32), the mean of post-test is (logit -.34), and the mean difference of the pre- and post-test is (logit -.66). Moreover, the mean of pre-test of the control group is (logit .25), the mean of post-test is (logit -.25), and the mean difference of the pre- and post-test is (logit -.50). This research also finds out that seven items have significant changes in the item difficulty level in the experimental group, lower than the pre- and post-test mean difference of (logit -.66), namely item 1, 2, 5, 7, 9, and 11. Eight items with a difficulty level greater than the mean are item 3, 4, 6, 8, 12, 13, 14, and 15. Item 10 has the same difficulty level as the mean. In the control group, eight items change significantly or less than the pre- and post-test mean difference of (logit -.50), including item 2, 3, 4, 5, 9, 11, 12, and 14; five items (item 1, 6, 7, 8, 10, 13) are greater than the mean; one item (item 15) has negative changes or becomes more difficult. The most difficult item in the experimental group is item 1 (.80 logit) and the easiest one is item 14 (logit -.10). Meanwhile, the most difficult item in the control group is item 13 (logit .64), and item 3 (logit -.15) is the easiest one. These findings indicate differences in the item difficulty level changes between students taught by the POGIL in the SSI context and the conventional model.

Table 8. Data of item measures of pre- and post-test of experimental and control groups

Item	Experimental (Mean)			Control (Mean)		
	Pre-test	Post-test	Difference Pre- and Post-test	Pre-test	Post-test	Difference Pre- and Post-test

Item1	.16	-1.00	-1.16	-.06	-.76	-.7
Item2	.80	.01	-.79	.39	-.40	-.79
Item3	.20	-.63	-.43	-.15	-.83	-.68
Item4	.62	.25	-.37	.54	.02	-.52
Item5	.14	-.78	-.92	.10	-.49	-.59
Item6	.26	.22	-.04	.41	.30	-.11
Item7	.66	-.33	-.99	.33	-.06	-.39
Item8	.59	.45	-.14	.49	.47	-.02
Item9	-.04	-.85	-.81	-.08	-.93	-.85
Item10	.40	-.26	-.66	.32	-.01	-.33
Item11	.13	-.91	-1.04	.05	-.78	-.83
Item12	.33	-.23	-.56	.25	-.51	-.76
Item13	.77	.16	-.61	.64	.33	-.31
Item14	-.10	-.80	-.7	.15	-.83	-.98
Item15	.25	-.40	-.65	.39	.72	.33
Mean	.32	-.34	-.66	.25	-.25	-.50

Conceptual Changes in Students' Ability and Item Difficulty Levels

Apart from the effect of learning interventions, there are three other factors that tend to influence the changes in students' ability and item difficulty levels, as follows: 1) guessing which happened to be correct or (lucky guess), 2) cheating, 3) carelessness. These factors can be identified from the student's item response pattern using a scalogram. For instance, the response pattern of post-test item 7 for student 353, 375, and 170 (Table 9). These three students, in the seventh and eighth row from the left, cannot understand item 12 (logit -.06) and item 10 (logit .08). Meanwhile, they can correctly explain the more difficult item, i.e., item 7 (logit .15). This situation implies a lucky guess, which in fact, these students have higher post-test abilities than the item 7 logit. Next is a cheating indication in the response pattern of student 128, 129, 134, 137, and 146. Such an indication is initially detected from the same post-test mean (logit 1.61) and item response pattern. The last one is carelessness, e.g., student 110, 118, and 139 are considered to be careless as they cannot correctly explain the easy item 4 (logit .33), yet can accurately understand item 13 (logit .46), which is harder than item 4. Moreover, they get very high post-test abilities.

Table 8. Scalogram

```
GUTTMAN SCALE OF RESPONSES:
Person |Item
       |   11 11   1 1   ID        Pre-  Post-test  Pre Post    Item
       |913415207265438  Person    Mean     Mean   Difference   Response Pattern
       |---------------
   353 +666555536665554  353MFCB    .8     .97      .17 ┐   Lucky Guess
   375 +166566516133664  375MMCB  -.28     .40      .68 ├─  Guessing answer
   170 +664666446566556  170NFEB   .33    1.17      .84 ┘   accidentaly correct

   128 +666666666646555  128DFEB   .76    1.61      .85 ┐
   129 +666666666646555  129DFEB   .51    1.61     1.10 │   Same response pattern
   134 +666666666646555  134JFEB   .17    1.61     1.44 ├─  Cheating indication
   137 +666666666646555  137MMEB   .04    1.61     1.21 │
   146 +666666666646555  146NFEB   .30    1.61     1.31 ┘

   110 +666666666666566  110NFEB   .85    3.00     2.15 ┐   Response pattern
   118 +666666666666565  118RFEB   .85    2.36     1.51 ├─  "Careless"
   139 +666666665666565  139MFEB   .62    2.01     1.39 ┘
```

Negative Changes

Negative changes in conceptual understanding are detected from the changes in students' post-test logit less than the pre-test logit. For example, two students from the experimental group (E18 and E75) and the control group (C225 and C247) are taken; they have negative changes (Table 9). This means that these four students experience decreased abilities after the intervention. The pre-test item mean and the post-item mean of student E18 are (logit .76) and (logit .04), sequentially, with the mean difference of pre- and post-test arriving at (logit -.72). Moreover, the pre- and post-test item standard errors of student E18 are (logit .22) and (logit .18), respectively, with the combined standard error of logit .40. On account of the higher combined standard error than the pre- and post-test measures, the ability of student E18 in both tests is not significantly different. This also applied to student E75, C225, and C247.

Table 9. Scalogram results of student E18, E75, C225, and 247

ID Person	Test	Item Response Pattern	Mean			
		| 11 11 1 1 |913415207265438 |---------------	Item Logit	S.E* Logit	Pre- test and post-test difference	Combined S.E
E18	Pre-test	+665666636366333	.76	.22	-.72	.40
	Post-test	+666661322521161	.04	.18		
E75	Pre-test	+562664552566426	.58	.20	-.35	.38
	Post-test	+655664322323463	.23	.18		
C225	Pre-test	+616665663261613	.36	.19	-.45	.37
	Post-test	+611622566131613	-.09	.18		
C247	Pre-test	+663636666666435	.97	.25	-.87	.43
	Post-test	+563345555314133	.10	.18		

Description: S.E = Standar Error

Discussion and Conclusion

The findings show changes in students' understanding abilities of the hydrolysis concept and items after the intervention. From the pre- and post-test mean difference, the

experimental group has better positive changes than the control group [58]. In addition to the effect of the intervention, there is another factor contributing to the positive conceptual changes mentioned above, in terms of students' ability and item difficulty levels [24, 58]. The factor refers to some students who "accidentally" give a correct response pattern (in the post-test). Even so, both groups have also experienced negative changes, implying that the intervention is specifically responded by students on account of the carelessness factor or a misconception-comprising response pattern [56, 58, 76]. Regarding this, not all learning objectives of the hydrolysis concept match the approach of POGIL in the SSI context. Negative changes of the students are because they are not epistemologically involved in the learning process, particularly in the observing, measuring, and calculating stages. These activities are interrelated up to group discussions as part of the stages of conceptual formation based on empirical facts [77]. Students are expected to explain and link the concepts they have learned following their epistemological reasoning [16, 78].

Furthermore, the interpretation of changes due to pedagogical interventions is exemplified by four students (Table 8) in item 5. In the pre-test, the ability of student E18 (logit .76), student E75 (logit .58), student C225 (logit .36), and student C247 (logit .96) is greater. They also respond to item 5 (-.25 logit) accurately. However, in the post-test item 5, the response of student E18, E75, C225, and C247 is incorrect due to their decreased post-test abilities. Therefore, the pre- and post-test mean difference is lower than item 5. Why do these changes occur? Such changes are exemplified by the response pattern of student E18 in item 5. This item measures students' ability in determining the reaction of NaOCl reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$, with the estimated pH = 7 and is alkaline. The question (Q1) of this item is, "is it correct that NaOCl is alkaline?". E18 answers "correct" in the pre-test, yet responds to "incorrect" in the post-test. The question (Q2) of this item is "what is your consideration for your answer in the Q1?". Four options are provided: (a) because NaOCl is

formed of strong acids and weak bases; (b) because NaOCl is formed of weak acids and strong bases; (c) because NaOCl is formed of weak acids and weak bases; (d) because NaOCl is formed of strong acids and strong bases. In the pre-test, E18 chooses the correct answer (b), yet selects the incorrect answer (a) in the post-test that comprises misconception. Next, in the Q3 of this item, E18 chooses "very sure" in the pre-test and "not sure" in the post-test. The item 5 response pattern of E18 becomes CCC (category of scientific knowledge - SK) in the pre-test and IIU (category of lack of knowledge - LOK) in the post-test. Accordingly, the response pattern changes from CCC to IIU. The pre- and post-test mean difference of E18 (logit -.72) lower than item 5 (-.25) signifies that the error of response pattern results from misconception. This also applies to the response pattern of E75 (logit -.35), C225 (logit -.45), and C247 (logit -.87).

The misconception refers to the inability to identify the NaOCl salt hydrolysis that is formed of weak acids and strong bases. In short, the four students tend to not understand the concept of acid and base and acid-base reaction. These findings strengthen several previous studies [51, 53,54, 79]. A study on the understanding of the acid-base concept of senior high school students in Malaysia concludes that some students have little understanding of the function of detergents as the cleaning agent, the difference between strong acids and strong bases, and the treatment for soil acidity using fertilizers [53]. In the same tune, such little understanding is because they do not conceptualize acid-base strength as a property that arises from the interaction of many reaction factors [51]. Additionally, research on an alternative conception of salt hydrolysis among senior high school students contends that the concept of hydrolysis is challenging for the students [54]. They are usually able to state the acidity of a salt solution correctly, yet writing a chemical equation to explain such a phenomenon is a great challenge. Most of the alternative conceptions are identifiable, rooted in the misunderstanding of equilibrium process, acid and base, material structure and other basic

problems, student tendency to use a wrong analogy, and the lack of laboratory practice.

This research findings and elaboration of negative changes (case E18) prove the advantages of the Rasch model, specifically its potential in linking the result of changes (pre- and post-test), the item difficulty level, and the content being measured [18]. Such information solely comes from the Rasch model-based stacking and racking analysis techniques. The stacking technique provides information regarding "who has changed"; in contrast, the racking technique offers information of "what has changed" [56, 58], allowing the researchers to spell out the effect of the applied pedagogical innovation [18, 33, 34]. Although the instrument measurement result of this work is not data-rich, the analysis strength of the Rasch model can describe in detail the conceptual changes, both in the students' ability and item difficulty levels.

Limitations and Further Studies

The primary limitation of this research is that it did not take into account the aspects of learning style, culture, and motivation that can change due to learning interventions. Future studies, therefore, can address these aspects. The present study can be continued by considering the context of a problem that closely connects with the parameter of item difficulty level. The analysis will be more interesting if it can prove that different item difficulty levels are influenced by problem contexts in each item [80]. Further studies are also expected to find an analysis technique that can integrate problem contexts, item characteristics, and item difficulty levels in a measurement model. It is assumed that different problem contexts in each item will be more likely to affect measurement results because problem contexts have conceptual linkage with items and student activities in doing experiments, measuring, interpreting data/graphs, and others. Thus, the linkages between the learning process during the intervention and conceptual changes in students' ability and item

difficulty levels can be explained in detail; which part of the process leads the students to change their understanding related to specific ideas taught to them.

Acknowledgment

We would like to express our deep gratitude to all parties for their assistance in data processing, especially the Head of Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo.

Conflicts of interest

The authors declare no conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material

All the data are presented in the manuscript. The data supporting the findings of the article is also available in the appendix section in the end part of the manuscript.

Author Contributions

Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing: Lukman Abdul Rauf Laliyo

Validation, supervision, writing—review and editing: Bambang Sumintono

Project administration, resources, funding acquisition: Citra Panigoro

References

1. Lu S, Bi H. Development of a measurement instrument to assess students' electrolyte conceptual understanding. Chem. Educ. Res. Pract. 2016;17(4):1030–1040. doi: 10.1039/c6rp00137h
2. Yildirir HE, Demirkol H. (2018). Identifying mental models of students for physical and chemical change. J. Balt. Sci. Educ. 2018;17(6): 986–1004. doi: 10.33225/jbse/18.17.986
3. Gette CR, Kryjevskaia M, Stetzer MR, Heron PRL. Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy. Phys. Rev. Phys. Educ. Res. 2018;14(1): 10113. doi: 10.1103/PhysRevPhysEducRes.14.010113
4. Johnstone AH. Chemical education research in Glasgow in perspective. Chem. Educ. Res. Pract. 2006;7(2): 49–63. doi: 10.1039/b5rp90021b
5. Johnstone AH. You can't get there from here. J. Chem. Educ. 2010;87(1): 22–29. doi: 10.1021/ed800026d
6. Taber KS. Chemical misconceptions—Prevention, diagnosis, and cure. J. Chem. Educ. 2003; 80(5): 491.
7. Taber KS. Challenging misconceptions in the chemistry classroom: Resources to support teachers. Educació Química EduQ. 2009;4:13–20. doi: 10.2346/20.2003.02.27
8. Alamina JI, Etokeren IS. Effectiveness of imagination stretch teaching strategy in correcting misconceptions of students about particulate nature of matter. J. Educ. Soc. Behav. Sci. 2018;27(1):1–11. doi: 10.9734/jesbs/2018/43063
9. Yaşar IZ, İnce E, Kırbaşlar FG. 7. Class science and technology course "structure of atom" subject readiness improvement test. 2014;152:662-667. Procedia Soc. Behav. Sci.

doi: 10.1016/j.sbspro.2014.09.259

10. Hoe KY, Subramaniam R. On the prevalence of alternative conceptions on acid-base chemistry among secondary students: Insights from cognitive and confidence measures. Chem. Educ. Res. Pract. 2016;17(2):263–282. doi: 10.1039/c5rp00146c

11. Aktan DC. Investigation of students' intermediate conceptual understanding levels: The case of direct current electricity concepts. Eur. J. Phys. 2013;34(1):33–43. doi: 10.1088/0143-0807/34/1/33

12. Allen M. Misconceptions in primary science. Maidenhead: Open University Press; 2014.

13. Soeharto, Csapó B, Sarimanah E, Dewi FI, Sabri T. A review of students' common misconceptions in science and their diagnostic assessment tools. J. Pendidik. IPA Indones. 2019;8(2):247–266. doi: 10.15294/jpii.v8i2.18649

14. Almuntasheri S, Gillies RM, Wright T. The effectiveness of a guided inquiry-based, teachers' professional development programme on saudi students' understanding of density. Sci. Educ. Int. 2016:27(1); 16–39.

15. Hashweh MZ. The complexity of teaching density in middle school. Res. Sci. Technol. Educ. 2016;34(1): 1–24. doi: 10.1080/02635143.2015.1042854

16. Zvoch K, Holveck S, Porter L. Teaching for conceptual change in a density unit provided to seventh graders: A comparison of teacher- and student-centered approaches. Res. Sci. Educ. 2019;51:1395–1421. doi: 10.1007/s11165-019-09907-8

17. Grob R, Holmeier M, Labudde P. Formative assessment to support students' competences in inquiry-based science education. Interdiscip. J. Probl-based. Learn. 2017:11(2). doi: 10.7771/1541-5015.1673

18. Pentecost TC, Barbera J. Measuring learning gains in chemical education: A comparison of two methods. J. Chem. Educ. 2013;90(7):839–845. doi:

10.1021/ed400018v

19. Sumintono B. Rasch model measurements as tools in assesment for learning. In: Proceedings of the 1st International Conference on Education Innovation (ICEI 2017); 2017 October 14; Surabaya. Paris: Atlantis Press; 2018. Available from: doi: 10.2991/icei-17.2018.11

20. Micelli R, Settanni M, Vidotto G. Measuring change in training programs : An empirical illustration Measurement of change and Item Response Theory. Psychol. Sci. Q. 2008;50(3): 433–447.

21. Willoughby SD, Metz A. Exploring gender differences with different gain calculations in astronomy and biology. Am. J. Phys. 2009;77(7):651–657. doi: 10.1119/1.3133087

22. Linn RL, Slinde JA. The determination of the significance of change between pre- and posttesting periods. Rev. Educ Res. 1977;47(1):121–150. doi: 10.3102/00346543047001121

23. He P, Liu X, Zheng C, Jia M. Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. Chem. Educ. Res. Pract. 2016;17(2):381–393. doi: 10.1039/C6RP00004E

24. Sumintono B, Widhiarso W. Aplikasi pemodelan Rasch pada assessment pendidikan [Application of Rasch modeling in educational assessment]. Bandung: Penerbit Trim Komunikata; 2015. https://www.researchgate.net/publication/282673464%0AAplikasi

25. Boone WJ, Yale MS, Staver JR. Rasch analysis in the human sciences. Dordrecht: Springer; 2014. doi: 10.1007/978-94-007-6857-4

26. Liu X. Developing measurement instruments for science education research. In: Fraser B, Tobin KG, McRobbie CJ, editors. Second International Handbook of Science Education. Netherlands: Springer; 2012. Pp. 651-665.

27. Linacre JM. 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. US: Winsteps; 2020.

28. Sabah S, Hammouri H, Akour M. Validation of a scale of attitudes toward science across countries using Rasch model: Findings from TIMSS. J. Balt. Sci. Educ. 2013;12(5):692–703.

29. Hadenfeldt JC, Bernholt S, Liu X, Neumann K, Parchmann I. Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter. J. Chem Educ. 2013; 90(12):1602–1608. doi: 10.1021/ed3006192

30. Herrmann-Abell CF, DeBoer GE. Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. Chem. Educ. Res. Pract. 2011;2(2):184–192. doi: 10.1039/c1rp90023d

31. Herrmann-Abell CF, Deboer GE. Using Rasch modeling and option probability curves to diagnose students' misconceptions. In: the 2016 AERA Annual Meeting; 2016 April 8-12; Washington DC. USA: American Educational Research Assossiation; 2016. Available from: https://www.aaas.org/sites/default/files/Project2061_CHA-AERA%20energy%20paper%204-7-16.pdf

32. Laliyo LAR, Tangio JS, Sumintono B, Jahja M, Panigoro C. Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter. J. Balt. Sci. Educ. 2020;19(5). doi: 10.33225/jbse/20.19.824

33. Davidowitz B, Potgieter M. Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. Int J. Sci. Educ. 2016;38(9):1483–1503. doi: 10.1080/09500693.2016.1196843

34. Park M, Liu X. An investigation of item difficulties in energy aspects across biology,

chemistry, environmental science, and physics. Res. Sci. Educ. 2019. doi: 10.1007/s11165-019-9819-y

35. Treagust DF, Qureshi SS, Vishnumolakala VR, Ojeil J, Mocerino M, Southam DC. Process-oriented guided inquiry learning (POGIL) as a culturally relevant pedagogy (CRP) in Qatar: A perspective from grade 10 chemistry classes. Res. Sci Educ. 2018:1–19. doi: 10.1007/s11165-018-9712-0

36. Kinslow AT, Sadler TD, Nguyen HT. Socio-scientific reasoning and environmental literacy in a field-based ecology class. Environ. Educ. Res. 2018;4622:1–23. doi: 10.1080/13504622.2018.1442418

37. Owens DC, Sadler TD, Friedrichsen P. Teaching practices for enactment of socio-scientific issues instruction: An instrumental case study of an experienced biology teacher. Res. Sci. Educ. 2019. doi: 10.1007/s11165-018-9799-3

38. Bruder R, Prescott A. Research evidence on the benefits of IBL. ZDM Int. J. Math. Educ. 2013;45(6):811–822. doi: 10.1007/s11858-013-0542-2

39. Zeidler DL. Socioscientific issues as a curriculum emphasis: theory, research and practice. In: Lederman NG, Abell SK editor. Handbook of research on science education. New York: Routledge; 2014. pp. 697–726.

40. Espeja AG, Lagarón DC. Socio-scientific Issues (SSI) in initial training of primary school teachers: Pre-service teachers' conceptualization of SSI and appreciation of the value of teaching SSI. Procedia Soc. Behav. Sci. 2015;196:80–88. doi: 10.1016/j.sbspro.2015.07.015

41. National Research Council. A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington DC: The National Academies Press; 2012. doi: 10.17226/13165

42. Sadler TD, Zeidler DL. Scientific literacy, PISA, and socioscientific discourse:

Assessment for progressive aims of science education. J. Res. Sci. Teach. 2009;46(8):909–921. doi: 10.1002/tea.20327

43. Lederman NG, Lederman JS, & Antink A. Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. Int. J. Educ. Math. Sci. Technol. 2013;1(3):138–147. doi: 10.18404/ijemst.19784

44. Cooke AN, Fielding KS, Louis WR. Environmentally active people: The role of autonomy, relatedness, competence and self-determined motivation. Environ. Educ. Res. 2016;22(5):631–657. doi: 10.1080/13504622.2015.1054262

45. Sadler TD. Moral and ethical dimensions of socioscientific decision-making as integral components of scientific literacy. Sci. Educ. 2004;13(1): 39–48.

46. DeBoer GE. Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. J. Res. Sci. Teach. 2000;37(6):582–601. doi: 10.1002/1098-2736(200008)37:6<582::AID-TEA5>3.0.CO;2-L

47. Sadler TD. Socio-scientific issues-based education: What we know about science education in the context of SSI. In: Sadler TD, editor. Socio-Scientific Issues in the Classroom: Teaching, Learning and Research. Dordrecht: Springer; 2011. pp. 355–369. doi: 10.1007/978-94-007-1159-4_16.

48. Gräber W, Nentwig P, Becker HJ, Sumfleth E, Pitton A, Wollweber K, Jorde D. Scientific literacy: From theory to practice. Res. Sci. Educ. Past. Present. Future. 2005;1996:61–70. doi: 10.1007/0-306-47639-8_6

49. Grooms J. A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction. J. Chem. Educ. 2020. doi: 10.1021/acs.jchemed.0c00026

50. Presley ML, Sickel AJ, Muslu N, Merle D. A framework for socio-scientific issues

based education. Sci. Educ. 2013;22:26–32.

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

52. Seçken N. Identifying student's misconceptions about SALT. Procedia Soc. Behav. Sci. 2010;2(2):234–245. doi: 10.1016/j.sbspro.2010.03.004

53. Damanhuri MIM., Treagust DF, Won M, Chandrasegaran AL. High school students' understanding of acid-base concepts: An ongoing challenge for teachers. Int. J. Environ. Sci. Educ. 2016;11(1):9–27. doi: 10.12973/ijese.2015.284a

54. Orwat K, Bernard P, Migdał-Mikuli A. Alternative conceptions of common salt hydrolysis among upper-secondary school students. J. Balt. Sci. Educ. 2017;16(1): 64–76.

55. Creswell JW. Educational research: Planning, conducting, and evaluating quantitative and qualitative research. 4th ed. New York: Pearson Education, Inc; 2012.

56. Wright BD. Rack and stack: Time 1 vs. time 2 or pre-test vs. post-test. Rasch Meas. Trans. 2003;17(1):905–906.

57. Anselmi P, Vidotto G, Bettinardi O, Bertolotti G. Measurement of change in health status with Rasch models. Health Qual. Life Outcomes. 2015;13(1):1–7. doi: 10.1186/s12955-014-0197-x

58. Ling M, Pang V, Ompok CC. Measuring change in early mathematics ability of children who learn using games: stacked analysis in rasch measurement. In: Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings, 2016 July 30-August 3; Singapore. Singapore: Springer; 2018. Available from: https://doi.org/10.1007/978-981-10-8138-5

59. Taber KS. Ethical considerations of chemistry education research involving "human

subjects”. Chem. Educ. Res. Pract. 2014;15(2):109–113. doi: 10.1039/c4rp90003k

60. Aldresti F, Rahayu S, Fajaroh F. The influence of inquiry-based chemistry learning with the contex of socio-scientific issues on high school students’ scientific explanation skills. J. Pendidik. IPA. 2019;23(2). doi:10.18269/jpmipa.v23i2.XXX

61. Mitarlis, Ibnu S, Rahayu S, Sutrisno. The effectiveness of new inquiry-based learning (NIBL) for improving multiple higher-order thinking skills (M-HOTS) of prospective chemistry teachers. Eur. J. Educ. Res. 2020;9(3):1309–1325. doi: 10.12973/eu-jer.9.3.1309

62. Pedaste M, Mäeots M, Siiman LA, de-Jong T, Van-Riesen SAN, Kamp ET, Manoli CC, Zacharia ZC, Tsourlidaki, E. Phases of inquiry-based learning: Definitions and the inquiry cycle. Educ. Res. Rev. 2015;14:47–61. doi: 10.1016/j.edurev.2015.02.003

63. Romine WL, Sadler TD. Measuring changes in interest in science and technology at the college level in response to two instructional interventions. Res. Sci. Educ. 2016;46(3):309–327. doi: 10.1007/s11165-014-9452-8

64. Wilson M. Constructing measures: an item response modeling approach. USA: Lawrence Erlbaum Associates, Inc; 2005. doi: doi: 10.4324/9781410611697

65. Wilson M. Cognitive diagnosis using item response models. Zeits Für Psych/J. Psych. 2008;216(2):74–88. doi: 10.1027/0044-3409.216.2.74

66. Wilson M. Measuring progressions: Assessment structures underlying a learning progression. J. Res. Sci. Teach. 2009;46(6):716–730. doi: 10.1002/tea.20318

67. Chittleborough G, Treagust D. Correct interpretation of chemical diagrams requires transforming from one level of representation to another. Res. Sci. Educ. 2008;38(4): 463–482. doi: 10.1007/s11165-007-9059-4

68. Treagust DF. Development and use of diagnostic tests to evaluate students’ misconceptions in science. Int. J. Sci. Educ. 1988;10(2):159–169. doi:

10.1080/0950069880100204

69. Arslan HO, Cigdemoglu C, Moseley C. A Three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. Int. J. Sci. Educ. 2012;34(11):1667–1686. doi: 10.1080/09500693.2012.680618

70. Hasan S, Bagayoko D, Kelley EL. Misconceptions and the certainty of response index (CRI). Phys. Educ. 1999;34(5):294–299. doi: 10.1088/0031-9120/34/5/304

71. Habiddin H, Page EM. Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). Indones. J. Chem. 2019;19(3):720–736. doi: 10.22146/ijc.39218

72. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences. 3rd ed. New York: Routledge Taylor & Francis Group; 2015. doi: 10.1088/1751-8113/44/8/085201

73. Herrmann-abell CF, Flanagan JC, Roseman JE. Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: Using student measures to evaluate the promise of the intervention. In: Proceeding of the 2013 NARST Annual International Conference. 2013 April 6-9; Puerto Rico. Puerto Rico: NARST; 2013. Available from: https://www.aaas.org/sites/default/files/project2061/files/NARST2013-Flanagan-Paper5-Usingteachermeasurestoe.pdf

74. Fisher WP. Rating scale instrument quality criteria. Rasch Meas. Trans. 2007;21(1): 1095. www.rasch.org/rmt/rmt211m.htm

75. Adams WK, Wieman CE. Development and validation of instruments to measure learning of expert-like thinking. Int. J. Sci. Educ. 2011;33(9):1289–1312. doi: 10.1080/09500693.2010.512369

76. Sumintono B, Widhiarso W. Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial [Application of Rasch Model for Social Science Studies]. In Trim B, editor.). Bandung: Trim Komunikata Publishing House; 2014. https://www.researchgate.net/publication/268688933%0AAplikasi

77. Rodriguez JG, Hunter KH, Scharlott LJ, Becker NM. (2020). A Review of Research on Process Oriented Guided Inquiry Learning: Implications for Research and Practice. J. Chem. Educ. 2020;97(10):3506–3520. doi: https://doi.org/10.1021/acs.jchemed.0c00355

78. Hancock TS, Friedrichsen PJ, Kinslow AT, Sadler TD. Selecting socio-scientific issues for teaching: A grounded theory study of how science teachers collaboratively design SSI-based curricula. Sci. Educ. 2019;28(6–7):639–667. doi: 10.1007/s11191-019-00065-x

79. Ültay N, Çalik M. A comparison of different teaching designs of "acids and bases" subject. Eurasia J. Math. Sci. Technol. Educ. 2016;12(1):57–86. doi: 10.12973/eurasia.2016.1422a

80. Hartig J, Frey A, Nold G, Klieme E. An application of explanatory item response modeling for model-based proficiency scaling. Educ. Psychol. Meas. 2012;72(4):665–686. doi: 10.1177/0013164411430707

Appendix 1. Standardized Residual Variance in Eigen value Units.

```
     Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item
information units
                                          Eigenvalue   Observed   Expected
Total raw variance in observations     =     22.7067 100.0%         100.0%
```

Raw variance explained by measures	=	7.7067	33.9%		35.9%
Raw variance explained by persons	=	2.7733	12.2%		12.9%
Raw Variance explained by items	=	4.9334	21.7%		23.0%
Raw unexplained variance (total)	=	15.0000	66.1%	100.0%	64.1%
Unexplned variance in 1st contrast	=	2.0698	9.1%	13.8%	
Unexplned variance in 2nd contrast	=	1.5312	6.7%	10.2%	
Unexplned variance in 3rd contrast	=	1.3696	6.0%	9.1%	
Unexplned variance in 4th contrast	=	1.3124	5.8%	8.7%	
Unexplned variance in 5th contrast	=	1.1945	5.3%	8.0%	

Appendix 2. Summary of Category Structure.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	317	6	-.18	-.20	1.06	1.08	NONE	(-1.73)	1
2	2	190	3	.10	.09	1.03	1.18	.46	-.77	2
3	3	963	17	.33	.31	1.02	.93	-1.43	-.22	3
4	4	542	10	.56	.52	1.02	.97	.98	.21	4
5	5	1262	22	.62	.73	1.27	.99	-.22	.74	5
6	6	2425	43	1.02	.98	.97	1.04	.20	(1.76)	6
MISSING		1	0	-.30						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

REVIEW 3

Gmail

Lukman Laliyo <lukman.laliyo019@gmail.com>

Decision on submission HELIYON-D-21-05440R2 to Heliyon

Heliyon <em@editorialmanager.com> Fri, Mar 4, 2022 at 6:11 PM
Reply-To: Heliyon <info@heliyon.com>
To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Ms. No.: HELIYON-D-21-05440R2
Title: Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model
Journal: Heliyon

Dear Dr. Laliyo,

Thank you for submitting your manuscript to Heliyon.

We have now received all of the editor and reviewer comments on your recent submission to Heliyon. Your paper will become acceptable for publication after implementation of minor formatting and/or administrative changes outlined below. To avoid unnecessary delays in the publication of your manuscript, please do not make any other additional changes during this revision.

To submit your revised manuscript, please log in as an author at https://www.editorialmanager.com/heliyon/, and navigate to the "Submissions Needing Revision" folder under the Author Main Menu. When submitting your revised manuscript, please ensure that you upload your most recent document with the "Revised manuscript file - highlighting revisions made" item type.

Kind regards,

Mengpei Yan
Editorial Assistant
Heliyon

Embargo
Embargos are not automatically set for papers published in Heliyon. Papers appear online a few days after acceptance. To request a media embargo and/or publication on a specific date to assist an institutional press release, please reach out to the Heliyon team (info@heliyon.com) as soon as possible and we will do our best to accommodate your request.

Editorial Office comments:

Please remove "Funding", "Author Contributions", "section from your manuscript file, as this information is handled separately.

*****

Data in Brief (optional):
We invite you to convert your supplementary data (or a part of it) into an additional journal publication in Data in Brief, a multi-disciplinary open access journal. Data in Brief articles are a fantastic way to describe supplementary data and associated metadata, or full raw datasets deposited in an external repository, which are otherwise unnoticed. A Data in Brief article (which will be reviewed, formatted, indexed, and given a DOI) will make your data easier to find, reproduce, and cite.

You can submit to Data in Brief when you upload your revised manuscript. To do so, complete the template and follow the co-submission instructions found here: www.elsevier.com/dib-template. If your manuscript is accepted, your Data in Brief submission will automatically be transferred to Data in Brief for editorial review and publication.

Please note: an open access Article Publication Charge (APC) is payable by the author or research funder to cover the costs associated with publication in Data in Brief and ensure your data article is immediately and permanently free to access by all. For the current APC see: www.elsevier.com/journals/data-in-brief/2352-3409/open-access-journal

Please contact the Data in Brief editorial office at dib-me@elsevier.com or visit the Data in Brief homepage (www.journals.elsevier.com/data-in-brief/) if you have questions or need further information.

More information and support
FAQ: How do I revise my submission in Editorial Manager?
https://service.elsevier.com/app/answers/detail/a_id/28463/supporthub/publishing/

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors
FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

---

REVISION 3

To
Editor-in-Chief
Heliyon

March 8, 2022

Dear editor-in-chief,

Thank you for your reply regarding our manuscript HELIYON-D-21-05440 entitled **"**Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model.**"**

We are grateful for the editor's comments, and the positive evaluation of our work. We have removed the mentioned sections according to the editor's suggestion. As a consequence we provide some changes and added many new and clarifying statements in all parts of the paper. These changes have improved the manuscript considerably and we hope that it can be published without delay.

Sincerely,
Corresponding Author on Behalf of All Authors,

Lukman A. R. Laliyo

Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model

Lukman Abdul Rauf Laliyo[1,*], Bambang Sumintono[2], Citra Panigoro[3]

[1] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo, 96128, Indonesia
[2] Institute of Educational Leadership, University of Malaya, Kuala Lumpur, 50603, Malaysia
[3] Department of Aquatic Resource Management, Faculty of Fisheries and Marine Science, Universitas Negeri Gorontalo, 96128, Indonesia

*Corresponding Author

Email: lukman.laliyo019@gmail.com

Authors' ORCID ID

Lukman Abdul Rauf Laliyo	: https://orcid.org/0000-0003-3281-7202
Bambang Sumintono	: https://orcid.org/0000-0002-5490-3665
Citra Panigoro	: -

Highlights

- Raw scores have a bias in a conventional psychometric measurement
- Stacking and racking measure students' ability and item difficulty level changes
- The learning process in socio-scientific issues improves students' understanding
- Misconceptions influence the negative values of students' pre-and post-test

Abstract

This research aimed to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the process-oriented guided inquiry learning (POGIL) model in the context of socio-scientific issues (SSI) with the pretest-posttest control group design. Such techniques were based on a person- and item-centered statistic to determine how students and items changed during interventions. Eleventh-grade students in one of the top-ranked senior high schools in the eastern part of Indonesia were involved as the participants. They provided written responses (pre- and post-test) to 15 three-tier multiple-choice items. Their responses were assessed through a rubric that combines diagnostic measurement and certainty of response index. Moreover, the data were analyzed following the Rasch Partial Credit Model, using the WINSTEPS 4.5.5 software. The results suggested that students in the experimental group taught by the POGIL approach in the SSI context had better positive conceptual changes than those in the control class learning with a conventional approach. Along with the intervention effect, in certain cases, it was found that positive conceptual changes were possibly due to student guessing, which happened to be correct (lucky guess), and cheating. In other cases, students who experienced negative conceptual changes may respond incorrectly due to carelessness, the boredom of problem-solving, or misconception. Such findings have also proven that some students tend to give specific responses after the intervention in certain items, indicating that not all students fit the intervention. Besides, stacking and racking analyses are highly significant in detailing every change in students' abilities, item difficulty levels, and learning progress.

Keywords: stacking, racking, Rasch model, hydrolysis conceptual changes, inquiry model

Introduction

Central to defining the quality of pedagogical innovation in science classes is conceptual changes. The changes refer to how ideas or conceptions the students understand according to their ways of thinking [1, 2] become scientifically accurate [3]. It is because such ideas generally comprise misconceptions [4, 5, 6, 7], are not in accordance with scientific concepts [8, 9], tend to be resistant [10], changeable and varied [11], so that they should be improved if the correct conceptual understanding is to be taught [12, 13].

Some studies have been conducted on learning innovation testing to form an accurate and scientific conceptual understanding of the students, e.g., inquiry-based learning. This model presents conceptual conflicts and participatory experiments to facilitate conceptual changes [14, 15, 16]. Conceptual understanding-based learning involves various strategies in identifying and analyzing students' comprehension so that the investigation process can be designed to lead them to a more accurate and scientific conception [16, 17]. This research relied on a quasi-experimental design that assessed students' pre-test and post-test, evaluated the changes in performances for testing significant differences. This type of testing informs the researcher about the presence of an effect, but does not provide detailed information on the level and trait of the changes [18]. What if the researcher is willing to compare the extent to which the pre- and post-test change (differences in learning outcomes) and interpret the changes (the reasoning why those changes occur) in terms of content? This is a core question regarding the changes in some latent traits or changes in traits measured after the intervention. In most studies, interpreting the changes in pre-test and post-test tends to be limited to identifying whether or not an effect prevails.

Pre- and post-test changes should be given in detail regarding the students' understanding ability and item difficulty levels. However, this has not been much revealed due

to the limitations of its measurement techniques and analyses and has not been the main focus in chemistry education research to date. One reason for this issue is the debate in the psychometric community regarding the ability to measure changes accurately [18]. This debate questions the use of raw scores in the conventional psychometric analysis, which largely contains measurement biases [19], as follows: 1) the difference in pre- and post-test scores will be negatively correlated with the pre-test score, especially for students with low pre-test scores [18, 20]; 2) the difference in pre- and post-test scores shows low test reliability [21]; 3) low measurement properties due to different scales [22].

Raw scores are not final data, so that they do not have a great deal of information for drawing conclusions [23, 24]. Around the 1950s, Dr. Georg Rasch, a mathematician from Denmark, introduced the formulation of the Rasch measurement model [24]. The model has been widely applied to analyze various types of data, e.g., dichotomous, polytomous, multi-rating, and multi-rater data. In the mid-2000s, the Rasch model was used as a probabilistic-based psychometric measurement that went beyond the use of raw scores [25, 26], and was used to overcome the limitations of conventional psychometric measurement [19, 27]. Its analyses, including item fit, PCA (Principal Component Analysis), and Wright map, are commonly used for international test analyses, namely TIMSS and PISA [28].

In chemistry education research, the Rasch model has been relied on to evaluate learning understanding and progress [29], to diagnose students' preconceptions [1], misconceptions [13, 30, 31, 32], link the measurement of content knowledge with pedagogical content knowledge [33], and investigate item difficulty patterns [13, 34]. Even so, studies on the Rasch model to reveal the chemistry conceptual changes in students' understanding and item difficulty levels are relatively hard to find as of today. The present study aims to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the POGIL approach in the context of SSI and

students who learn conventionally. Such techniques are based on a person- and item-centered statistic to estimate how students and items change during the intervention.

POGIL is a student-centered learning strategy that teaches content or process skills. The philosophical foundation of POGIL is the involvement of an interactive process of careful thinking, discussing ideas, perfecting understanding, practicing skills, reflecting progress, and evaluating performances [35]. POGIL is able to lead the process of designing a participatory experiment that presents a conceptual conflict as a strategy to encourage students to form an accurate concept [14]. Therefore, POGIL intervention is more likely to be potential in driving epistemological understanding and reasoning [36], making students have opportunities to change their conceptions to be more accurate and scientific [16]. Nevertheless, it is also worth noting that some students potentially have misconceptions resistant to changes [3].

SSI functions as a learning context through the intergration of social problems that students are familiar with. It also has a conceptual connection with salt hydrolysis [37, 38], and its resolution requires many perspectives [39], including the dimension of moral and ethical evaluation of students [40]. The SSI context is a socio-scientific phenomenon that the students should explain based on their conceptual viewpoints. It encourages them to actively get involved in grasping problems [41], developing and utilizing their knowledge [42], improving their critical thinking [43], and being able to scientifically describe the discussed socio-scientific phenomenon [36, 44, 45]. For such reasons, the integration of SSI can build up students' scientific literacy [39, 46, 47]. In the end, this integration enables the learning process to be more significant in enhancing students' understanding [45, 48]. Besides, they are skilled in negotiating the social aspect of the studied phenomenon [49, 50]. For instance, the issues of global warming, climate change, and pollution [36].

Salt hydrolysis is a learning topic in high school that is strongly related to SSI. Students with a good understanding of hydrolysis will manage to clarify scientifically why

detergents, bleaching agents (NaOCl), and fertilizers can pollute the environment. Despite this, the linkage of this issue as the problem in learning hydrolysis is rarely carried out. The learning process is more emphasized on mastering theoretical concepts [36]. As a consequence, students find it challenging to use their hydrolysis understanding to explain socio-scientific phenomena around them [37]. This challenge is on account of their misconceptions regarding acid-base reaction [51], making them unable to elaborate the concept of salt hydrolysis [52] and determine acid and base strength [53]. In addition, they are struggling with correctly explaining the dissolving process and the reaction of ionic compounds with water, writing down chemical equations, and having different interpretations of the dissolving process mentioned earlier [54]. On this ground, it is essential to reveal how the hydrolysis concept changes if intervened with the POGIL approach in the SSI context, through the following specific questions: (1) is there a significant hydrolysis conceptual change of the students after the learning process in experimental and control groups? (2) if compared, how is the hydrolysis conceptual change through the intervention of POGIL in the SSI context and conventional learning? (3) in addition to intervention, is there any other factor that also contributes to the students' hydrolysis conceptual changes?

Method of Study

This study relied on a quantitative approach with a quasi-experimental and pretest-posttest control group design [55] by comparing the extent to which the hydrolysis concept changes after the intervention. Researchers carried out the learning process for 12 meetings, gave tests, and collected data on the results of the intervention and measurement.

The changes of students and items were analyzed using the stacking and racking techniques in the Rasch model [56]. As standard techniques, racking and stacking were introduced by Benjamin Wright to measure the extent to which conceptual understanding of students and items change before and after interventions [57]. The referred changes are cases (item and student levels) caused by the learning intervention and can be diagnosed based on the estimated changes.

In regards to students' understanding, the measurement was to identify students who had specific hydrolysis conceptual changes in responding to the learning intervention. In terms of items, the measurement was done to identify which items had special characteristics and been understood by students differently during the learning intervention [57]. Thus, the scientific inquiry approach might not be suitable for some students, or some items might be too hard after the intervention. This insightful information is immensely helpful for researchers and education practitioners, especially in evaluating the weaknesses of pedagogical innovations being applied and devising learning strategies that meet students' needs in learning [58].

Participants

Eleventh-grade students aged 16-17 years in one of the senior high schools in the eastern part of Indonesia were involved as the sample. This top-ranked school gets an "A" accreditation (excellent) from the National Accreditation Board for High School. The sample was determined by convenience sampling in six randomly assigned classes. Three classes (N=97) were experimental groups that applied the POGIL model in the SSI context. The other three classes (N=93), as control groups, applied conventional learning without the SSI context. The same teacher taught these classes following the Curriculum 2013 of Chemistry Subject (revised in 2016). There was no special classroom for learning the concept of hydrolysis, i.e., taking up the regular learning process at school. Before learning the hydrolysis concept, the students had previously learned the concept of acid and base to understand the concept of salt hydrolysis way better. Research permission was obtained from the government and school administrators. In accordance with principles of research ethics, research purpose and procedures were informed to all the students being involved and that they were voluntarily participating. Additionally, their information is confidential and only used for science development [59].

Learning implementation

Students in the experimental group studied employing the process-oriented guided inquiry learning (POGIL) in the SSI context [35]. Meanwhile, in the control class, the learning process was performed conventionally; the teacher facilitated learning initiatives. The learning process focused more on content mastery and problem-solving practice. Applying the POGIL model in the SSI context highlights teacher assistance to guide the students to prepare their conceptual understanding based on epistemological reasoning they get from experiments, discussions, and collaborations [49, 60]. Researchers carried out the learning process for eight weeks to apply the intervention to the sample, gave tests, collected data on the results of the intervention and measurement. The first three weeks were the preparation stages when researchers and the teacher shared perceptions, and asked the teacher to perform a learning simulation under the scenario, including different assistance techniques in leading the students to conduct experiments, and to ask analytical questions. The pre-test was carried out in the third week. Further, the learning implementation was done for four weeks, and the post-test was executed in the eighth week.

The learning stages with POGIL in the SSI context consist of orientation, exploration, concept formation, application, and closing. During the orientation stage, the teacher presented familiar contextual phenomena related to the concept of hydrolysis. The teacher asked initial questions to provoke curiosity and arouse motivation and interest of the students. While watching the video, had the students responded and explained the relationship between the phenomena and acids and bases, hydrolysis, and buffers. In the exploration stage, the teacher developed analytical questions with data, images, and multiple video clips to give perspectives on learning objectives and to delve into the concept that had been and would be learned. Next, the teacher assisted the students in doing experiments guided by a worksheet, and at the same time, asked analytical questions to lead them and strengthen their conceptual understanding. In the concept formation stage, the teacher asked students to build their conceptual understanding based on the exploration results, accompanied by critical and fundamental questions to guide students in building a conceptual understanding of the salt hydrolysis and buffer solution.

Following the formation stage was the application stage when the teacher presented contextual problems in the SSI context, particularly those comprising social problems in society, that closely linked with the understanding of salt hydrolysis and buffer solution concepts. Such problems included 1) the use of bleaching agents (detergents), 2) the functions of alum $KAl(SO_4)_2 \cdot 12H_2O$ for water purification, 3) the harmful effects of detergent waste, 4) the beneficial and harmful effects of artificial fertilizer $(NH_4)_2SO_4$ for soil fertility, and 5) the harmful effects of monosodium glutamate (MSG) for health. In this stage, the teacher guided the students through collaborative discussions and critical questions, intending to give them perspectives on SSI phenomena and encourage them to collect information and do experiments following student activity sheets. Thereupon, the students had presentation and discussion sessions, during which they reported their experiment results and drew conclusions [61, 62]. The teacher asked them to describe the possible problems and solutions from their understanding of the studied concepts. This enabled the students to form their conceptual understanding that is closely related to contexts; the learning process was from contextual to abstract [37, 63]**.** From such a condition, the teacher led the students to apply their knowledge in different contexts and situations and solve problems. The final stage was closing or teacher assistance in guiding the students to explain the conclusion and reflection on the learning process as the end of the learning activities.

Instrument

Table 1 displays 15 items of diagnostic three-tier multiple choice test to measure students' hydrolysis conceptual understanding. The test was constructed following the Competence Standard of 2013 Chemistry Curriculum of Senior High School under Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 37 of 2018. The procedures of developing the instrument followed the recommendation by [64, 65, 66].

Table 1. Conceptual Map of Hydrolysis Concept Understanding [67]

Problem Context	Item	Conceptual Understanding	Ability Level	
Bleaching agents are formed of weak acid HOCl and strong base NaOH. Sodium hypochlorite salt (NaOCl) is reactive and dissolves the dye. In the water, the ion OCl^{-} will be hydrolyzed to HOCl and OH^{-} -	1	Balancing the salt (NaOCl) hydrolysis reaction in the water	2	Level 3: Students are able to calculate the pH of the hydrolyzed salt solution.
	2	Stating the partial hydrolysis reaction: $NaOCl \rightarrow Na^{+} + OCl^{-}$	2	
	3	Determining corrosive alkali of sodium hypochlorite salt (NaOCl)	1	
	4	Calculating the pH of hydrolysis of sodium hypochlorite salt (NaOCl) with NaOCl = 0.1 M; Ka = 10^{-5})	3	Level 2: Students are able to determine the hydrolysis
	5	Determining the property of NaOCl, in the reaction: $OCl^{-} + H_2O \rightarrow HOCl + OH^{-}$	2	

	6	Calculating the pH of sodium hypochlorite salt (NaOCl) that comes from a mixture of HOCl and NaOH (partially hydrolyzed), if the Ka HOCl is 10^{-5} and there is an increase in the pH of the solution mixture.	3	reaction from different types of salt Level 1: Students are able to analyze the properties of the hydrolyzed salt
Water purification with alum $KAl(SO_4)_2 \cdot 12H_2O$ is the concept of salt hydrolysis, formed of H_2SO_4 and $Al(OH)_3$.	7	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water	1	
	8	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water that is partially hydrolyzed by the Al^{3+} ion	1	
The sodium tripolyphosphate (STPP) in detergents can pollute the environment, a eutrophication process.	9	Determining the properties of detergent solution causing eutrophication	1	
	10	Determining the properties of detergent solution (sodium tripolyphosphate salt) that is partially hydrolyzed	1	
	11	Determining the impact of the disposal of detergent waste on the environment	2	

ZA fertilizer $(NH4)_2SO_4$ is an acidic salt.	12	Determining the properties of ammonium sulfate salt $(NH4)_2SO_4$	1	
	13	Stating the equation of $(NH_4)_2SO_4$ reaction in the water, partially hydrolyzed	2	
Monosodium glutamate $(C_5H_8NO_4Na)$ is L-glutamic acid salt, adversely impactful on human health	14	Students' attitude towards the use of monosodium glutamate $(C_5H_8NO_4Na)$	2	
	15	Determining the properties of monosodium glutamate salt $(C_5H_8NO_4Na)$	1	

Each item was designed in three questions (Q1, Q2, Q3) that integrated diagnostic [68, 69] and summative measurements [10] and certainty of response index (CRI) [70, 71]. Students' responses to items (Q1, Q2, Q3) were evaluated based on the rubric (Table 2). For example, students' responses to items were as follows: Q1, Q2 "correct", and Q3 "very sure" under the code CCC. Such a code indicated that students' conceptual understanding was in level 6, category of Scientific Knowledge (SK). On the other hand, if the response patterns in Q1, Q2 "incorrect" and Q3 "not sure", the code would be IIU, implying that students' conceptual understanding was in the category of Lack of Knowledge (LOK), or level 1. This instrument had been validated from the aspects of item conformity with the construct variable and language. The validity results by three experts were stated under Fleiss' kappa (K = .96), meaning that the experts agreed that the item validity was categorized good.

Table 2 All Possibilities of Responses [70, 71, 72]

(Q1)	(Q2)	(Q3)	Code	Conceptual Understanding Category	Level
Correct	Correct	Certain	CCC	Scientific Knowledge (SK)	6
Correct	Incorrect	Certain	CIC	Misconception False Positive (MFP)	5
Incorrect	Correct	Certain	ICC	Misconception False Negative (MFN)	4
Incorrect	Incorrect	Certain	IIC	All-Misconception (ALM)	3
Correct	Correct	Uncertain	CCU	Lack of Confidence/Lucky Guess. (LG)	2
Correct	Incorrect	Uncertain	CIU	Lack of Knowledge (LOK)	1
Incorrect	Correct	Uncertain	ICU	Lack of Knowledge (LOK)	1
Incorrect	Incorrect	Uncertain	IIU	Lack of Knowledge (LOK)	1

Data collection and analysis

Before the intervention, this research underwent pre-test data collection; whereas, the post-test data collection was done after the intervention. The construction of pre- and post-test items was the same. Students wrote down their responses on the provided answer sheet. Both tests were supervised by teachers in the school. The students must work on all items according to the allocated time (45 minutes). The instrument was immediately collected and should have the same number as the total participants.

The pre- and post-test measurement data were still ordinal data. The Rasch Partial Credit Model with WINSTEPS 4.5.5 software [27, 73] was used to convert ordinal data into

interval data to have the same logit scale. The result was a data calibration of the levels of student's ability and item difficulty in the same interval.

The stacking analysis technique put pre-test and post-test data vertically [74]; meanwhile, the items appeared once in the experimental and control groups, allowing the researchers to check out any changes of the students after the intervention [56]. The examination was based on the same item, making the changes in students' ability during the pre- and post-test be measured [56]. Hence, each student created two measures of abilities, namely pre-test and post-test, and one measure for each item. The research hypothesis is that the students' conceptual understanding from pre-test to post-test changes, both in the experimental and control groups.

Conversely, the racking analysis technique put both pre- and post-test data horizontally, in which each item appeared twice in data collection, and students' ability only emerged once. This enabled the researchers to check out the effects of learning implementation on each student's ability from the tests, especially the changes in item difficulty levels before and after the intervention [56].

Results

Rasch analysis properties of instrument

The summary of changes in concepts and items analyzed by the Rasch model is presented in Table 1. Table 2 provides the item fit statistic. An item is considered to experience a misfit if the measurement result is not in line with the following three criteria: Outfit mean-square residual (MNSQ): $.5 < y < 1.5$; Outfit standardized mean-square residual (ZSTD): $-2 < Z < +2$; and point measure correlation (PTMEA CORR): $.4 < x < .8$ [25]. All items comply with the Outfit MNSQ criterion; item 15 does not meet the Outfit MNSQ criterion; five items (item 1, 6, 12, 13, and 15) are not in accordance with the Outfit (ZSTD)

criterion; all items meet the PTMEA CORR criterion. Simply put, all items fulfill those criteria mentioned previously (none having a misfit), and are fit and valid.

Table 3. Item Statistics: Misfit Order

Item	Difficult	Error	Outfit MNSQ	Outfit ZSTD	PTMEA CORR.
1	-.38	.05	1.36	2.87	.47
2	.20	.04	1.13	1.56	.49
3	-.36	.05	.91	-.79	.43
4	.33	.04	1.09	.77	.55
5	-.25	.05	.94	-.55	.56
6	.26	.04	1.20	2.44	.41
7	.15	.04	.91	-1.17	.54
8	.47	.04	.90	-1.45	.44
9	-.47	.05	1.19	1.49	.46
10	.08	.04	1.09	1.04	.55
11	-.34	.05	1.04	.42	.51
12	-.06	.04	.71	-3.50	.60
13	.46	.04	.74	-4.12	.55
14	-.36	.05	1.00	.77	.55
15	.26	.04	1.31	3.74	.47

This instrument has a good unidimensionality (Appendix 1). Raw variant index arrives at above the standard of 20% (33.9%), indicating that the instrument can effectively measure students' understanding of the hydrolysis concept [75]. The assessment scale analysis (Appendix 2) informs that the observation mean starts from logit -1.73 for category 1 (LOK)

to logit +1.76 (category 6, SK). This signifies that the category of students' understanding takes place consistently [27]. In addition, the high item separation index (logit 6.71) and the high item reliability (logit .98) (Table 3) indicate that the respondents (students) are sufficient to confirm the level of item difficulty, strengthening the instrument construct validity [27]. The higher the item separation and reliability index, the more confident the researchers are about replicating item placement in other suitable sample students [25, 27]. Person separation index and person reliability that reach logit 2.0 and logit .75 (Table 4), respectively, imply that the instrument is quite sensitive to differentiate the high and low abilities of the students [25, 27]. According to the Rasch model calculation, the coefficient of Cronbach Alpha of logit .81 (Table 4) reflects an interaction between 380 students and 15 items with an excellent category [24, 76]. In other words, the interaction between students and items is very significant. The instrument has an excellent internal psychometric consistency and is considered very reliable.

Table 4. Person Separation and Reliability Statistics

Parameter	Measure	SD	Separation	Reliability	INFIT		OUTFIT		KR-20
					MNSQ	ZSTD	MNSQ	ZSTD	
Person (N=380)	.67	.52	1.72	.75	1.00	.04	1.02	.10	.81
Item (N=15)	.00	.32	6.71	.98	1.07	.41	1.02	-.01	

The Difference in Students' Understanding Ability of Hydrolysis Concept

The result of the Mann-Whitney test (Table 5) brings out the fact that statistically, there is a significant difference in the results of pre-test (U=3459.000), $p<0.05$) and post-test (U=1723.000, $p<0.05$) among students in experimental and control groups. Further, the Wilcoxon test result (Table 6) shows that the results of pre-test and post-test of students in the experimental group (Z=-8.076) and the control group (Z=-6.690) at the significant level *(p)* < 0.05 are significantly different. This suggests that students' understanding of the hydrolysis concept after the intervention (post-test) is higher than before the intervention (pre-test), both in experimental and control groups. However, the abilities of students in the experimental group are better than those in the control group. Accordingly, the learning process with the POGIL in the SSI context is better than the conventional learning.

Table 5. The result of the Mann-Whitney U test based on Students' Pre-Test and Post-Test Abilities in Experimental and Control Groups ($p<0.05$)

Test	Experimental Group (N=97)	Control Group (N=93)	U	p
Pre-test	0.5026(-0.57-1.26)[a]	0.3029(-1.61-1.03)[a]	3459.000	0.005
Post-test	1.1722(-0.09-3.00)[a]	0.7052(-1.06-1.47)[a]	1723.000	0.000

Table 6. The result of the Wilcoxon test of Students' Pre-Test and Post-Test in Experimental and Control Groups ($p<0.05$)

Group	Pre-test	Post-test	Z	p*
Experimental	0.5026(-0.57-1.26)[a]	1.1722(-0.09-3.00)[a]	-8.076	0.000
Control	0.3029(-1.61-1.03)[a]	0.7052(-1.06-1.47)[a]	-6.690	0.000

The Changes in Students' Understanding Ability of the Hydrolysis Concept

From the different changes in pre- and post-test (Table 7), students in the experimental and control groups have improved their understanding of the hydrolysis concept. The experimental group's mean of pre-test and post-test is logit .51 (S.E = logit .21) and logit 1.50 (S.E = logit .32), respectively, with the mean difference of both tests is (logit .99). In contrast, the mean of pre-test and post-test of the control group gets logit .26 (S.E = logit .20) and logit .87 (S.E = logit .26), respectively, with the mean difference of pre- and post-test is logit .61. Such differences indicate different effects of interventions in the experimental and control group.

Table 7. Logit of Mean of Pre- and Post-Test Items of Experimental and Control Groups

Group	Student	Item	Mean/SE (logit)		
			Pre-test	Post-test	Pre- and Post-test Difference
Experimental	97	15	.51/(.21)	1.50/(.32)	.99
Control	93	15	.26/(.20)	.87/(.24)	.61

Description: SE = Standar Error.

If the pre-test and post-test results of the experimental group are plotted in pairs (Figure 1), so that the mean difference in the sample pre- and post-test (logit +.99) is displayed as an intercept on the horizontal axis with the plotted slope = 1, several facts obtained: First, two lines that form the upper and lower curves separate 66 students around the empirical plot line, in which the pre-test and post-test mean is not significantly different from the mean difference in the pre- and post-test in the experimental group. Second, above the curve, 23 students experience significant changes; the mean of pre- and post-test is greater than the mean difference in sample pre-test and post-test. Third, seven students do not change, and ten students have negative changes (under the curve), so that they are under the curve. Similarly, the results of pre- and post-test of the control group (Figure 2) show that 53

students are around the empirical plot line; the abilities of 25 students change significantly (greater than the mean of sample pre- and post-test (logit +.61); two students do not change; 13 students experience negative changes in abilities. The difference in the plotting of pre-test and post-test results signifies different effects of interventions in the experimental and control groups.

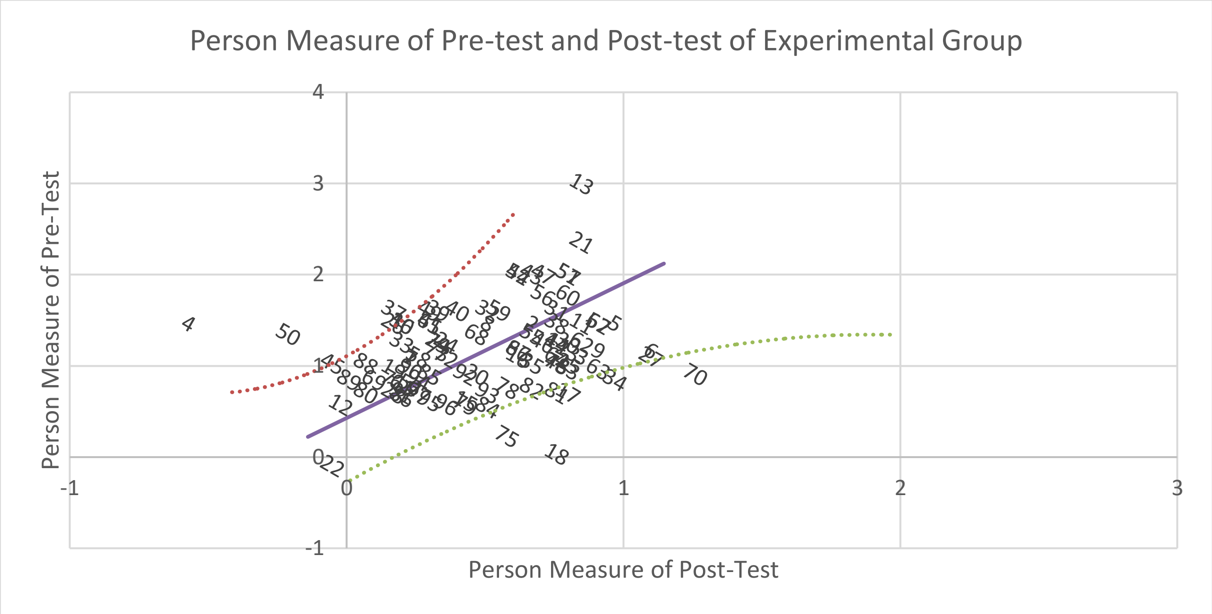

Figure 1. Scatter Plots of Person Measures in Pre- and Post-test of the Experimental Group

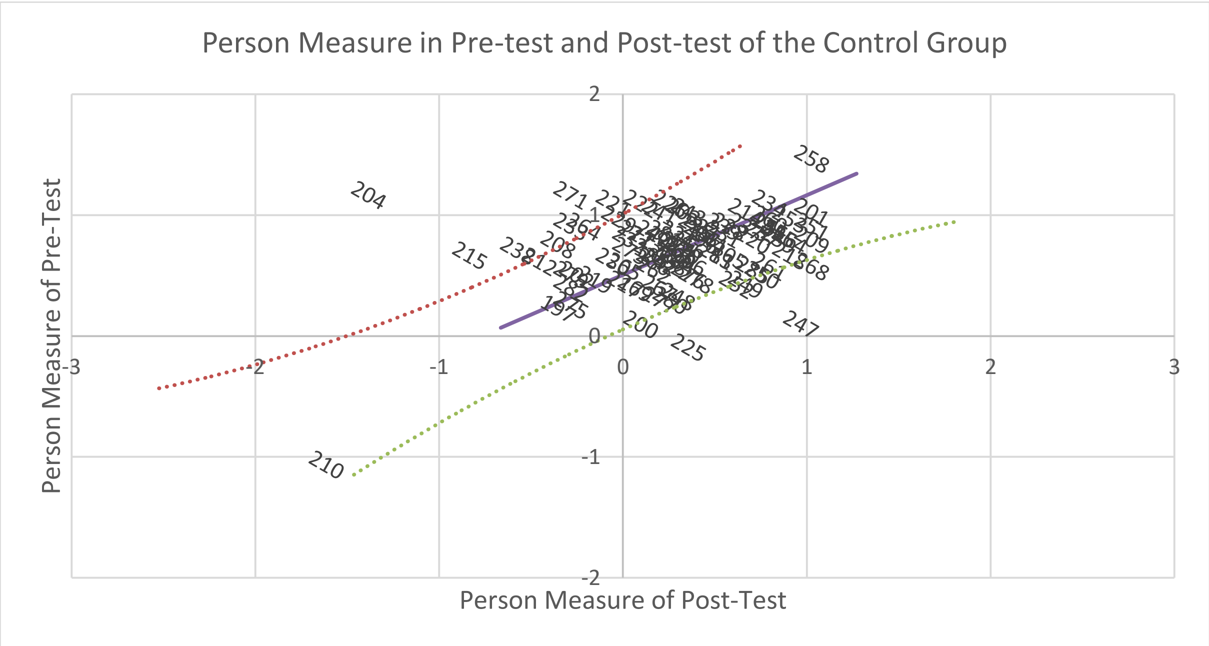

Figure 2. Scatter Plots of Person Measures in Pre- and Post-test of the Control Group

The Changes in Item Difficulty Level

Table 8 presents the results of the racking analysis in connection with the changes in item difficulty level in the pre- and post-test of experimental and control groups. It is shown that in terms of item difficulty level, the mean of pre-test of the experimental group is (logit .32), the mean of post-test is (logit -.34), and the mean difference of the pre- and post-test is (logit -.66). Moreover, the mean of pre-test of the control group is (logit .25), the mean of post-test is (logit -.25), and the mean difference of the pre- and post-test is (logit -.50). This research also finds out that seven items have significant changes in the item difficulty level in the experimental group, lower than the pre- and post-test mean difference of (logit -.66), namely item 1, 2, 5, 7, 9, and 11. Eight items with a difficulty level greater than the mean are item 3, 4, 6, 8, 12, 13, 14, and 15. Item 10 has the same difficulty level as the mean. In the control group, eight items change significantly or less than the pre- and post-test mean difference of (logit -.50), including item 2, 3, 4, 5, 9, 11, 12, and 14; five items (item 1, 6, 7, 8, 10, 13) are greater than the mean; one item (item 15) has negative changes or becomes more difficult. The most difficult item in the experimental group is item 1 (.80 logit) and the easiest one is item 14 (logit -.10). Meanwhile, the most difficult item in the control group is item 13 (logit .64), and item 3 (logit -.15) is the easiest one. These findings indicate differences in the item difficulty level changes between students taught by the POGIL in the SSI context and the conventional model.

Table 8. Data of item measures of pre- and post-test of experimental and control groups

Item	Experimental (Mean)			Control (Mean)		
	Pre-test	Post-test	Difference Pre- and Post-test	Pre-test	Post-test	Difference Pre- and Post-test

Item1	.16	-1.00	-1.16	-.06	-.76	-.7
Item2	.80	.01	-.79	.39	-.40	-.79
Item3	.20	-.63	-.43	-.15	-.83	-.68
Item4	.62	.25	-.37	.54	.02	-.52
Item5	.14	-.78	-.92	.10	-.49	-.59
Item6	.26	.22	-.04	.41	.30	-.11
Item7	.66	-.33	-.99	.33	-.06	-.39
Item8	.59	.45	-.14	.49	.47	-.02
Item9	-.04	-.85	-.81	-.08	-.93	-.85
Item10	.40	-.26	-.66	.32	-.01	-.33
Item11	.13	-.91	-1.04	.05	-.78	-.83
Item12	.33	-.23	-.56	.25	-.51	-.76
Item13	.77	.16	-.61	.64	.33	-.31
Item14	-.10	-.80	-.7	.15	-.83	-.98
Item15	.25	-.40	-.65	.39	.72	.33
Mean	.32	-.34	-.66	.25	-.25	-.50

Conceptual Changes in Students' Ability and Item Difficulty Levels

Apart from the effect of learning interventions, there are three other factors that tend to influence the changes in students’ ability and item difficulty levels, as follows: 1) guessing which happened to be correct or (lucky guess), 2) cheating, 3) carelessness. These factors can be identified from the student's item response pattern using a scalogram. For instance, the response pattern of post-test item 7 for student 353, 375, and 170 (Table 9). These three students, in the seventh and eighth row from the left, cannot understand item 12 (logit -.06) and item 10 (logit .08). Meanwhile, they can correctly explain the more difficult item, i.e., item 7 (logit .15). This situation implies a lucky guess, which in fact, these students have higher post-test abilities than the item 7 logit. Next is a cheating indication in the response pattern of student 128, 129, 134, 137, and 146. Such an indication is initially detected from the same post-test mean (logit 1.61) and item response pattern. The last one is carelessness, e.g., student 110, 118, and 139 are considered to be careless as they cannot correctly explain the easy item 4 (logit .33), yet can accurately understand item 13 (logit .46), which is harder than item 4. Moreover, they get very high post-test abilities.

Table 8. Scalogram

```
GUTTMAN SCALE OF RESPONSES:
Person |Item
       |   11 11   1 1   ID        Pre- Post-test  Pre Post   Item
       |913415207265438  Person    Mean    Mean   Difference   Response Pattern
       |---------------
   353 +666555536665554  353MFCB     .8    .97     .17    Lucky Guess
   375 +166566516133664  375MMCB   -.28    .40     .68    Guessing answer
   170 +664666446566556  170NFEB    .33   1.17     .84    accidentaly correct

   128 +666666666646555  128DFEB    .76   1.61     .85
   129 +666666666646555  129DFEB    .51   1.61    1.10    Same response pattern
   134 +666666666646555  134JFEB    .17   1.61    1.44    Cheating indication
   137 +666666666646555  137MMEB    .04   1.61    1.21
   146 +666666666646555  146NFEB    .30   1.61    1.31

   110 +666666666666566  110NFEB    .85   3.00    2.15    Response pattern
   118 +666666666666565  118RFEB    .85   2.36    1.51    “Careless”
   139 +666666665666565  139MFEB    .62   2.01    1.39
```

Negative Changes

Negative changes in conceptual understanding are detected from the changes in students' post-test logit less than the pre-test logit. For example, two students from the experimental group (E18 and E75) and the control group (C225 and C247) are taken; they have negative changes (Table 9). This means that these four students experience decreased abilities after the intervention. The pre-test item mean and the post-item mean of student E18 are (logit .76) and (logit .04), sequentially, with the mean difference of pre- and post-test arriving at (logit -.72). Moreover, the pre- and post-test item standard errors of student E18 are (logit .22) and (logit .18), respectively, with the combined standard error of logit .40. On account of the higher combined standard error than the pre- and post-test measures, the ability of student E18 in both tests is not significantly different. This also applied to student E75, C225, and C247.

Table 9. Scalogram results of student E18, E75, C225, and 247

ID Person	Test	Item Response Pattern	Mean			
		| 11 11 1 1 |913415207265438 |---------------	Item Logit	S.E* Logit	Pre- test and post-test difference	Combined S.E
E18	Pre-test	+665666636366333	.76	.22	-.72	.40
	Post-test	+666661322521161	.04	.18		
E75	Pre-test	+562664552566426	.58	.20	-.35	.38
	Post-test	+655664322323463	.23	.18		
C225	Pre-test	+616665663261613	.36	.19	-.45	.37
	Post-test	+611622566131613	-.09	.18		
C247	Pre-test	+663636666666435	.97	.25	-.87	.43
	Post-test	+563345555314133	.10	.18		

Description: S.E = Standar Error

Discussion and Conclusion

The findings show changes in students' understanding abilities of the hydrolysis concept and items after the intervention. From the pre- and post-test mean difference, the

experimental group has better positive changes than the control group [58]. In addition to the effect of the intervention, there is another factor contributing to the positive conceptual changes mentioned above, in terms of students' ability and item difficulty levels [24, 58]. The factor refers to some students who "accidentally" give a correct response pattern (in the post-test). Even so, both groups have also experienced negative changes, implying that the intervention is specifically responded by students on account of the carelessness factor or a misconception-comprising response pattern [56, 58, 77]. Regarding this, not all learning objectives of the hydrolysis concept match the approach of POGIL in the SSI context. Negative changes of the students are because they are not epistemologically involved in the learning process, particularly in the observing, measuring, and calculating stages. These activities are interrelated up to group discussions as part of the stages of conceptual formation based on empirical facts [78]. Students are expected to explain and link the concepts they have learned following their epistemological reasoning [16, 79].

Furthermore, the interpretation of changes due to pedagogical interventions is exemplified by four students (Table 8) in item 5. In the pre-test, the ability of student E18 (logit .76), student E75 (logit .58), student C225 (logit .36), and student C247 (logit .96) is greater. They also respond to item 5 (-.25 logit) accurately. However, in the post-test item 5, the response of student E18, E75, C225, and C247 is incorrect due to their decreased post-test abilities. Therefore, the pre- and post-test mean difference is lower than item 5. Why do these changes occur? Such changes are exemplified by the response pattern of student E18 in item 5. This item measures students' ability in determining the reaction of NaOCl reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$, with the estimated pH = 7 and is alkaline. The question (Q1) of this item is, "is it correct that NaOCl is alkaline?". E18 answers "correct" in the pre-test, yet responds to "incorrect" in the post-test. The question (Q2) of this item is "what is your consideration for your answer in the Q1?". Four options are provided: (a) because NaOCl is

formed of strong acids and weak bases; (b) because NaOCl is formed of weak acids and strong bases; (c) because NaOCl is formed of weak acids and weak bases; (d) because NaOCl is formed of strong acids and strong bases. In the pre-test, E18 chooses the correct answer (b), yet selects the incorrect answer (a) in the post-test that comprises misconception. Next, in the Q3 of this item, E18 chooses "very sure" in the pre-test and "not sure" in the post-test. The item 5 response pattern of E18 becomes CCC (category of scientific knowledge - SK) in the pre-test and IIU (category of lack of knowledge - LOK) in the post-test. Accordingly, the response pattern changes from CCC to IIU. The pre- and post-test mean difference of E18 (logit -.72) lower than item 5 (-.25) signifies that the error of response pattern results from misconception. This also applies to the response pattern of E75 (logit -.35), C225 (logit -.45), and C247 (logit -.87).

The misconception refers to the inability to identify the NaOCl salt hydrolysis that is formed of weak acids and strong bases. In short, the four students tend to not understand the concept of acid and base and acid-base reaction. These findings strengthen several previous studies [51, 53,54, 80]. A study on the understanding of the acid-base concept of senior high school students in Malaysia concludes that some students have little understanding of the function of detergents as the cleaning agent, the difference between strong acids and strong bases, and the treatment for soil acidity using fertilizers [53]. In the same tune, such little understanding is because they do not conceptualize acid-base strength as a property that arises from the interaction of many reaction factors [51]. Additionally, research on an alternative conception of salt hydrolysis among senior high school students contends that the concept of hydrolysis is challenging for the students [54]. They are usually able to state the acidity of a salt solution correctly, yet writing a chemical equation to explain such a phenomenon is a great challenge. Most of the alternative conceptions are identifiable, rooted in the misunderstanding of equilibrium process, acid and base, material structure and other basic

problems, student tendency to use a wrong analogy, and the lack of laboratory practice.

This research findings and elaboration of negative changes (case E18) prove the advantages of the Rasch model, specifically its potential in linking the result of changes (pre- and post-test), the item difficulty level, and the content being measured [18]. Such information solely comes from the Rasch model-based stacking and racking analysis techniques. The stacking technique provides information regarding “who has changed”; in contrast, the racking technique offers information of “what has changed” [56, 58], allowing the researchers to spell out the effect of the applied pedagogical innovation [18, 33, 34]. Although the instrument measurement result of this work is not data-rich, the analysis strength of the Rasch model can describe in detail the conceptual changes, both in the students’ ability and item difficulty levels.

Limitations and Further Studies

The primary limitation of this research is that it did not take into account the aspects of learning style, culture, and motivation that can change due to learning interventions. Future studies, therefore, can address these aspects. The present study can be continued by considering the context of a problem that closely connects with the parameter of item difficulty level. The analysis will be more interesting if it can prove that different item difficulty levels are influenced by problem contexts in each item [81]. Further studies are also expected to find an analysis technique that can integrate problem contexts, item characteristics, and item difficulty levels in a measurement model. It is assumed that different problem contexts in each item will be more likely to affect measurement results because problem contexts have conceptual linkage with items and student activities in doing experiments, measuring, interpreting data/graphs, and others. Thus, the linkages between the learning process during the intervention and conceptual changes in students’ ability and item

difficulty levels can be explained in detail; which part of the process leads the students to change their understanding related to specific ideas taught to them.

Acknowledgment

We would like to express our deep gratitude to all parties for their assistance in data processing, especially the Head of Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo.

Conflicts of interest

The authors declare no conflict of interest.

Availability of data and material

All the data are presented in the manuscript. The data supporting the findings of the article is also available in the appendix section in the end part of the manuscript.

References

1. Lu S, Bi H. Development of a measurement instrument to assess students' electrolyte conceptual understanding. Chem. Educ. Res. Pract. 2016;17(4):1030–1040. doi: 10.1039/c6rp00137h
2. Yildirir HE, Demirkol H. (2018). Identifying mental models of students for physical and chemical change. J. Balt. Sci. Educ. 2018;17(6): 986–1004. doi: 10.33225/jbse/18.17.986
3. Gette CR, Kryjevskaia M, Stetzer MR, Heron PRL. Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy. Phys. Rev. Phys. Educ. Res. 2018;14(1): 10113. doi: 10.1103/PhysRevPhysEducRes.14.010113
4. Johnstone AH. Chemical education research in Glasgow in perspective. Chem. Educ.

Res. Pract. 2006;7(2): 49–63. doi: 10.1039/b5rp90021b

5. Johnstone AH. You can't get there from here. J. Chem. Educ. 2010;87(1): 22–29. doi: 10.1021/ed800026d

6. Taber KS. Chemical misconceptions—Prevention, diagnosis, and cure. J. Chem. Educ. 2003; 80(5): 491.

7. Taber KS. Challenging misconceptions in the chemistry classroom: Resources to support teachers. Educació Química EduQ. 2009;4:13–20. doi: 10.2346/20.2003.02.27

8. Alamina JI, Etokeren IS. Effectiveness of imagination stretch teaching strategy in correcting misconceptions of students about particulate nature of matter. J. Educ. Soc. Behav. Sci. 2018;27(1):1–11. doi: 10.9734/jesbs/2018/43063

9. Yaşar IZ, İnce E, Kırbaşlar FG. 7. Class science and technology course "structure of atom" subject readiness improvement test. 2014;152:662-667. Procedia Soc. Behav. Sci. doi: 10.1016/j.sbspro.2014.09.259

10. Hoe KY, Subramaniam R. On the prevalence of alternative conceptions on acid-base chemistry among secondary students: Insights from cognitive and confidence measures. Chem. Educ. Res. Pract. 2016;17(2):263–282. doi: 10.1039/c5rp00146c

11. Aktan DC. Investigation of students' intermediate conceptual understanding levels: The case of direct current electricity concepts. Eur. J. Phys. 2013;34(1):33–43. doi: 10.1088/0143-0807/34/1/33

12. Allen M. Misconceptions in primary science. Maidenhead: Open University Press; 2014.

13. Soeharto, Csapó B, Sarimanah E, Dewi FI, Sabri T. A review of students' common misconceptions in science and their diagnostic assessment tools. J. Pendidik. IPA Indones. 2019;8(2):247–266. doi: 10.15294/jpii.v8i2.18649

14. Almuntasheri S, Gillies RM, Wright T. The effectiveness of a guided inquiry-based,

teachers' professional development programme on saudi students' understanding of density. Sci. Educ. Int. 2016:27(1); 16–39.

15. Hashweh MZ. The complexity of teaching density in middle school. Res. Sci. Technol. Educ. 2016;34(1): 1–24. doi: 10.1080/02635143.2015.1042854

16. Zvoch K, Holveck S, Porter L. Teaching for conceptual change in a density unit provided to seventh graders: A comparison of teacher- and student-centered approaches. Res. Sci. Educ. 2019;51:1395–1421. doi: 10.1007/s11165-019-09907-8

17. Grob R, Holmeier M, Labudde P. Formative assessment to support students' competences in inquiry-based science education. Interdiscip. J. Probl-based. Learn. 2017:11(2). doi: 10.7771/1541-5015.1673

18. Pentecost TC, Barbera J. Measuring learning gains in chemical education: A comparison of two methods. J. Chem. Educ. 2013;90(7):839–845. doi: 10.1021/ed400018v

19. Sumintono B. Rasch model measurements as tools in assesment for learning. In: Proceedings of the 1st International Conference on Education Innovation (ICEI 2017); 2017 October 14; Surabaya. Paris: Atlantis Press; 2018. Available from: doi: 10.2991/icei-17.2018.11

20. Micelli R, Settanni M, Vidotto G. Measuring change in training programs : An empirical illustration Measurement of change and Item Response Theory. Psychol. Sci. Q. 2008;50(3): 433–447.

21. Willoughby SD, Metz A. Exploring gender differences with different gain calculations in astronomy and biology. Am. J. Phys. 2009;77(7):651–657. doi: 10.1119/1.3133087

22. Linn RL, Slinde JA. The determination of the significance of change between pre- and posttesting periods. Rev. Educ Res. 1977;47(1):121–150. doi: 10.3102/00346543047001121

23. He P, Liu X, Zheng C, Jia M. Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. Chem. Educ. Res. Pract. 2016;17(2):381–393. doi: 10.1039/C6RP00004E

24. Sumintono B, Widhiarso W. Aplikasi pemodelan Rasch pada assessment pendidikan [Application of Rasch modeling in educational assessment]. Bandung: Penerbit Trim Komunikata; 2015. https://www.researchgate.net/publication/282673464%0AAplikasi

25. Boone WJ, Yale MS, Staver JR. Rasch analysis in the human sciences. Dordrecht: Springer; 2014. doi: 10.1007/978-94-007-6857-4

26. Liu X. Developing measurement instruments for science education research. In: Fraser B, Tobin KG, McRobbie CJ, editors. Second International Handbook of Science Education. Netherlands: Springer; 2012. Pp. 651-665.

27. Linacre JM. 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. US: Winsteps; 2020.

28. Sabah S, Hammouri H, Akour M. Validation of a scale of attitudes toward science across countries using Rasch model: Findings from TIMSS. J. Balt. Sci. Educ. 2013;12(5):692–703.

29. Hadenfeldt JC, Bernholt S, Liu X, Neumann K, Parchmann I. Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter. J. Chem Educ. 2013; 90(12):1602–1608. doi: 10.1021/ed3006192

30. Herrmann-Abell CF, DeBoer GE. Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. Chem. Educ. Res. Pract. 2011;2(2):184–192. doi: 10.1039/c1rp90023d

31. Herrmann-Abell CF, Deboer GE. Using Rasch modeling and option probability curves to diagnose students' misconceptions. In: the 2016 AERA Annual Meeting; 2016 April

8-12; Washington DC. USA: American Educational Research Assossiation; 2016. Available from: https://www.aaas.org/sites/default/files/Project2061_CHA-AERA%20energy%20paper%204-7-16.pdf

32. Laliyo LAR, Tangio JS, Sumintono B, Jahja M, Panigoro C. Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter. J. Balt. Sci. Educ. 2020;19(5). doi: 10.33225/jbse/20.19.824

33. Davidowitz B, Potgieter M. Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. Int J. Sci. Educ. 2016;38(9):1483–1503. doi: 10.1080/09500693.2016.1196843

34. Park M, Liu X. An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. Res. Sci. Educ. 2019. doi: 10.1007/s11165-019-9819-y

35. Treagust DF, Qureshi SS, Vishnumolakala VR, Ojeil J, Mocerino M, Southam DC. Process-oriented guided inquiry learning (POGIL) as a culturally relevant pedagogy (CRP) in Qatar: A perspective from grade 10 chemistry classes. Res. Sci Educ. 2018:1–19. doi: 10.1007/s11165-018-9712-0

36. Kinslow AT, Sadler TD, Nguyen HT. Socio-scientific reasoning and environmental literacy in a field-based ecology class. Environ. Educ. Res. 2018;4622:1–23. doi: 10.1080/13504622.2018.1442418

37. Owens DC, Sadler TD, Friedrichsen P. Teaching practices for enactment of socio-scientific issues instruction: An instrumental case study of an experienced biology teacher. Res. Sci. Educ. 2019. doi: 10.1007/s11165-018-9799-3

38. Bruder R, Prescott A. Research evidence on the benefits of IBL. ZDM Int. J. Math.

Educ. 2013;45(6):811–822. doi: 10.1007/s11858-013-0542-2

39. Zeidler DL. Socioscientific issues as a curriculum emphasis: theory, research and practice. In: Lederman NG, Abell SK editor. Handbook of research on science education. New York: Routledge; 2014. pp. 697–726.

40. Espeja AG, Lagarón DC. Socio-scientific Issues (SSI) in initial training of primary school teachers: Pre-service teachers' conceptualization of SSI and appreciation of the value of teaching SSI. Procedia Soc. Behav. Sci. 2015;196:80–88. doi: 10.1016/j.sbspro.2015.07.015

41. National Research Council. A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington DC: The National Academies Press; 2012. doi: 10.17226/13165

42. Sadler TD, Zeidler DL. Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. J. Res. Sci. Teach. 2009;46(8):909–921. doi: 10.1002/tea.20327

43. Lederman NG, Lederman JS, & Antink A. Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. Int. J. Educ. Math. Sci. Technol. 2013;1(3):138–147. doi: 10.18404/ijemst.19784

44. Cooke AN, Fielding KS, Louis WR. Environmentally active people: The role of autonomy, relatedness, competence and self-determined motivation. Environ. Educ. Res. 2016;22(5):631–657. doi: 10.1080/13504622.2015.1054262

45. Sadler TD. Moral and ethical dimensions of socioscientific decision-making as integral components of scientific literacy. Sci. Educ. 2004;13(1): 39–48.

46. DeBoer GE. Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. J. Res. Sci. Teach. 2000;37(6):582–601. doi: 10.1002/1098-2736(200008)37:6<582::AID-

TEA5>3.0.CO;2-L

47. Sadler TD. Socio-scientific issues-based education: What we know about science education in the context of SSI. In: Sadler TD, editor. Socio-Scientific Issues in the Classroom: Teaching, Learning and Research. Dordrecht: Springer; 2011. pp. 355–369. doi: 10.1007/978-94-007-1159-4_16.

48. Gräber W, Nentwig P, Becker HJ, Sumfleth E, Pitton A, Wollweber K, Jorde D. Scientific literacy: From theory to practice. Res. Sci. Educ. Past. Present. Future. 2005;1996:61–70. doi: 10.1007/0-306-47639-8_6

49. Grooms J. A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction. J. Chem. Educ. 2020. doi: 10.1021/acs.jchemed.0c00026

50. Presley ML, Sickel AJ, Muslu N, Merle D. A framework for socio-scientific issues based education. Sci. Educ. 2013;22:26–32.

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

52. Seçken N. Identifying student's misconceptions about SALT. Procedia Soc. Behav. Sci. 2010;2(2):234–245. doi: 10.1016/j.sbspro.2010.03.004

53. Damanhuri MIM., Treagust DF, Won M, Chandrasegaran AL. High school students' understanding of acid-base concepts: An ongoing challenge for teachers. Int. J. Environ. Sci. Educ. 2016;11(1):9–27. doi: 10.12973/ijese.2015.284a

54. Orwat K, Bernard P, Migdał-Mikuli A. Alternative conceptions of common salt hydrolysis among upper-secondary school students. J. Balt. Sci. Educ. 2017;16(1): 64–76.

55. Creswell JW. Educational research: Planning, conducting, and evaluating quantitative

and qualitative research. 4th ed. New York: Pearson Education, Inc; 2012.

56. Wright BD. Rack and stack: Time 1 vs. time 2 or pre-test vs. post-test. Rasch Meas. Trans. 2003;17(1):905–906.

57. Anselmi P, Vidotto G, Bettinardi O, Bertolotti G. Measurement of change in health status with Rasch models. Health Qual. Life Outcomes. 2015;13(1):1–7. doi: 10.1186/s12955-014-0197-x

58. Ling M, Pang V, Ompok CC. Measuring change in early mathematics ability of children who learn using games: stacked analysis in rasch measurement. In: Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings, 2016 July 30-August 3; Singapore. Singapore: Springer; 2018. Available from: https://doi.org/10.1007/978-981-10-8138-5

59. Taber KS. Ethical considerations of chemistry education research involving “human subjects”. Chem. Educ. Res. Pract. 2014;15(2):109–113. doi: 10.1039/c4rp90003k

60. Aldresti F, Rahayu S, Fajaroh F. The influence of inquiry-based chemistry learning with the contex of socio-scientific issues on high school students’ scientific explanation skills. J. Pendidik. IPA. 2019;23(2). doi:10.18269/jpmipa.v23i2.XXX

61. Mitarlis, Ibnu S, Rahayu S, Sutrisno. The effectiveness of new inquiry-based learning (NIBL) for improving multiple higher-order thinking skills (M-HOTS) of prospective chemistry teachers. Eur. J. Educ. Res. 2020;9(3):1309–1325. doi: 10.12973/eu-jer.9.3.1309

62. Pedaste M, Mäeots M, Siiman LA, de-Jong T, Van-Riesen SAN, Kamp ET, Manoli CC, Zacharia ZC, Tsourlidaki, E. Phases of inquiry-based learning: Definitions and the inquiry cycle. Educ. Res. Rev. 2015;14:47–61. doi: 10.1016/j.edurev.2015.02.003

63. Romine WL, Sadler TD. Measuring changes in interest in science and technology at the college level in response to two instructional interventions. Res. Sci. Educ.

2016;46(3):309–327. doi: 10.1007/s11165-014-9452-8

64. Wilson M. Constructing measures: an item response modeling approach. USA: Lawrence Erlbaum Associates, Inc; 2005. doi: doi: 10.4324/9781410611697

65. Wilson M. Cognitive diagnosis using item response models. Zeits Für Psych/J. Psych. 2008;216(2):74–88. doi: 10.1027/0044-3409.216.2.74

66. Wilson M. Measuring progressions: Assessment structures underlying a learning progression. J. Res. Sci. Teach. 2009;46(6):716–730. doi: 10.1002/tea.20318

67. Laliyo LAR. Mendiagnosis sifat perubahan konseptual siswa: Penerapan teknik analisis stacking dan racking Rasch Model [Diagnosing the nature of students' conceptual change: Application of stacking and racking analysis techniques Rasch Model]. Yogyakarta: Penerbit Deepublish; 2021.

68. Chittleborough G, Treagust D. Correct interpretation of chemical diagrams requires transforming from one level of representation to another. Res. Sci. Educ. 2008;38(4): 463–482. doi: 10.1007/s11165-007-9059-4

69. Treagust DF. Development and use of diagnostic tests to evaluate students' misconceptions in science. Int. J. Sci. Educ. 1988;10(2):159–169. doi: 10.1080/0950069880100204

70. Arslan HO, Cigdemoglu C, Moseley C. A Three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. Int. J. Sci. Educ. 2012;34(11):1667–1686. doi: 10.1080/09500693.2012.680618

71. Hasan S, Bagayoko D, Kelley EL. Misconceptions and the certainty of response index (CRI). Phys. Educ. 1999;34(5):294–299. doi: 10.1088/0031-9120/34/5/304

72. Habiddin H, Page EM. Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). Indones. J. Chem. 2019;19(3):720–736. doi:

10.22146/ijc.39218

73. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences. 3rd ed. New York: Routledge Taylor & Francis Group; 2015. doi: 10.1088/1751-8113/44/8/085201

74. Herrmann-abell CF, Flanagan JC, Roseman JE. Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: Using student measures to evaluate the promise of the intervention. In: Proceeding of the 2013 NARST Annual International Conference. 2013 April 6-9; Puerto Rico. Puerto Rico: NARST; 2013. Available from: https://www.aaas.org/sites/default/files/project2061/files/NARST2013-Flanagan-Paper5-Usingteachermeasurestoe.pdf

75. Fisher WP. Rating scale instrument quality criteria. Rasch Meas. Trans. 2007;21(1): 1095. www.rasch.org/rmt/rmt211m.htm

76. Adams WK, Wieman CE. Development and validation of instruments to measure learning of expert-like thinking. Int. J. Sci. Educ. 2011;33(9):1289–1312. doi: 10.1080/09500693.2010.512369

77. Sumintono B, Widhiarso W. Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial [Application of Rasch Model for Social Science Studies]. In Trim B, editor.). Bandung: Trim Komunikata Publishing House; 2014. https://www.researchgate.net/publication/268688933%0AAplikasi

78. Rodriguez JG, Hunter KH, Scharlott LJ, Becker NM. (2020). A Review of Research on Process Oriented Guided Inquiry Learning: Implications for Research and Practice. J. Chem. Educ. 2020;97(10):3506–3520. doi: https://doi.org/10.1021/acs.jchemed.0c00355

79. Hancock TS, Friedrichsen PJ, Kinslow AT, Sadler TD. Selecting socio-scientific issues

for teaching: A grounded theory study of how science teachers collaboratively design SSI-based curricula. Sci. Educ. 2019;28(6–7):639–667. doi: 10.1007/s11191-019-00065-x

80. Ültay N, Çalik M. A comparison of different teaching designs of "acids and bases" subject. Eurasia J. Math. Sci. Technol. Educ. 2016;12(1):57–86. doi: 10.12973/eurasia.2016.1422a

81. Hartig J, Frey A, Nold G, Klieme E. An application of explanatory item response modeling for model-based proficiency scaling. Educ. Psychol. Meas. 2012;72(4):665–686. doi: 10.1177/0013164411430707

Appendix 1. Standardized Residual Variance in Eigen value Units.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	22.7067	100.0%		100.0%
Raw variance explained by measures	=	7.7067	33.9%		35.9%
Raw variance explained by persons	=	2.7733	12.2%		12.9%
Raw Variance explained by items	=	4.9334	21.7%		23.0%
Raw unexplained variance (total)	=	15.0000	66.1%	100.0%	64.1%
Unexplned variance in 1st contrast	=	2.0698	9.1%	13.8%	
Unexplned variance in 2nd contrast	=	1.5312	6.7%	10.2%	
Unexplned variance in 3rd contrast	=	1.3696	6.0%	9.1%	
Unexplned variance in 4th contrast	=	1.3124	5.8%	8.7%	
Unexplned variance in 5th contrast	=	1.1945	5.3%	8.0%	

Appendix 2. Summary of Category Structure.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	317	6	-.18	-.20	1.06	1.08	NONE	(-1.73)	1
2	2	190	3	.10	.09	1.03	1.18	.46	-.77	2
3	3	963	17	.33	.31	1.02	.93	-1.43	-.22	3
4	4	542	10	.56	.52	1.02	.97	.98	.21	4
5	5	1262	22	.62	.73	1.27	.99	-.22	.74	5
6	6	2425	43	1.02	.98	.97	1.04	.20	(1.76)	6
MISSING		1	0	-.30						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

ACCEPTANCE

Gmail Lukman Laliyo <lukman.laliyo019@gmail.com>

Decision on submission to Heliyon

Heliyon <em@editorialmanager.com> Tue, Mar 15, 2022 at 3:21 PM
Reply-To: Heliyon <info@heliyon.com>
To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Manuscript Number: HELIYON-D-21-05440R3
Title: Measuring Changes in Hydrolysis Concept of Students Taught by Inquiry Model: Stacking and Racking Analysis Techniques in Rasch Model
Journal: Heliyon

Dear Dr. Laliyo,

Thank you for submitting your manuscript to Heliyon.

I am pleased to inform you that your manuscript has been accepted for publication.

Your accepted manuscript will now be transferred to our production department. We will create a proof which you will be asked to check, and you will also be asked to complete a number of online forms required for publication. If we need additional information from you during the production process, we will contact you directly.

We appreciate and value your contribution to Heliyon. We regularly invite authors of recently published manuscript to participate in the peer review process. If you were not already part of the journal's reviewer pool, you have now been added to it. We look forward to your continued participation in our journal, and we hope you will consider us again for future submissions.

Kind regards,
Mengpei Yan
Editorial Assistant
Heliyon

Embargo
Embargos are not automatically set for papers published in Heliyon. Papers appear online a few days after acceptance. To request a media embargo and/or publication on a specific date to assist an institutional press release, please reach out to the Heliyon team (info@heliyon.com) as soon as possible and we will do our best to accommodate your request.

Editorial Office comments:

Ready to accept.

More information and support

FAQ: When and how will I receive the proofs of my article?
https://service.elsevier.com/app/answers/detail/a_id/6007/p/10592/supporthub/publishing/related/

You will find information relevant for you as an author on Elsevier’s Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live

chat and email

GALLEY PROOF

Gmail

Lukman Laliyo <lukman.laliyo019@gmail.com>

Corrections received - [HLY_9126]

optteam@elsevierproofcentral.com <optteam@elsevierproofcentral.com> Thu, Mar 24, 2022 at 1:37 PM
To: lukman.laliyo019@gmail.com

This is an automatically generated message. Please do not reply because this mailbox is not monitored.

Dear Dr. Lukman Abdul Rauf Laliyo,

Thank you very much for using the Proof Central application for your article "Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in Rasch model" in the journal "HLY"

All your corrections have been saved in our system. The PDF summary of your corrections, generated from Proof Central, can be downloaded from the following site for your reference:
https://pcv3-elsevier-live.s3.amazonaws.com/a713aa88a74196084f34a9d673d367/HLY_9126_edit_report.pdf

To track the status of your article throughout the publication process, please use our article tracking service:

http://authors.elsevier.com/TrackPaper.html?trk_article=HLY9126&trk_surname=

For help with article tracking:
http://support.elsevier.com/app/answers/detail/a_id/90

Kindly note that now we have received your corrections, your article is considered finalised and further amendments are no longer possible.

For further assistance, please visit our customer support site at http://support.elsevier.com. Here you can search for solutions on a range of topics. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

Yours sincerely,
Elsevier Proof Central team

Gmail

Lukman Laliyo <lukman.laliyo019@gmail.com>

Proofs of [HLY_9126]

corrections.esch@elsevier.tnq.co.in <corrections.esch@elsevier.tnq.co.in>
To: lukman.laliyo019@gmail.com

Thu, Mar 17, 2022 at 6:36 PM

PLEASE DO NOT ALTER THE SUBJECT LINE OF THIS E-MAIL ON REPLY

Dear Dr. Lukman Abdul Rauf Laliyo,

Thank you for publishing with HELIYON. We are pleased to inform you that the proof for your upcoming publication is ready for review via the link below. You will find instructions on the start page on how to make corrections directly on-screen or through PDF.

https://elsevier.proofcentral.com/en-us/landing-page.html?token=afdc80df483d5bb1550d85688cf31e

Please open this hyperlink using one of the following browser versions:

- Google Chrome 68+
- Mozilla Firefox 61+
- Mac OS Safari 11+
- Microsoft Edge 79+

We ask you to check that you are satisfied with the accuracy of the copy-editing, and with the completeness and correctness of the text, tables and figures. To assist you with this, copy-editing changes have been highlighted.

You can save and return to your article at any time during the correction process. Once you make corrections and hit the SUBMIT button you can no longer make further corrections.

Please review the proof and submit any corrections within 48 hours to help us publish your article as quickly and accurately as possible.

We very much look forward to your response.

Yours sincerely,

Elsevier

E-mail: corrections.esch@elsevier.tnq.co.in

For further assistance, please visit our customer support site at https://service.elsevier.com. Here you can search for solutions on a range of topics. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

Heliyon xxx (xxxx) e09126

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.elsevier.com/locate/heliyon

Research article

Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in Rasch model

Lukman Abdul Rauf Laliyo [a,*], Bambang Sumintono [b], Citra Panigoro [c]

[a] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo, 96128, Indonesia
[b] Faculty of Education, Universitas Islam Internasional Indonesia, 16416, Indonesia
[c] Department of Aquatic Resource Management, Faculty of Fisheries and Marine Science, Universitas Negeri Gorontalo, 96128, Indonesia

ARTICLE INFO

Keywords:
Stacking
Racking
Rasch model
Hydrolysis conceptual changes
Inquiry model

ABSTRACT

This research aimed to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the process-oriented guided inquiry learning (POGIL) model in the context of socio-scientific issues (SSI) with the pretest-posttest control group design. Such techniques were based on a person- and item-centered statistic to determine how students and items changed during interventions. Eleventh-grade students in one of the top-ranked senior high schools in the eastern part of Indonesia were involved as the participants. They provided written responses (pre- and post-test) to 15 three-tier multiple-choice items. Their responses were assessed through a rubric that combines diagnostic measurement and certainty of response index. Moreover, the data were analyzed following the Rasch Partial Credit Model, using the WINSTEPS 4.5.5 software. The results suggested that students in the experimental group taught by the POGIL approach in the SSI context had better positive conceptual changes than those in the control class learning with a conventional approach. Along with the intervention effect, in certain cases, it was found that positive conceptual changes were possibly due to student guessing, which happened to be correct (lucky guess), and cheating. In other cases, students who experienced negative conceptual changes may respond incorrectly due to carelessness, the boredom of problem-solving, or misconception. Such findings have also proven that some students tend to give specific responses after the intervention in certain items, indicating that not all students fit the intervention. Besides, stacking and racking analyses are highly significant in detailing every change in students' abilities, item difficulty levels, and learning progress.

1. Introduction

Central to defining the quality of pedagogical innovation in science classes is conceptual changes. The changes refer to how ideas or conceptions the students understand according to their ways of thinking [1, 2] become scientifically accurate [3]. It is because such ideas generally comprise misconceptions [4, 5, 6, 7], are not in accordance with scientific concepts [8, 9], tend to be resistant [10], changeable and varied [11], so that they should be improved if the correct conceptual understanding is to be taught [12, 13].

Some studies have been conducted on learning innovation testing to form an accurate and scientific conceptual understanding of the students, e.g., inquiry-based learning. This model presents conceptual conflicts and participatory experiments to facilitate conceptual changes [14, 15, 16]. Conceptual understanding-based learning involves various strategies in identifying and analyzing students' comprehension so that the investigation process can be designed to lead them to a more accurate and scientific conception [16, 17]. This research relied on a quasi-experimental design that assessed students' pre-test and post-test, evaluated the changes in performances for testing significant differences. This type of testing informs the researcher about the presence of an effect, but does not provide detailed information on the level and trait of the changes [18]. What if the researcher is willing to compare the extent to which the pre- and post-test change (differences in learning outcomes) and interpret the changes (the reasoning why those changes occur) in terms of content? This is a core question regarding the changes in some latent traits or changes in traits measured after the intervention. In most studies, interpreting the changes in pre-test and post-test tends to be limited to identifying whether or not an effect prevails.

Pre- and post-test changes should be given in detail regarding the students' understanding ability and item difficulty levels. However, this has not been much revealed due to the limitations of its measurement

* Corresponding author.
E-mail address: lukman.laliyo019@gmail.com (L.A. Rauf Laliyo).

https://doi.org/10.1016/j.heliyon.2022.e09126
Received 7 July 2021; Received in revised form 26 January 2022; Accepted 15 March 2022

techniques and analyses and has not been the main focus in chemistry education research to date. One reason for this issue is the debate in the psychometric community regarding the ability to measure changes accurately [18]. This debate questions the use of raw scores in the conventional psychometric analysis, which largely contains measurement biases [19], as follows: 1) the difference in pre- and post-test scores will be negatively correlated with the pre-test score, especially for students with low pre-test scores [18,20]; 2) the difference in pre- and post-test scores shows low test reliability [21]; 3) low measurement properties due to different scales [22].

Raw scores are not final data, so that they do not have a great deal of information for drawing conclusions [23, 24]. Around the 1950s, Dr. Georg Rasch, a mathematician from Denmark, introduced the formulation of the Rasch measurement model [24]. The model has been widely applied to analyze various types of data, e.g., dichotomous, polytomous, multi-rating, and multi-rater data. In the mid-2000s, the Rasch model was used as a probabilistic-based psychometric measurement that went beyond the use of raw scores [25, 26], and was used to overcome the limitations of conventional psychometric measurement [19, 27]. Its analyses, including item fit, PCA (Principal Component Analysis), and Wright map, are commonly used for international test analyses, namely TIMSS and PISA [28].

In chemistry education research, the Rasch model has been relied on to evaluate learning understanding and progress [29], to diagnose students' preconceptions [1], misconceptions [13, 30, 31, 32], link the measurement of content knowledge with pedagogical content knowledge [33], and investigate item difficulty patterns [13, 34]. Even so, studies on the Rasch model to reveal the chemistry conceptual changes in students' understanding and item difficulty levels are relatively hard to find as of today. The present study aims to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the POGIL approach in the context of SSI and students who learn conventionally. Such techniques are based on a person- and item-centered statistic to estimate how students and items change during the intervention.

POGIL is a student-centered learning strategy that teaches content or process skills. The philosophical foundation of POGIL is the involvement of an interactive process of careful thinking, discussing ideas, perfecting understanding, practicing skills, reflecting progress, and evaluating performances [35]. POGIL is able to lead the process of designing a participatory experiment that presents a conceptual conflict as a strategy to encourage students to form an accurate concept [14]. Therefore, POGIL intervention is more likely to be potential in driving epistemological understanding and reasoning [36], making students have opportunities to change their conceptions to be more accurate and scientific [16]. Nevertheless, it is also worth noting that some students potentially have misconceptions resistant to changes [3].

SSI functions as a learning context through the integration of social problems that students are familiar with. It also has a conceptual connection with salt hydrolysis [37, 38], and its resolution requires many perspectives [39], including the dimension of moral and ethical evaluation of students [40]. The SSI context is a socio-scientific phenomenon that the students should explain based on their conceptual viewpoints. It encourages them to actively get involved in grasping problems [41], developing and utilizing their knowledge [42], improving their critical thinking [43], and being able to scientifically describe the discussed socio-scientific phenomenon [36, 44, 45]. For such reasons, the integration of SSI can build up students' scientific literacy [39, 46, 47]. In the end, this integration enables the learning process to be more significant in enhancing students' understanding [45, 48]. Besides, they are skilled in negotiating the social aspect of the studied phenomenon [49, 50]. For instance, the issues of global warming, climate change, and pollution [36].

Salt hydrolysis is a learning topic in high school that is strongly related to SSI. Students with a good understanding of hydrolysis will manage to clarify scientifically why detergents, bleaching agents (NaOCl), and fertilizers can pollute the environment. Despite this, the linkage of this issue as the problem in learning hydrolysis is rarely carried out. The learning process is more emphasized on mastering theoretical concepts [36]. As a consequence, students find it challenging to use their hydrolysis understanding to explain socio-scientific phenomena around them [37]. This challenge is on account of their misconceptions regarding acid-base reaction [51], making them unable to elaborate the concept of salt hydrolysis [52] and determine acid and base strength [53]. In addition, they are struggling with correctly explaining the dissolving process and the reaction of ionic compounds with water, writing down chemical equations, and having different interpretations of the dissolving process mentioned earlier [54]. On this ground, it is essential to reveal how the hydrolysis concept changes if intervened with the POGIL approach in the SSI context, through the following specific questions: (1) is there a significant hydrolysis conceptual change of the students after the learning process in experimental and control groups? (2) if compared, how is the hydrolysis conceptual change through the intervention of POGIL in the SSI context and conventional learning? (3) in addition to intervention, is there any other factor that also contributes to the students' hydrolysis conceptual changes?

2. Method of study

This study relied on a quantitative approach with a quasi-experimental and pretest-posttest control group design [55] by comparing the extent to which the hydrolysis concept changes after the intervention. Researchers carried out the learning process for 12 meetings, gave tests, and collected data on the results of the intervention and measurement.

The changes of students and items were analyzed using the stacking and racking techniques in the Rasch model [56]. As standard techniques, racking and stacking were introduced by Benjamin Wright to measure the extent to which conceptual understanding of students and items change before and after interventions [57]. The referred changes are cases (item and student levels) caused by the learning intervention and can be diagnosed based on the estimated changes.

In regards to students' understanding, the measurement was to identify students who had specific hydrolysis conceptual changes in responding to the learning intervention. In terms of items, the measurement was done to identify which items had special characteristics and been understood by students differently during the learning intervention [57]. Thus, the scientific inquiry approach might not be suitable for some students, or some items might be too hard after the intervention. This insightful information is immensely helpful for researchers and education practitioners, especially in evaluating the weaknesses of pedagogical innovations being applied and devising learning strategies that meet students' needs in learning [58].

2.1. Participants

Eleventh-grade students aged 16–17 years in one of the senior high schools in the eastern part of Indonesia were involved as the sample. This top-ranked school gets an “A” accreditation (excellent) from the National Accreditation Board for High School. The sample was determined by convenience sampling in six randomly assigned classes. Three classes (N = 97) were experimental groups that applied the POGIL model in the SSI context. The other three classes (N = 93), as control groups, applied conventional learning without the SSI context. The same teacher taught these classes following the Curriculum 2013 of Chemistry Subject (revised in 2016). There was no special classroom for learning the concept of hydrolysis, i.e., taking up the regular learning process at school. Before learning the hydrolysis concept, the students had previously learned the concept of acid and base to understand the concept of salt hydrolysis way better. Research permission was obtained from the government and school administrators. In accordance with principles of research ethics, research purpose and procedures were informed to all the students being involved and that they were voluntarily participating. Additionally, their information is confidential and only used for science development [59].

2.2. Learning implementation

Students in the experimental group studied employing the process-oriented guided inquiry learning (POGIL) in the SSI context [35]. Meanwhile, in the control class, the learning process was performed conventionally; the teacher facilitated learning initiatives. The learning process focused more on content mastery and problem-solving practice. Applying the POGIL model in the SSI context highlights teacher assistance to guide the students to prepare their conceptual understanding based on epistemological reasoning they get from experiments, discussions, and collaborations [49, 60]. Researchers carried out the learning process for eight weeks to apply the intervention to the sample, gave tests, collected data on the results of the intervention and measurement. The first three weeks were the preparation stages when researchers and the teacher shared perceptions, and asked the teacher to perform a learning simulation under the scenario, including different assistance techniques in leading the students to conduct experiments, and to ask analytical questions. The pre-test was carried out in the third week. Further, the learning implementation was done for four weeks, and the post-test was executed in the eighth week.

The learning stages with POGIL in the SSI context consist of orientation, exploration, concept formation, application, and closing. During the orientation stage, the teacher presented familiar contextual phenomena related to the concept of hydrolysis. The teacher asked initial questions to provoke curiosity and arouse motivation and interest of the students. While watching the video, had the students responded and explained the relationship between the phenomena and acids and bases, hydrolysis, and buffers. In the exploration stage, the teacher developed analytical questions with data, images, and multiple video clips to give perspectives on learning objectives and to delve into the concept that had been and would be learned. Next, the teacher assisted the students in doing experiments guided by a worksheet, and at the same time, asked analytical questions to lead them and strengthen their conceptual understanding. In the concept formation stage, the teacher asked students to build their conceptual understanding based on the exploration results, accompanied by critical and fundamental questions to guide students in building a conceptual understanding of the salt hydrolysis and buffer solution.

Following the formation stage was the application stage when the teacher presented contextual problems in the SSI context, particularly those comprising social problems in society, that closely linked with the understanding of salt hydrolysis and buffer solution concepts. Such problems included 1) the use of bleaching agents (detergents), 2) the functions of alum $KAl(SO_4)_2{\cdot}12H_2O$ for water purification, 3) the harmful effects of detergent waste, 4) the beneficial and harmful effects of artificial fertilizer $(NH_4)_2SO_4$ for soil fertility, and 5) the harmful effects of monosodium glutamate (MSG) for health. In this stage, the teacher guided the students through collaborative discussions and critical questions, intending to give them perspectives on SSI phenomena and encourage them to collect information and do experiments following student activity sheets. Thereupon, the students had presentation and discussion sessions, during which they reported their experiment results and drew conclusions [61, 62]. The teacher asked them to describe the possible problems and solutions from their understanding of the studied concepts. This enabled the students to form their conceptual understanding that is closely related to contexts; the learning process was from contextual to abstract [37, 63]. From such a condition, the teacher led the students to apply their knowledge in different contexts and situations and solve problems. The final stage was closing or teacher assistance in guiding the students to explain the conclusion and reflection on the learning process as the end of the learning activities.

2.3. Instrument

Table 1 displays 15 items of diagnostic three-tier multiple choice test to measure students' hydrolysis conceptual understanding. The test was constructed following the Competence Standard of 2013 Chemistry Curriculum of Senior High School under Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 37 of 2018. The procedures of developing the instrument followed the recommendation by [64, 65, 66].

Each item was designed in three questions (Q1, Q2, Q3) that integrated diagnostic [68, 69] and summative measurements [10] and cer-

Table 1
Conceptual map of hydrolysis concept understanding [67].

Problem Context	Item	Conceptual Understanding	Ability Level
Bleaching agents are formed of weak acid HOCl and strong base NaOH. Sodium hypochlorite salt (NaOCl) is reactive and dissolves the dye. In the water, the ion OCl^- will be hydrolyzed to HOCl and OH^- .	1	Balancing the salt (NaOCl) hydrolysis reaction in the water	2 Level 3: Students are able to calculate the pH of the hydrolyzed salt solution.
	2	Stating the partial hydrolysis reaction: $NaOCl \rightarrow Na^+ + OCl^-$	2
	3	Determining corrosive alkali of sodium hypochlorite salt (NaOCl)	1
	4	Calculating the pH of hydrolysis of sodium hypochlorite salt (NaOCl) with NaOCl = 0.1 M; Ka = 10^{-5})	3 Level 2: Students are able to determine the hydrolysis reaction from different types of salt
	5	Determining the property of NaOCl, in the reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$	2
	6	Calculating the pH of sodium hypochlorite salt (NaOCl) that comes from a mixture of HOCl and NaOH (partially hydrolyzed), if the Ka HOCl is 10^{-5} and there is an increase in the pH of the solution mixture.	3 Level 1: Students are able to analyze the properties of the hydrolyzed salt
Water purification with alum $KAl(SO_4)_2{\cdot}12H_2O$ is the concept of salt hydrolysis, formed of H_2SO_4 and $Al(OH)_3$.	7	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water	1
	8	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water that is partially hydrolyzed by the Al^{3+} ion	1
The sodium tripolyphosphate (STPP) in detergents can pollute the environment, a eutrophication process.	9	Determining the properties of detergent solution causing eutrophication	1
	10	Determining the properties of detergent solution (sodium tripolyphosphate salt) that is partially hydrolyzed	1
	11	Determining the impact of the disposal of detergent waste on the environment	2
ZA fertilizer $(NH4)_2SO_4$ is an acidic salt.	12	Determining the properties of ammonium sulfate salt $(NH4)_2SO_4$	1
	13	Stating the equation of $(NH_4)_2SO_4$ reaction in the water, partially hydrolyzed	2
Monosodium glutamate ($C_5H_8NO_4Na$) is L-glutamic acid salt, adversely impactful on human health	14	Students' attitude towards the use of monosodium glutamate ($C_5H_8NO_4Na$)	2
	15	Determining the properties of monosodium glutamate salt ($C_5H_8NO_4Na$)	1

tainty of response index (CRI) [70, 71]. Students' responses to items (Q1, Q2, Q3) were evaluated based on the rubric (Table 2). For example, students' responses to items were as follows: Q1, Q2 “correct”, and Q3 “very sure” under the code CCC. Such a code indicated that students' conceptual understanding was in level 6, category of Scientific Knowledge (SK). On the other hand, if the response patterns in Q1, Q2 “incorrect” and Q3 “not sure”, the code would be IIU, implying that students' conceptual understanding was in the category of Lack of Knowledge (LOK), or level 1. This instrument had been validated from the aspects of item conformity with the construct variable and language. The validity results by three experts were stated under Fleiss’ kappa (K = .96), meaning that the experts agreed that the item validity was categorized good.

2.4. Data collection and analysis

Before the intervention, this research underwent pre-test data collection; whereas, the post-test data collection was done after the intervention. The construction of pre- and post-test items was the same. Students wrote down their responses on the provided answer sheet. Both tests were supervised by teachers in the school. The students must work on all items according to the allocated time (45 min). The instrument was immediately collected and should have the same number as the total participants.

The pre- and post-test measurement data were still ordinal data. The Rasch Partial Credit Model with WINSTEPS 4.5.5 software [27, 73] was used to convert ordinal data into interval data to have the same logit scale. The result was a data calibration of the levels of student's ability and item difficulty in the same interval.

The stacking analysis technique put pre-test and post-test data vertically [74]; meanwhile, the items appeared once in the experimental and control groups, allowing the researchers to check out any changes of the students after the intervention [56]. The examination was based on the same item, making the changes in students’ ability during the pre- and post-test be measured [56]. Hence, each student created two measures of abilities, namely pre-test and post-test, and one measure for each item. The research hypothesis is that the students' conceptual understanding from pre-test to post-test changes, both in the experimental and control groups.

Conversely, the racking analysis technique put both pre- and post-test data horizontally, in which each item appeared twice in data collection, and students' ability only emerged once. This enabled the researchers to check out the effects of learning implementation on each student's ability from the tests, especially the changes in item difficulty levels before and after the intervention [56].

Table 2
All possibilities of responses [70, 71, 72].

(Q1)	(Q2)	(Q3)	Code	Conceptual Understanding Category	Level
Correct	Correct	Certain	CCC	Scientific Knowledge (SK)	6
Correct	Incorrect	Certain	CIC	Misconception False Positive (MFP)	5
Incorrect	Correct	Certain	ICC	Misconception False Negative (MFN)	4
Incorrect	Incorrect	Certain	IIC	All-Misconception (ALM)	3
Correct	Correct	Uncertain	CCU	Lack of Confidence/Lucky Guess. (LG)	2
Correct	Incorrect	Uncertain	CIU	Lack of Knowledge (LOK)	1
Incorrect	Correct	Uncertain	ICU	Lack of Knowledge (LOK)	1
Incorrect	Incorrect	Uncertain	IIU	Lack of Knowledge (LOK)	1

3. Results

3.1. Rasch analysis properties of instrument

The summary of changes in concepts and items analyzed by the Rasch model is presented in Table 1. Table 2 provides the item fit statistic. An item is considered to experience a misfit if the measurement result is not in line with the following three criteria: 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 [25]. All items comply with the Outfit MNSQ criterion; item 15 does not meet the Outfit MNSQ criterion; five items (item 1, 6, 12, 13, and 15) are not in accordance with the Outfit (ZSTD) criterion; all items meet the PTMEA CORR criterion. Simply put, all items fulfill those criteria mentioned previously (none having a misfit), and are fit and valid.

This instrument has a good unidimensionality (Appendix 1). Raw variant index arrives at above the standard of 20% (33.9%), indicating that the instrument can effectively measure students' understanding of the hydrolysis concept [75]. The assessment scale analysis (Appendix 2) informs that the observation mean starts from logit -1.73 for category 1 (LOK) to logit +1.76 (category 6, SK). This signifies that the category of students’ understanding takes place consistently [27]. In addition, the high item separation index (logit 6.71) and the high item reliability (logit .98) (Table 3) indicate that the respondents (students) are sufficient to confirm the level of item difficulty, strengthening the instrument construct validity [27]. The higher the item separation and reliability index, the more confident the researchers are about replicating item placement in other suitable sample students [25, 27]. Person separation index and person reliability that reach logit 2.0 and logit .75 (Table 4), respectively, imply that the instrument is quite sensitive to differentiate the high and low abilities of the students [25, 27]. According to the Rasch model calculation, the coefficient of Cronbach Alpha of logit .81 (Table 4) reflects an interaction between 380 students and 15 items with an excellent category [24, 76]. In other words, the interaction between students and items is very significant. The instrument has

Table 3
Item statistics: Misfit order.

Item	Difficult	Error	Outfit MNSQ	Outfit ZSTD	PTMEA CORR.
1	-.38	.05	1.36	2.87	.47
2	.20	.04	1.13	1.56	.49
3	-.36	.05	.91	-.79	.43
4	.33	.04	1.09	.77	.55
5	-.25	.05	.94	-.55	.56
6	.26	.04	1.20	2.44	.41
7	.15	.04	.91	-1.17	.54
8	.47	.04	.90	-1.45	.44
9	-.47	.05	1.19	1.49	.46
10	.08	.04	1.09	1.04	.55
11	-.34	.05	1.04	.42	.51
12	-.06	.04	.71	-3.50	.60
13	.46	.04	.74	-4.12	.55
14	-.36	.05	1.00	.77	.55
15	.26	.04	1.31	3.74	.47

Table 4
Person separation and reliability statistics.

Parameter	Measure	SD	Separation	Reliability	INFIT		OUTFIT		KR-20
					MNSQ	ZSTD	MNSQ	ZSTD	
Person (N = 380)	.67	.52	1.72	.75	1.00	.04	1.02	.10	.81
Item (N = 15)	.00	.32	6.71	.98	1.07	.41	1.02	-.01	

an excellent internal psychometric consistency and is considered very reliable.

3.2. The difference in students' understanding ability of hydrolysis concept

The result of the Mann-Whitney test (Table 5) brings out the fact that statistically, there is a significant difference in the results of pre-test (U = 3459.000), $p < 0.05$) and post-test (U = 1723.000, $p < 0.05$) among students in experimental and control groups. Further, the Wilcoxon test result (Table 6) shows that the results of pre-test and post-test of students in the experimental group (Z = -8.076) and the control group (Z = -6.690) at the significant level *(p)* < 0.05 are significantly different. This suggests that students' understanding of the hydrolysis concept after the intervention (post-test) is higher than before the intervention (pre-test), both in experimental and control groups. However, the abilities of students in the experimental group are better than those in the control group. Accordingly, the learning process with the POGIL in the SSI context is better than the conventional learning.

Table 5
The result of the Mann-Whitney U test based on Students' Pre-Test and Post-Test Abilities in Experimental and Control Groups ($p < 0.05$).

Test	Experimental Group (N = 97)	Control Group (N = 93)	U	p
Pre-test	0.5026 (-0.57–1.26)[a]	0.3029 (-1.61–1.03)[a]	3459.000	0.005
Post-test	1.1722 (-0.09–3.00)[a]	0.7052 (-1.06–1.47)[a]	1723.000	0.000

Table 6
The result of the Wilcoxon test of Students' Pre-Test and Post-Test in Experimental and Control Groups ($p < 0.05$).

Group	Pre-test	Post-test	Z	p*
Experimental	0.5026 (-0.57–1.26)[a]	1.1722 (-0.09–3.00)[a]	-8.076	0.000
Control	0.3029 (-1.61–1.03)[a]	0.7052 (-1.06–1.47)[a]	-6.690	0.000

Table 7
Logit of mean of pre- and post-test items of experimental and control groups.

Group	Student	Item	Mean/SE (logit)		
			Pre-test	Post-test	Pre- and Post-test Difference
Experimental	97	15	.51/(.21)	1.50/(.32)	.99
Control	93	15	.26/(.20)	.87/(.24)	.61

Description: SE = Standard Error.

3.3. The changes in students' understanding ability of the hydrolysis concept

From the different changes in pre- and post-test (Table 7), students in the experimental and control groups have improved their understanding of the hydrolysis concept. The experimental group's mean of pre-test and post-test is logit .51 (S.E = logit .21) and logit 1.50 (S.E = logit .32), respectively, with the mean difference of both tests is (logit .99). In contrast, the mean of pre-test and post-test of the control group gets logit .26 (S.E = logit .20) and logit .87 (S.E = logit .26), respectively, with the mean difference of pre- and post-test is logit .61. Such differences indicate different effects of interventions in the experimental and control group.

If the pre-test and post-test results of the experimental group are plotted in pairs (Figure 1), so that the mean difference in the sample pre- and post-test (logit +.99) is displayed as an intercept on the horizontal axis with the plotted slope = 1, several facts obtained: First, two lines that form the upper and lower curves separate 66 students around the empirical plot line, in which the pre-test and post-test mean is not significantly different from the mean difference in the pre- and post-test in the experimental group. Second, above the curve, 23 students experience significant changes; the mean of pre- and post-test is greater than the mean difference in sample pre-test and post-test. Third, seven students do not change, and ten students have negative changes (under the curve), so that they are under the curve. Similarly, the results of pre- and post-test of the control group (Figure 2) show that 53 students are around the empirical plot line; the abilities of 25 students change significantly (greater than the mean of sample pre- and post-test (logit +.61); two students do not change; 13 students experience negative changes in abilities. The difference in the plotting of pre-test and post-test results signifies different effects of interventions in the experimental and control groups.

3.4. The changes in item difficulty level

Table 8 presents the results of the racking analysis in connection with the changes in item difficulty level in the pre- and post-test of experimental and control groups. It is shown that in terms of item difficulty level, the mean of pre-test of the experimental group is (logit .32), the mean of post-test is (logit -.34), and the mean difference of the pre- and post-test is (logit -.66). Moreover, the mean of pre-test of

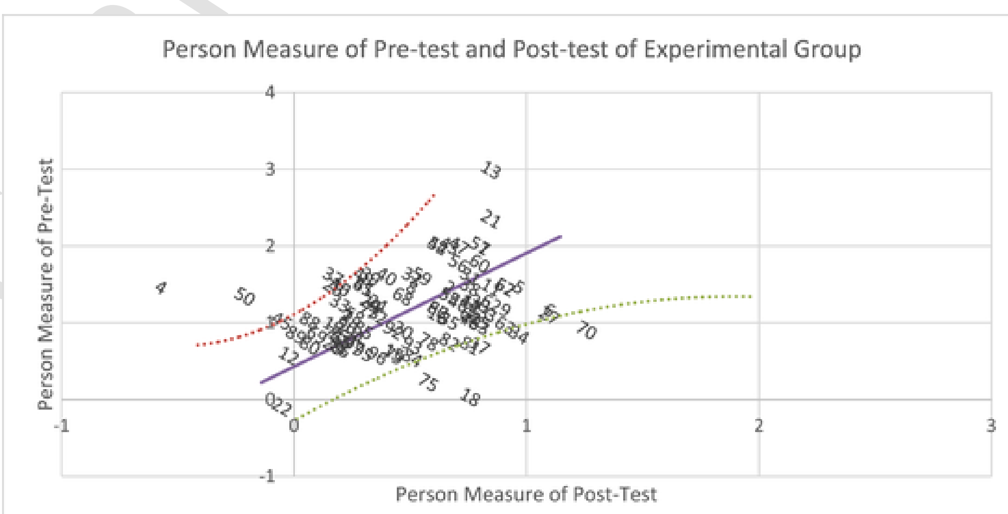

Figure 1. Scatter plots of person measures in pre- and post-test of the experimental group.

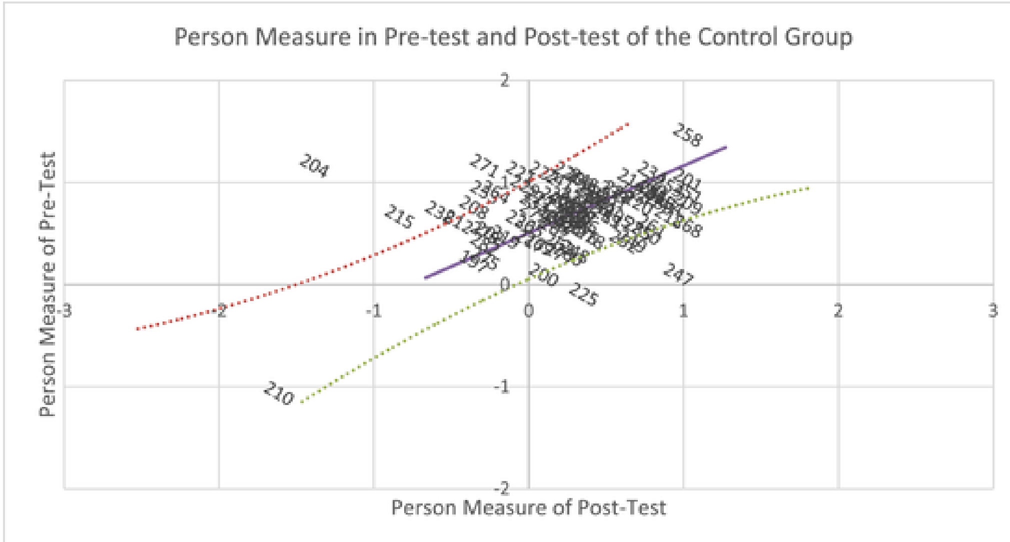

Figure 2. Scatter plots of person measures in pre- and post-test of the control group.

Table 8
Data of item measures of pre- and post-test of experimental and control groups.

Item	Experimental (Mean)			Control (Mean)		
	Pre-test	Post-test	Difference Pre- and Post-test	Pre-test	Post-test	Difference Pre- and Post-test
Item1	.16	-1.00	-1.16	-.06	-.76	-.7
Item2	.80	.01	-.79	.39	-.40	-.79
Item3	.20	-.63	-.43	-.15	-.83	-.68
Item4	.62	.25	-.37	.54	.02	-.52
Item5	.14	-.78	-.92	.10	-.49	-.59
Item6	.26	.22	-.04	.41	.30	-.11
Item7	.66	-.33	-.99	.33	-.06	-.39
Item8	.59	.45	-.14	.49	.47	-.02
Item9	-.04	-.85	-.81	-.08	-.93	-.85
Item10	.40	-.26	-.66	.32	-.01	-.33
Item11	.13	-.91	-1.04	.05	-.78	-.83
Item12	.33	-.23	-.56	.25	-.51	-.76
Item13	.77	.16	-.61	.64	.33	-.31
Item14	-.10	-.80	-.7	.15	-.83	-.98
Item15	.25	-.40	-.65	.39	.72	.33
Mean	.32	-.34	-.66	.25	-.25	-.50

the control group is (logit .25), the mean of post-test is (logit -.25), and the mean difference of the pre- and post-test is (logit -.50). This research also finds out that seven items have significant changes in the item difficulty level in the experimental group, lower than the pre- and post-test mean difference of (logit -.66), namely item 1, 2, 5, 7, 9, and 11. Eight items with a difficulty level greater than the mean are item 3, 4, 6, 8, 12, 13, 14, and 15. Item 10 has the same difficulty level as the mean. In the control group, eight items change significantly or less than the pre- and post-test mean difference of (logit -.50), including item 2, 3, 4, 5, 9, 11, 12, and 14; five items (item 1, 6, 7, 8, 10, 13) are greater than the mean; one item (item 15) has negative changes or becomes more difficult. The most difficult item in the experimental group is item 1 (.80 logit) and the easiest one is item 14 (logit -.10). Meanwhile, the most difficult item in the control group is item 13 (logit .64), and item 3 (logit -.15) is the easiest one. These findings indicate differences in the item difficulty level changes between students taught by the POGIL in the SSI context and the conventional model.

3.5. Conceptual changes in students' ability and item difficulty levels

Apart from the effect of learning interventions, there are three other factors that tend to influence the changes in students' ability and item difficulty levels, as follows: 1) guessing which happened to be correct or (lucky guess), 2) cheating, 3) carelessness. These factors can be identified from the student's item response pattern using a scalogram. For instance, the response pattern of post-test item 7 for student 353, 375, and 170 (Table 9). These three students, in the seventh and eighth row from the left, cannot understand item 12 (logit -.06) and item 10 (logit .08). Meanwhile, they can correctly explain the more difficult item, i.e., item 7 (logit .15). This situation implies a lucky guess, which in fact, these students have higher post-test abilities than the item 7 logit. Next is a cheating indication in the response pattern of student 128, 129, 134, 137, and 146. Such an indication is initially detected from the same post-test mean (logit 1.61) and item response pattern. The last one is carelessness, e.g., student 110, 118, and 139 are considered to be careless as they cannot correctly explain the easy item 4 (logit .33), yet can accurately understand item 13 (logit .46), which is harder than item 4. Moreover, they get very high post-test abilities.

3.6. Negative changes

Negative changes in conceptual understanding are detected from the changes in students' post-test logit less than the pre-test logit. For example, two students from the experimental group (E18 and E75) and the control group (C225 and C247) are taken; they have negative changes (Table 10). This means that these four students experience decreased abilities after the intervention. The pre-test item mean and the post-item mean of student E18 are (logit .76) and (logit .04), sequentially, with the mean difference of pre- and post-test arriving at (logit -.72). Moreover, the pre- and post-test item standard errors of student E18 are (logit .22) and (logit .18), respectively, with the combined standard error of logit .40. On account of the higher combined standard error than the pre- and post-test measures, the ability of student E18 in both tests is not significantly different. This also applied to student E75, C225, and C247.

4. Discussion and conclusion

The findings show changes in students' understanding abilities of the hydrolysis concept and items after the intervention. From the pre- and post-test mean difference, the experimental group has better posi-

Table 9
Scalogram.

```
GUTTMAN SCALE OF RESPONSES:
Person |Item
       |    11 11    1 1   ID       Pre-  Post-test  Pre Post    Item
       |913415207265438   Person   Mean    Mean    Difference   Response Pattern
       |---------------
   353 +666555536665554  353MFCB    .8     .97      .17    Lucky Guess
   375 +166566516133664  375MMCB  -.28     .40      .68    Guessing answer
   170 +664666446566556  170NFEB   .33    1.17      .84    accidentaly correct

   128 +666666666646555  128DFEB   .76    1.61      .85
   129 +666666666646555  129DFEB   .51    1.61     1.10    Same response pattern
   134 +666666666646555  134JFEB   .17    1.61     1.44    Cheating indication
   137 +666666666646555  137MMEB   .04    1.61     1.21
   146 +666666666646555  146NFEB   .30    1.61     1.31

   110 +666666666666566  110NFEB   .85    3.00     2.15    Response pattern
   118 +666666666666565  118RFEB   .85    2.36     1.51    "Careless"
   139 +666666665666565  139MFEB   .62    2.01     1.39
```


```
GUTTMAN SCALE OF RESPONSES:
Person |Item
       |    11 11    1 1   ID       Pre-  Post-test  Pre Post    Item
       |913415207265438   Person   Mean    Mean    Difference   Response Pattern
       |---------------
   353 +666555536665554  353MFCB    .8     .97      .17    Lucky Guess
   375 +166566516133664  375MMCB  -.28     .40      .68    Guessing answer
   170 +664666446566556  170NFEB   .33    1.17      .84    accidentaly correct

   128 +666666666646555  128DFEB   .76    1.61      .85
   129 +666666666646555  129DFEB   .51    1.61     1.10    Same response pattern
   134 +666666666646555  134JFEB   .17    1.61     1.44    Cheating indication
   137 +666666666646555  137MMEB   .04    1.61     1.21
   146 +666666666646555  146NFEB   .30    1.61     1.31

   110 +666666666666566  110NFEB   .85    3.00     2.15    Response pattern
   118 +666666666666565  118RFEB   .85    2.36     1.51    "Careless"
   139 +666666665666565  139MFEB   .62    2.01     1.39
```


tive changes than the control group [58]. In addition to the effect of the intervention, there is another factor contributing to the positive conceptual changes mentioned above, in terms of students' ability and item difficulty levels [24, 58]. The factor refers to some students who "accidently" give a correct response pattern (in the post-test). Even so, both groups have also experienced negative changes, implying that the intervention is specifically responded by students on account of the carelessness factor or a misconception-comprising response pattern [56, 58, 77]. Regarding this, not all learning objectives of the hydrolysis concept match the approach of POGIL in the SSI context. Negative changes of the students are because they are not epistemologically involved in the learning process, particularly in the observing, measuring, and calculating stages. These activities are interrelated up to group discussions as part of the stages of conceptual formation based on empirical facts [78]. Students are expected to explain and link the concepts they have learned following their epistemological reasoning [16, 79].

Furthermore, the interpretation of changes due to pedagogical interventions is exemplified by four students (Table 8) in item 5. In the pre-test, the ability of student E18 (logit .76), student E75 (logit .58), student C225 (logit .36), and student C247 (logit .96) is greater. They also respond to item 5 (-.25 logit) accurately. However, in the post-test item 5, the response of student E18, E75, C225, and C247 is incorrect due to their decreased post-test abilities. Therefore, the pre- and post-test mean difference is lower than item 5. Why do these changes occur? Such changes are exemplified by the response pattern of student E18 in item 5. This item measures students' ability in determining the reaction of NaOCl reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$, with the estimated pH = 7 and is alkaline. The question (Q1) of this item is, "is it correct that NaOCl is alkaline?". E18 answers "correct" in the pre-test, yet responds to "incorrect" in the post-test. The question (Q2) of this item is "what is your consideration for your answer in the Q1?". Four options are provided: (a) because NaOCl is formed of strong acids and weak bases; (b) because NaOCl is formed of weak acids and strong bases; (c) because NaOCl is formed of weak acids and weak bases; (d) because NaOCl is formed of strong acids and strong bases. In the pre-test, E18 chooses the correct answer (b), yet selects the incorrect answer (a) in the post-test that comprises misconception. Next, in the Q3 of this item, E18 chooses "very sure" in the pre-test and "not sure" in the post-test. The item 5 response pattern of E18 becomes CCC (category of scientific knowledge - SK) in the pre-test and IIU (category of lack of knowledge - LOK) in the post-test. Accordingly, the response pattern changes from CCC to IIU. The pre- and post-test mean difference of E18 (logit -.72) lower than item 5 (-.25) signifies that the error of response pattern results from misconception. This also applies to the response pattern of E75 (logit -.35), C225 (logit -.45), and C247 (logit -.87).

The misconception refers to the inability to identify the NaOCl salt hydrolysis that is formed of weak acids and strong bases. In short, the four students tend to not understand the concept of acid and base and acid-base reaction. These findings strengthen several previous studies [51, 53, 54, 80]. A study on the understanding of the acid-base concept of senior high school students in Malaysia concludes that some students have little understanding of the function of detergents as the cleaning agent, the difference between strong acids and strong bases, and the treatment for soil acidity using fertilizers [53]. In the same tune, such little understanding is because they do not conceptualize acid-base strength as a property that arises from the interaction of many reaction

Table 10
Scalogram results of student E18, E75, C225, and 247.

<table>
<tr><th rowspan="2">ID Person</th><th rowspan="2">Test</th><th>Item Response Pattern</th><th colspan="4">Mean</th></tr>
<tr><th>| 11 11 1 1
|913415207265438
|---------------</th><th>Item Logit</th><th>S.E* Logit</th><th>Pre- test and post-test difference</th><th>Combined S.E</th></tr>
<tr><td rowspan="2">E18</td><td>Pre-test</td><td>+665666636366333</td><td>.76</td><td>.22</td><td rowspan="2">-.72</td><td rowspan="2">.40</td></tr>
<tr><td>Post-test</td><td>+666661322521161</td><td>.04</td><td>.18</td></tr>
<tr><td rowspan="2">E75</td><td>Pre-test</td><td>+562664552566426</td><td>.58</td><td>.20</td><td rowspan="2">-.35</td><td rowspan="2">.38</td></tr>
<tr><td>Post-test</td><td>+655664322323463</td><td>.23</td><td>.18</td></tr>
<tr><td rowspan="2">C225</td><td>Pre-test</td><td>+616665663261613</td><td>.36</td><td>.19</td><td rowspan="2">-.45</td><td rowspan="2">.37</td></tr>
<tr><td>Post-test</td><td>+611622566131613</td><td>-.09</td><td>.18</td></tr>
<tr><td rowspan="2">C247</td><td>Pre-test</td><td>+663636666666435</td><td>.97</td><td>.25</td><td rowspan="2">-.87</td><td rowspan="2">.43</td></tr>
<tr><td>Post-test</td><td>+563345555314133</td><td>.10</td><td>.18</td></tr>
</table>

Description: S.E = Standar Error

<table>
<tr><th rowspan="2">ID Person</th><th rowspan="2">Test</th><th>Item Response Pattern</th><th colspan="4">Mean</th></tr>
<tr><th>| 11 11 1 1
|913415207265438
|---------------</th><th>Item Logit</th><th>S.E* Logit</th><th>Pre- test and post-test difference</th><th>Combined S.E</th></tr>
<tr><td rowspan="2">E18</td><td>Pre-test</td><td>+665666636366333</td><td>.76</td><td>.22</td><td rowspan="2">-.72</td><td rowspan="2">.40</td></tr>
<tr><td>Post-test</td><td>+666661322521161</td><td>.04</td><td>.18</td></tr>
<tr><td rowspan="2">E75</td><td>Pre-test</td><td>+562664552566426</td><td>.58</td><td>.20</td><td rowspan="2">-.35</td><td rowspan="2">.38</td></tr>
<tr><td>Post-test</td><td>+655664322323463</td><td>.23</td><td>.18</td></tr>
<tr><td rowspan="2">C225</td><td>Pre-test</td><td>+616665663261613</td><td>.36</td><td>.19</td><td rowspan="2">-.45</td><td rowspan="2">.37</td></tr>
<tr><td>Post-test</td><td>+611622566131613</td><td>-.09</td><td>.18</td></tr>
<tr><td rowspan="2">C247</td><td>Pre-test</td><td>+663636666666435</td><td>.97</td><td>.25</td><td rowspan="2">-.87</td><td rowspan="2">.43</td></tr>
<tr><td>Post-test</td><td>+563345555314133</td><td>.10</td><td>.18</td></tr>
</table>

Description: S.E = Standar Error

factors [51]. Additionally, research on an alternative conception of salt hydrolysis among senior high school students contends that the concept of hydrolysis is challenging for the students [54]. They are usually able to state the acidity of a salt solution correctly, yet writing a chemical equation to explain such a phenomenon is a great challenge. Most of the alternative conceptions are identifiable, rooted in the misunderstanding of equilibrium process, acid and base, material structure and other basic problems, student tendency to use a wrong analogy, and the lack of laboratory practice.

This research findings and elaboration of negative changes (case E18) prove the advantages of the Rasch model, specifically its potential in linking the result of changes (pre- and post-test), the item difficulty level, and the content being measured [18]. Such information solely comes from the Rasch model-based stacking and racking analysis techniques. The stacking technique provides information regarding “who has changed”; in contrast, the racking technique offers information of “what has changed” [56, 58], allowing the researchers to spell out the effect of the applied pedagogical innovation [18, 33, 34]. Although the instrument measurement result of this work is not data-rich, the analysis strength of the Rasch model can describe in detail the conceptual changes, both in the students’ ability and item difficulty levels.

4.1. Limitations and further studies

The primary limitation of this research is that it did not take into account the aspects of learning style, culture, and motivation that can change due to learning interventions. Future studies, therefore, can address these aspects. The present study can be continued by considering the context of a problem that closely connects with the parameter of item difficulty level. The analysis will be more interesting if it can prove that different item difficulty levels are influenced by problem contexts in each item [81]. Further studies are also expected to find an analysis technique that can integrate problem contexts, item characteristics, and item difficulty levels in a measurement model. It is assumed that different problem contexts in each item will be more likely to affect measurement results because problem contexts have conceptual linkage with items and student activities in doing experiments, measuring, interpreting data/graphs, and others. Thus, the linkages between the learning process during the intervention and conceptual changes in students’ ability and item difficulty levels can be explained in detail; which part of the process leads the students to change their understanding related to specific ideas taught to them.

Declarations

Author contribution statement

Lukman Abdul Rauf Laliyo: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Bambang Sumintono: Analyzed and interpreted the data; Wrote the paper.

Citra Panigoro: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

We would like to express our deep gratitude to all parties for their assistance in data processing, especially the Head of Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo.

Appendix

Appendix 1. Standardized Residual Variance in Eigen value Units.

```
   Table of STANDARDIZED RESIDUAL variance in Eigenvalue uni
information units
                                             Eigenvalue   Obser\
Total raw variance in observations     =       22.7067 100.0%
  Raw variance explained by measures   =        7.7067  33.9%
    Raw variance explained by persons  =        2.7733  12.2%
    Raw Variance explained by items    =        4.9334  21.7%
  Raw unexplained variance (total)     =       15.0000  66.1% 1
    Unexplned variance in 1st contrast =        2.0698   9.1%
    Unexplned variance in 2nd contrast =        1.5312   6.7%
    Unexplned variance in 3rd contrast =        1.3696   6.0%
    Unexplned variance in 4th contrast =        1.3124   5.8%
    Unexplned variance in 5th contrast =        1.1945   5.3%
```

Appendix 2. Summary of Category Structure.

```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
-------------------------------------------------------------
|CATEGORY      OBSERVED|OBSVD SAMPLE|INFIT OUTFIT|| ANDRICH |CATE
|LABEL   SCORE COUNT  %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEA
|----------------------+------------+------------++---------+----
|    1   1      317   6|  -.18  -.20|  1.06  1.08||  NONE   |( -1
|    2   2      190   3|   .10   .09|  1.03  1.18||     .46 |
|    3   3      963  17|   .33   .31|  1.02   .93||   -1.43 |
|    4   4      542  10|   .56   .52|  1.02   .97||     .98 |
|    5   5     1262  22|   .62   .73|  1.27   .99||    -.22 |
|    6   6     2425  43|  1.02   .98|   .97  1.04||     .20 |(  1
|----------------------+------------+------------++---------+----
| MISSING           1   0|  -.30      |            ||         |
-------------------------------------------------------------
 OBSERVED AVERAGE is mean of measures in category. It is not a p
```


References

[1] S. Lu, H. Bi, Development of a measurement instrument to assess students' electrolyte conceptual understanding, Chem. Educ. Res. Pract. 17 (4) (2016) 1030–1040.
[2] H.E. Yildirir, H. Demirkol, Identifying mental models of students for physical and chemical change, J. Baltic Sci. Educ. 17 (6) (2018) 986–1004.
[3] C.R. Gette, M. Kryjevskaia, M.R. Stetzer, P.R.L. Heron, Probing student reasoning approaches through the lens of dual-process theories: a case study in buoyancy, Phys. Rev. Phys. Educ. Res. 14 (1) (2018) 10113.
[4] A.H. Johnstone, Chemical education research in Glasgow in perspective, Chem. Educ. Res. Pract. 7 (2) (2006) 49–63.
[5] A.H. Johnstone, You can't get there from here, J. Chem. Educ. 87 (1) (2010) 22–29.
[6] K.S. Taber, Chemical misconceptions—prevention, diagnosis, and cure, J. Chem. Educ. 80 (5) (2003) 491.
[7] K.S. Taber, Challenging misconceptions in the chemistry classroom: resources to support teachers, Educació Química EduQ. 4 (2009) 13–20.
[8] J.I. Alamina, I.S. Etokeren, Effectiveness of imagination stretch teaching strategy in correcting misconceptions of students about particulate nature of matter, J. Educ. Soc. Behav. Sci. 27 (1) (2018) 1–11.
[9] I.Z. Yaşar, E. İnce, F.G. Kırbaşlar, 7. Class Science and Technology Course "Structure of Atom" Subject Readiness Improvement Test, 152, Procedia Soc. Behav. Sci., 2014, pp. 662–667.
[10] K.Y. Hoe, R. Subramaniam, On the prevalence of alternative conceptions on acid-base chemistry among secondary students: insights from cognitive and confidence measures, Chem. Educ. Res. Pract. 17 (2) (2016) 263–282.
[11] D.C. Aktan, Investigation of students' intermediate conceptual understanding levels: the case of direct current electricity concepts, Eur. J. Phys. 34 (1) (2013) 33–43.
[12] M. Allen, Misconceptions in Primary Science, Open University Press, Maidenhead, 2014.
[13] Soeharto, B. Csapó, E. Sarimanah, F.I. Dewi, T. Sabri, A review of students' common misconceptions in science and their diagnostic assessment tools, J. Pendidik. IPA Indones. 8 (2) (2019) 247–266.
[14] S. Almuntasheri, R.M. Gillies, T. Wright, The effectiveness of a guided inquiry-based, teachers' professional development programme on saudi students' understanding of density, Sci. Educ. Int. 27 (1) (2016) 16–39.
[15] M.Z. Hashweh, The complexity of teaching density in middle school, Res. Sci. Technol. Educ. 34 (1) (2016) 1–24.
[16] K. Zvoch, S. Holveck, L. Porter, Teaching for conceptual change in a density unit provided to seventh graders: a comparison of teacher- and student-centered approaches, Res. Sci. Educ. 51 (2019) 1395–1421.
[17] R. Grob, M. Holmeier, P. Labudde, Formative assessment to support students' competences in inquiry-based science education, Interdiscip. J. Probl-based. Learn. 11 (2) (2017).
[18] T.C. Pentecost, J. Barbera, Measuring learning gains in chemical education: a comparison of two methods, J. Chem. Educ. 90 (7) (2013) 839–845.
[19] B. Sumintono, Rasch model measurements as tools in assesment for learning, in: Proceedings of the 1st International Conference on Education Innovation (ICEI 2017); 2017 October 14; Surabaya, Atlantis Press, Paris, 2018.
[20] R. Micelli, M. Settanni, G. Vidotto, Measuring change in training programs : an empirical illustration Measurement of change and Item Response Theory, Psychol. Sci. Q. 50 (3) (2008) 433–447.
[21] S.D. Willoughby, A. Metz, Exploring gender differences with different gain calculations in astronomy and biology, Am. J. Phys. 77 (7) (2009) 651–657.
[22] R.L. Linn, J.A. Slinde, The determination of the significance of change between pre- and posttesting periods, Rev. Educ. Res. 47 (1) (1977) 121–150.
[23] P. He, X. Liu, C. Zheng, M. Jia, Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons, Chem. Educ. Res. Pract. 17 (2) (2016) 381–393.
[24] B. Sumintono, W. Widhiarso, Aplikasi Pemodelan Rasch Pada Assessment Pendidikan [Application of Rasch Modeling in Educational Assessment], Bandung: Penerbit Trim Komunikata, 2015. https://www.researchgate.net/publication/282673464%0AAplikasi.
[25] W.J. Boone, M.S. Yale, J.R. Staver, Rasch Analysis in the Human Sciences, Springer, Dordrecht, 2014.
[26] X. Liu, Developing measurement instruments for science education research, in: B. Fraser, K.G. Tobin, C.J. McRobbie (Eds.), Second International Handbook of Science Education, Springer, Netherlands, 2012, pp. 651–665.
[27] J.M. Linacre, 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. US, Winsteps, 2020.
[28] S. Sabah, H. Hammouri, M. Akour, Validation of a scale of attitudes toward science across countries using Rasch model: findings from TIMSS, J. Baltic Sci. Educ. 12 (5) (2013) 692–703.
[29] J.C. Hadenfeldt, S. Bernholt, X. Liu, K. Neumann, I. Parchmann, Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter, J Chem. Educ. 90 (12) (2013) 1602–1608.
[30] C.F. Herrmann-Abell, G.E. DeBoer, Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items, Chem. Educ. Res. Pract. 2 (2) (2011) 184–192.
[31] C.F. Herrmann-Abell, G.E. Deboer, Using Rasch modeling and option probability curves to diagnose students' misconceptions, in: The 2016 AERA Annual Meeting; 2016 April 8-12, American Educational Research Assossiation, Washington DC. USA, 2016, Available from:. https://www.aaas.org/sites/default/files/Project2061_CHA-AERA%20energy%20paper%204-7-16.pdf.
[32] L.A.R. Laliyo, J.S. Tangio, B. Sumintono, M. Jahja, C. Panigoro, Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter, J. Baltic Sci. Educ. 19 (5) (2020).
[33] B. Davidowitz, M. Potgieter, Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry, Int. J. Sci. Educ. 38 (9) (2016) 1483–1503.
[34] M. Park, X. Liu, An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics, Res. Sci. Educ. (2019).
[35] D.F. Treagust, S.S. Qureshi, V.R. Vishnumolakala, J. Ojeil, M. Mocerino, D.C. Southam, Process-oriented guided inquiry learning (POGIL) as a culturally relevant pedagogy (CRP) in Qatar: a perspective from grade 10 chemistry classes, Res. Sci. Educ. 1–19 (2018).
[36] A.T. Kinslow, T.D. Sadler, H.T. Nguyen, Socio-scientific reasoning and environmental literacy in a field-based ecology class, Environ. Educ. Res. 4622 (2018) 1–23.
[37] D.C. Owens, T.D. Sadler, P. Friedrichsen, Teaching practices for enactment of socio-scientific issues instruction: an instrumental case study of an experienced biology teacher, Res. Sci. Educ. (2019).
[38] R. Bruder, A. Prescott, Research evidence on the benefits of IBL, ZDM Int. J. Math. Educ. 45 (6) (2013) 811–822.
[39] D.L. Zeidler, Socioscientific issues as a curriculum emphasis: theory, research and practice, in: N.G. Lederman, S.K. Abell (Eds.), Handbook of Research on Science Education, Routledge, New York, 2014, pp. 697–726.
[40] A.G. Espeja, D.C. Lagarón, Socio-scientific Issues (SSI) in initial training of primary school teachers: pre-service teachers' conceptualization of SSI and appreciation of the value of teaching SSI, Procedia Soc. Behav. Sci. 196 (2015) 80–88.
[41] National Research Council, A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas, The National Academies Press, Washington DC, 2012.
[42] T.D. Sadler, D.L. Zeidler, Scientific literacy, PISA, and socioscientific discourse:

assessment for progressive aims of science education, J. Res. Sci. Teach. 46 (8) (2009) 909–921.
[43] N.G. Lederman, J.S. Lederman, A. Antink, Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy, Int. J. Educ. Math. Sci. Technol. 1 (3) (2013) 138–147.
[44] A.N. Cooke, K.S. Fielding, W.R. Louis, Environmentally active people: the role of autonomy, relatedness, competence and self-determined motivation, Environ. Educ. Res. 22 (5) (2016) 631–657.
[45] T.D. Sadler, Moral and ethical dimensions of socioscientific decision-making as integral components of scientific literacy, Sci. Educ. 13 (1) (2004) 39–48.
[46] G.E. DeBoer, Scientific literacy: another look at its historical and contemporary meanings and its relationship to science education reform, J. Res. Sci. Teach. 37 (6) (2000) 582–601.
[47] T.D. Sadler, Socio-scientific issues-based education: what we know about science education in the context of SSI, in: T.D. Sadler (Ed.), Socio-Scientific Issues in the Classroom: Teaching, Learning and Research, Springer, Dordrecht, 2011, pp. 355–369.
[48] W. Gräber, P. Nentwig, H.J. Becker, E. Sumfleth, A. Pitton, K. Wollweber, D. Jorde, Scientific literacy: from theory to practice, Res. Sci. Educ. Past. Present. Future. 1996 (2005) 61–70.
[49] J. Grooms, A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction, J. Chem. Educ. (2020).
[50] M.L. Presley, A.J. Sickel, N. Muslu, D. Merle, A framework for socio-scientific issues based education, Sci. Educ. 22 (2013) 26–32.
[51] H. Tümay, Emergence, learning difficulties, and misconceptions in chemistry undergraduate students' conceptualizations of acid strength, Sci. Educ. 25 (1–2) (2016) 21–46.
[52] N. Seçken, Identifying student's misconceptions about SALT, Procedia Soc. Behav. Sci. 2 (2) (2010) 234–245.
[53] M.I.M. Damanhuri, D.F. Treagust, M. Won, A.L. Chandrasegaran, High school students' understanding of acid-base concepts: an ongoing challenge for teachers, Int. J. Environ. Sci. Educ. 11 (1) (2016) 9–27.
[54] K. Orwat, P. Bernard, A. Migdał-Mikuli, Alternative conceptions of common salt hydrolysis among upper-secondary school students, J. Baltic Sci. Educ. 16 (1) (2017) 64–76.
[55] J.W. Creswell, Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, fourth ed., Pearson Education, Inc, New York, 2012.
[56] B.D. Wright, Rack and stack: time 1 vs. time 2 or pre-test vs. post-test, Rasch Meas. Trans. 17 (1) (2003) 905–906.
[57] P. Anselmi, G. Vidotto, O. Bettinardi, G. Bertolotti, Measurement of change in health status with Rasch models, Health Qual. Life Outcome 13 (1) (2015) 1–7.
[58] M. Ling, V. Pang, C.C. Ompok, Measuring change in early mathematics ability of children who learn using games: stacked analysis in rasch measurement, in: Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings, 2016 July 30-August 3; Singapore, Springer, Singapore, 2018.
[59] K.S. Taber, Ethical considerations of chemistry education research involving "human subjects", Chem. Educ. Res. Pract. 15 (2) (2014) 109–113.
[60] F. Aldresti, S. Rahayu, F. Fajaroh, The influence of inquiry-based chemistry learning with the contex of socio-scientific issues on high school students' scientific explanation skills, J. Pendidik. IPA. 23 (2) (2019).
[61] Ibnu S. Mitarlis, S. Rahayu, Sutrisno, The effectiveness of new inquiry-based learning (NIBL) for improving multiple higher-order thinking skills (M-HOTS) of prospective chemistry teachers, Eur. J. Educ. Res. 9 (3) (2020) 1309–1325.
[62] M. Pedaste, M. Mäeots, L.A. Siiman, T. de-Jong, S.A.N. Van-Riesen, E.T. Kamp, C.C. Manoli, Z.C. Zacharia, E. Tsourlidaki, Phases of inquiry-based learning: definitions and the inquiry cycle, Educ. Res. Rev. 14 (2015) 47–61.
[63] W.L. Romine, T.D. Sadler, Measuring changes in interest in science and technology at the college level in response to two instructional interventions, Res. Sci. Educ. 46 (3) (2016) 309–327.
[64] M. Wilson, Constructing Measures: an Item Response Modeling Approach, Lawrence Erlbaum Associates, USA, 2005.
[65] M. Wilson, Cognitive diagnosis using item response models, Zeits Für Psych/J. Psych. 216 (2) (2008) 74–88.
[66] M. Wilson, Measuring progressions: assessment structures underlying a learning progression, J. Res. Sci. Teach. 46 (6) (2009) 716–730.
[67] L.A.R. Laliyo, Mendiagnosis Sifat Perubahan Konseptual Siswa: Penerapan Teknik Analisis Stacking Dan Racking Rasch Model [Diagnosing the Nature of Students' Conceptual Change: Application of Stacking and Racking Analysis Techniques Rasch Model], Penerbit Deepublish, Yogyakarta, 2021.
[68] G. Chittleborough, D. Treagust, Correct interpretation of chemical diagrams requires transforming from one level of representation to another, Res. Sci. Educ. 38 (4) (2008) 463–482.
[69] D.F. Treagust, Development and use of diagnostic tests to evaluate students' misconceptions in science, Int. J. Sci. Educ. 10 (2) (1988) 159–169.
[70] H.O. Arslan, C. Cigdemoglu, C. Moseley, A Three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain, Int. J. Sci. Educ. 34 (11) (2012) 1667–1686.
[71] S. Hasan, D. Bagayoko, E.L. Kelley, Misconceptions and the certainty of response index (CRI), Phys. Educ. 34 (5) (1999) 294–299.
[72] H. Habiddin, E.M. Page, Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK), Indones. J. Chem. 19 (3) (2019) 720–736.
[73] T.G. Bond, C.M. Fox, Applying the Rasch Model: Fundamental Measurement in the Human Sciences, third ed., Routledge Taylor & Francis Group, New York, 2015.
[74] C.F. Herrmann-abell, J.C. Flanagan, J.E. Roseman, Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: using student measures to evaluate the promise of the intervention, in: Proceeding of the 2013 NARST Annual International Conference. 2013 April 6-9; Puerto Rico, NARST, Puerto Rico, 2013, Available from:. https://www.aaas.org/sites/default/files/project2061/files/NARST2013-Flanagan-Paper5-Usingteachermeasurestoe.pdf.
[75] W.P. Fisher, Rating scale instrument quality criteria, Rasch Meas. Trans. 21 (1) (2007) 1095. www.rasch.org/rmt/rmt211m.htm.
[76] W.K. Adams, C.E. Wieman, Development and validation of instruments to measure learning of expert-like thinking, Int. J. Sci. Educ. 33 (9) (2011) 1289–1312.
[77] B. Sumintono, W. Widhiarso, Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [application of rasch model for social science studies], in: B. Trim (Ed.), Bandung, Trim Komunikata Publishing House, 2014. https://www.researchgate.net/publication/268688933%0AAplikasi.
[78] J.G. Rodriguez, K.H. Hunter, L.J. Scharlott, N.M. Becker, A review of research on process oriented guided inquiry learning: implications for research and practice, J. Chem. Educ. 97 (10) (2020) 3506–3520.
[79] T.S. Hancock, P.J. Friedrichsen, A.T. Kinslow, T.D. Sadler, Selecting socio-scientific issues for teaching: a grounded theory study of how science teachers collaboratively design SSI-based curricula, Sci. Educ. 28 (6–7) (2019) 639–667.
[80] N. Ültay, M. Çalik, A comparison of different teaching designs of "acids and bases" subject, Eurasia J. Math. Sci. Technol. Educ. 12 (1) (2016) 57–86.
[81] J. Hartig, A. Frey, G. Nold, E. Klieme, An application of explanatory item response modeling for model-based proficiency scaling, Educ. Psychol. Meas. 72 (4) (2012) 665–686.

PUBLISHED

Heliyon 8 (2022) e09126

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in Rasch model

Lukman Abdul Rauf Laliyo [a,*], Bambang Sumintono [b], Citra Panigoro [c]

[a] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Gorontalo, 96128, Indonesia
[b] Faculty of Education, Universitas Islam Internasional Indonesia, 16416, Indonesia
[c] Department of Aquatic Resource Management, Faculty of Fisheries and Marine Science, Universitas Negeri Gorontalo, 96128, Indonesia

HIGHLIGHTS

- Raw scores have a bias in a conventional psychometric measurement
- Stacking and racking measure students' ability and item difficulty level changes
- The learning process in socio-scientific issues improves students' understanding
- Misconceptions influence the negative values of students' pre-and post-test

ARTICLE INFO

Keywords:
Stacking
Racking
Rasch model
Hydrolysis conceptual changes
Inquiry model

ABSTRACT

This research aimed to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the process-oriented guided inquiry learning (POGIL) model in the context of socio-scientific issues (SSI) with the pretest-posttest control group design. Such techniques were based on a person- and item-centered statistic to determine how students and items changed during interventions. Eleventh-grade students in one of the top-ranked senior high schools in the eastern part of Indonesia were involved as the participants. They provided written responses (pre- and post-test) to 15 three-tier multiple-choice items. Their responses were assessed through a rubric that combines diagnostic measurement and certainty of response index. Moreover, the data were analyzed following the Rasch Partial Credit Model, using the WINSTEPS 4.5.5 software. The results suggested that students in the experimental group taught by the POGIL approach in the SSI context had better positive conceptual changes than those in the control class learning with a conventional approach. Along with the intervention effect, in certain cases, it was found that positive conceptual changes were possibly due to student guessing, which happened to be correct (lucky guess), and cheating. In other cases, students who experienced negative conceptual changes may respond incorrectly due to carelessness, the boredom of problem-solving, or misconception. Such findings have also proven that some students tend to give specific responses after the intervention in certain items, indicating that not all students fit the intervention. Besides, stacking and racking analyses are highly significant in detailing every change in students' abilities, item difficulty levels, and learning progress.

1. Introduction

Central to defining the quality of pedagogical innovation in science classes is conceptual changes. The changes refer to how ideas or conceptions the students understand according to their ways of thinking [1, 2] become scientifically accurate [3]. It is because such ideas generally comprise misconceptions [4, 5, 6, 7], are not in accordance with scientific concepts [8, 9], tend to be resistant [10], changeable and varied [11], so that they should be improved if the correct conceptual understanding is to be taught [12, 13].

Some studies have been conducted on learning innovation testing to form an accurate and scientific conceptual understanding of the students, e.g., inquiry-based learning. This model presents conceptual conflicts and participatory experiments to facilitate conceptual changes

* Corresponding author.
E-mail address: lukman.laliyo019@gmail.com (L.A.R. Laliyo).

https://doi.org/10.1016/j.heliyon.2022.e09126
Received 7 July 2021; Received in revised form 26 January 2022; Accepted 15 March 2022

[14, 15, 16]. Conceptual understanding-based learning involves various strategies in identifying and analyzing students' comprehension so that the investigation process can be designed to lead them to a more accurate and scientific conception [16, 17]. This research relied on a quasi-experimental design that assessed students' pre-test and post-test, evaluated the changes in performances for testing significant differences. This type of testing informs the researcher about the presence of an effect, but does not provide detailed information on the level and trait of the changes [18]. What if the researcher is willing to compare the extent to which the pre- and post-test change (differences in learning outcomes) and interpret the changes (the reasoning why those changes occur) in terms of content? This is a core question regarding the changes in some latent traits or changes in traits measured after the intervention. In most studies, interpreting the changes in pre-test and post-test tends to be limited to identifying whether or not an effect prevails.

Pre- and post-test changes should be given in detail regarding the students' understanding ability and item difficulty levels. However, this has not been much revealed due to the limitations of its measurement techniques and analyses and has not been the main focus in chemistry education research to date. One reason for this issue is the debate in the psychometric community regarding the ability to measure changes accurately [18]. This debate questions the use of raw scores in the conventional psychometric analysis, which largely contains measurement biases [19], as follows: 1) the difference in pre- and post-test scores will be negatively correlated with the pre-test score, especially for students with low pre-test scores [18,20]; 2) the difference in pre- and post-test scores shows low test reliability [21]; 3) low measurement properties due to different scales [22].

Raw scores are not final data, so that they do not have a great deal of information for drawing conclusions [23, 24]. Around the 1950s, Dr. Georg Rasch, a mathematician from Denmark, introduced the formulation of the Rasch measurement model [24]. The model has been widely applied to analyze various types of data, e.g., dichotomous, polytomous, multi-rating, and multi-rater data. In the mid-2000s, the Rasch model was used as a probabilistic-based psychometric measurement that went beyond the use of raw scores [25, 26], and was used to overcome the limitations of conventional psychometric measurement [19, 27]. Its analyses, including item fit, PCA (Principal Component Analysis), and Wright map, are commonly used for international test analyses, namely TIMSS and PISA [28].

In chemistry education research, the Rasch model has been relied on to evaluate learning understanding and progress [29], to diagnose students' preconceptions [1], misconceptions [13, 30, 31, 32], link the measurement of content knowledge with pedagogical content knowledge [33], and investigate item difficulty patterns [13, 34]. Even so, studies on the Rasch model to reveal the chemistry conceptual changes in students' understanding and item difficulty levels are relatively hard to find as of today. The present study aims to employ stacking and racking analysis techniques in the Rasch model to measure the hydrolysis conceptual changes of students taught by the POGIL approach in the context of SSI and students who learn conventionally. Such techniques are based on a person- and item-centered statistic to estimate how students and items change during the intervention.

POGIL is a student-centered learning strategy that teaches content or process skills. The philosophical foundation of POGIL is the involvement of an interactive process of careful thinking, discussing ideas, perfecting understanding, practicing skills, reflecting progress, and evaluating performances [35]. POGIL is able to lead the process of designing a participatory experiment that presents a conceptual conflict as a strategy to encourage students to form an accurate concept [14]. Therefore, POGIL intervention is more likely to be potential in driving epistemological understanding and reasoning [36], making students have opportunities to change their conceptions to be more accurate and scientific [16]. Nevertheless, it is also worth noting that some students potentially have misconceptions resistant to changes [3].

SSI functions as a learning context through the integration of social problems that students are familiar with. It also has a conceptual connection with salt hydrolysis [37, 38], and its resolution requires many perspectives [39], including the dimension of moral and ethical evaluation of students [40]. The SSI context is a socio-scientific phenomenon that the students should explain based on their conceptual viewpoints. It encourages them to actively get involved in grasping problems [41], developing and utilizing their knowledge [42], improving their critical thinking [43], and being able to scientifically describe the discussed socio-scientific phenomenon [36, 44, 45]. For such reasons, the integration of SSI can build up students' scientific literacy [39, 46, 47]. In the end, this integration enables the learning process to be more significant in enhancing students' understanding [45, 48]. Besides, they are skilled in negotiating the social aspect of the studied phenomenon [49, 50]. For instance, the issues of global warming, climate change, and pollution [36].

Salt hydrolysis is a learning topic in high school that is strongly related to SSI. Students with a good understanding of hydrolysis will manage to clarify scientifically why detergents, bleaching agents ($NaOCl$), and fertilizers can pollute the environment. Despite this, the linkage of this issue as the problem in learning hydrolysis is rarely carried out. The learning process is more emphasized on mastering theoretical concepts [36]. As a consequence, students find it challenging to use their hydrolysis understanding to explain socio-scientific phenomena around them [37]. This challenge is on account of their misconceptions regarding acid-base reaction [51], making them unable to elaborate the concept of salt hydrolysis [52] and determine acid and base strength [53]. In addition, they are struggling with correctly explaining the dissolving process and the reaction of ionic compounds with water, writing down chemical equations, and having different interpretations of the dissolving process mentioned earlier [54]. On this ground, it is essential to reveal how the hydrolysis concept changes if intervened with the POGIL approach in the SSI context, through the following specific questions: (1) is there a significant hydrolysis conceptual change of the students after the learning process in experimental and control groups? (2) if compared, how is the hydrolysis conceptual change through the intervention of POGIL in the SSI context and conventional learning? (3) in addition to intervention, is there any other factor that also contributes to the students' hydrolysis conceptual changes?

2. Method of study

This study relied on a quantitative approach with a quasi-experimental and pretest-posttest control group design [55] by comparing the extent to which the hydrolysis concept changes after the intervention. Researchers carried out the learning process for 12 meetings, gave tests, and collected data on the results of the intervention and measurement.

The changes of students and items were analyzed using the stacking and racking techniques in the Rasch model [56]. As standard techniques, racking and stacking were introduced by Benjamin Wright to measure the extent to which conceptual understanding of students and items change before and after interventions [57]. The referred changes are cases (item and student levels) caused by the learning intervention and can be diagnosed based on the estimated changes.

In regards to students' understanding, the measurement was to identify students who had specific hydrolysis conceptual changes in responding to the learning intervention. In terms of items, the measurement was done to identify which items had special characteristics and been understood by students differently during the learning intervention [57]. Thus, the scientific inquiry approach might not be suitable for some students, or some items might be too hard after the intervention. This insightful information is immensely helpful for researchers and education practitioners, especially in evaluating the weaknesses of pedagogical innovations being applied and devising learning strategies that meet students' needs in learning [58].

2.1. Participants

Eleventh-grade students aged 16–17 years in one of the senior high schools in the eastern part of Indonesia were involved as the sample. This top-ranked school gets an "A" accreditation (excellent) from the National Accreditation Board for High School. The sample was determined by convenience sampling in six randomly assigned classes. Three classes (N = 97) were experimental groups that applied the POGIL model in the SSI context. The other three classes (N = 93), as control groups, applied conventional learning without the SSI context. The same teacher taught these classes following the Curriculum 2013 of Chemistry Subject (revised in 2016). There was no special classroom for learning the concept of hydrolysis, i.e., taking up the regular learning process at school. Before learning the hydrolysis concept, the students had previously learned the concept of acid and base to understand the concept of salt hydrolysis way better. Research permission was obtained from the government and school administrators. In accordance with principles of research ethics, research purpose and procedures were informed to all the students being involved and that they were voluntarily participating. Additionally, their information is confidential and only used for science development [59].

2.2. Learning implementation

Students in the experimental group studied employing the process-oriented guided inquiry learning (POGIL) in the SSI context [35]. Meanwhile, in the control class, the learning process was performed conventionally; the teacher facilitated learning initiatives. The learning process focused more on content mastery and problem-solving practice. Applying the POGIL model in the SSI context highlights teacher assistance to guide the students to prepare their conceptual understanding based on epistemological reasoning they get from experiments, discussions, and collaborations [49, 60]. Researchers carried out the learning process for eight weeks to apply the intervention to the sample, gave tests, collected data on the results of the intervention and measurement. The first three weeks were the preparation stages when researchers and the teacher shared perceptions, and asked the teacher to perform a learning simulation under the scenario, including different assistance techniques in leading the students to conduct experiments, and to ask analytical questions. The pre-test was carried out in the third week. Further, the learning implementation was done for four weeks, and the post-test was executed in the eighth week.

The learning stages with POGIL in the SSI context consist of orientation, exploration, concept formation, application, and closing. During the orientation stage, the teacher presented familiar contextual phenomena related to the concept of hydrolysis. The teacher asked initial questions to provoke curiosity and arouse motivation and interest of the students. While watching the video, had the students responded and explained the relationship between the phenomena and acids and bases, hydrolysis, and buffers. In the exploration stage, the teacher developed analytical questions with data, images, and multiple video clips to give perspectives on learning objectives and to delve into the concept that had been and would be learned. Next, the teacher assisted the students in doing experiments guided by a worksheet, and at the same time, asked analytical questions to lead them and strengthen their conceptual understanding. In the concept formation stage, the teacher asked students to build their conceptual understanding based on the exploration results, accompanied by critical and fundamental questions to guide students in building a conceptual understanding of the salt hydrolysis and buffer solution.

Following the formation stage was the application stage when the teacher presented contextual problems in the SSI context, particularly those comprising social problems in society, that closely linked with the understanding of salt hydrolysis and buffer solution concepts. Such problems included 1) the use of bleaching agents (detergents), 2) the functions of alum $KAl(SO_4)_2{\cdot}12H_2O$ for water purification, 3) the harmful effects of detergent waste, 4) the beneficial and harmful effects of artificial fertilizer $(NH_4)_2SO_4$ for soil fertility, and 5) the harmful effects of monosodium glutamate (MSG) for health. In this stage, the teacher guided the students through collaborative discussions and critical questions, intending to give them perspectives on SSI phenomena and encourage them to collect information and do experiments following student activity sheets. Thereupon, the students had presentation and discussion sessions, during which they reported their experiment results and drew conclusions [61, 62]. The teacher asked them to describe the possible problems and solutions from their understanding of the studied concepts. This enabled the students to form their conceptual understanding that is closely related to contexts; the learning process was from contextual to abstract [37, 63]. From such a condition, the teacher led the students to apply their knowledge in different contexts and situations and solve problems. The final stage was closing or teacher assistance in guiding the students to explain the conclusion and reflection on the learning process as the end of the learning activities.

2.3. Instrument

Table 1 displays 15 items of diagnostic three-tier multiple choice test to measure students' hydrolysis conceptual understanding. The test was constructed following the Competence Standard of 2013 Chemistry Curriculum of Senior High School under Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 37 of 2018. The procedures of developing the instrument followed the recommendation by [64, 65, 66].

Each item was designed in three questions (Q1, Q2, Q3) that integrated diagnostic [68, 69] and summative measurements [10] and certainty of response index (CRI) [70, 71]. Students' responses to items (Q1, Q2, Q3) were evaluated based on the rubric (Table 2). For example, students' responses to items were as follows: Q1, Q2 "correct", and Q3 "very sure" under the code CCC. Such a code indicated that students' conceptual understanding was in level 6, category of Scientific Knowledge (SK). On the other hand, if the response patterns in Q1, Q2 "incorrect" and Q3 "not sure", the code would be IIU, implying that students' conceptual understanding was in the category of Lack of Knowledge (LOK), or level 1. This instrument had been validated from the aspects of item conformity with the construct variable and language. The validity results by three experts were stated under Fleiss' kappa (K = .96), meaning that the experts agreed that the item validity was categorized good.

2.4. Data collection and analysis

Before the intervention, this research underwent pre-test data collection; whereas, the post-test data collection was done after the intervention. The construction of pre- and post-test items was the same. Students wrote down their responses on the provided answer sheet. Both tests were supervised by teachers in the school. The students must work on all items according to the allocated time (45 min). The instrument was immediately collected and should have the same number as the total participants.

The pre- and post-test measurement data were still ordinal data. The Rasch Partial Credit Model with WINSTEPS 4.5.5 software [27, 73] was used to convert ordinal data into interval data to have the same logit scale. The result was a data calibration of the levels of student's ability and item difficulty in the same interval.

The stacking analysis technique put pre-test and post-test data vertically [74]; meanwhile, the items appeared once in the experimental and control groups, allowing the researchers to check out any changes of the students after the intervention [56]. The examination was based on the same item, making the changes in students' ability during the pre- and post-test be measured [56]. Hence, each student created two measures of abilities, namely pre-test and post-test, and one measure for each item.

Table 1. Conceptual map of hydrolysis concept understanding [67].

Problem Context	Item	Conceptual Understanding	Ability Level	
Bleaching agents are formed of weak acid HOCl and strong base NaOH. Sodium hypochlorite salt (NaOCl) is reactive and dissolves the dye. In the water, the ion OCl^- will be hydrolyzed to HOCl and OH^- ˉ	1	Balancing the salt (NaOCl) hydrolysis reaction in the water	2	Level 3: Students are able to calculate the pH of the hydrolyzed salt solution. Level 2: Students are able to determine the hydrolysis reaction from different types of salt Level 1: Students are able to analyze the properties of the hydrolyzed salt
	2	Stating the partial hydrolysis reaction: $NaOCl \rightarrow Na^+ + OCl^-$	2	
	3	Determining corrosive alkali of sodium hypochlorite salt (NaOCl)	1	
	4	Calculating the pH of hydrolysis of sodium hypochlorite salt (NaOCl) with NaOCl = 0.1 M; $Ka = 10^{-5}$)	3	
	5	Determining the property of NaOCl, in the reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$	2	
	6	Calculating the pH of sodium hypochlorite salt (NaOCl) that comes from a mixture of HOCl and NaOH (partially hydrolyzed), if the Ka HOCl is 10^{-5} and there is an increase in the pH of the solution mixture.	3	
Water purification with alum $KAl(SO_4)_2{\cdot}12H_2O$ is the concept of salt hydrolysis, formed of H_2SO_4 and $Al(OH)_3$.	7	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water	1	
	8	Determining aluminum salt ($Al_2(SO_4)_3$) properties in the water that is partially hydrolyzed by the Al^{3+} ion	1	
The sodium tripolyphosphate (STPP) in detergents can pollute the environment, a eutrophication process.	9	Determining the properties of detergent solution causing eutrophication	1	
	10	Determining the properties of detergent solution (sodium tripolyphosphate salt) that is partially hydrolyzed	1	
	11	Determining the impact of the disposal of detergent waste on the environment	2	
ZA fertilizer $(NH4)_2SO_4$ is an acidic salt.	12	Determining the properties of ammonium sulfate salt $(NH4)_2SO_4$	1	
	13	Stating the equation of $(NH_4)_2SO_4$ reaction in the water, partially hydrolyzed	2	
Monosodium glutamate ($C_5H_8NO_4Na$) is L-glutamic acid salt, adversely impactful on human health	14	Students' attitude towards the use of monosodium glutamate ($C_5H_8NO_4Na$)	2	
	15	Determining the properties of monosodium glutamate salt ($C_5H_8NO_4Na$)	1	

Table 2. All possibilities of responses [70, 71, 72].

(Q1)	(Q2)	(Q3)	Code	Conceptual Understanding Category	Level
Correct	Correct	Certain	CCC	Scientific Knowledge (SK)	6
Correct	Incorrect	Certain	CIC	Misconception False Positive (MFP)	5
Incorrect	Correct	Certain	ICC	Misconception False Negative (MFN)	4
Incorrect	Incorrect	Certain	IIC	All-Misconception (ALM)	3
Correct	Correct	Uncertain	CCU	Lack of Confidence/Lucky Guess. (LG)	2
Correct	Incorrect	Uncertain	CIU	Lack of Knowledge (LOK)	1
Incorrect	Correct	Uncertain	ICU	Lack of Knowledge (LOK)	1
Incorrect	Incorrect	Uncertain	IIU	Lack of Knowledge (LOK)	1

The research hypothesis is that the students' conceptual understanding from pre-test to post-test changes, both in the experimental and control groups.

Conversely, the racking analysis technique put both pre- and post-test data horizontally, in which each item appeared twice in data collection, and students' ability only emerged once. This enabled the researchers to check out the effects of learning implementation on each student's ability from the tests, especially the changes in item difficulty levels before and after the intervention [56].

3. Results

3.1. Rasch analysis properties of instrument

The summary of changes in concepts and items analyzed by the Rasch model is presented in Table 1. Table 2 provides the item fit statistic. An item is considered to experience a misfit if the measurement result is not in line with the following three criteria: Outfit mean-square residual (MNSQ): $.5 < y < 1.5$; Outfit standardized mean-square residual (ZSTD): $-2 < Z < +2$; and point measure correlation (PTMEA CORR): $.4 < x < .8$ [25]. All items comply with the Outfit MNSQ criterion; item 15 does not meet the Outfit MNSQ criterion; five items (item 1, 6, 12, 13, and 15) are not in accordance with the Outfit (ZSTD) criterion; all items meet the PTMEA CORR criterion. Simply put, all items fulfill those criteria mentioned previously (none having a misfit), and are fit and valid.

This instrument has a good unidimensionality (Appendix 1). Raw variant index arrives at above the standard of 20% (33.9%), indicating that the instrument can effectively measure students' understanding of the hydrolysis concept [75]. The assessment scale analysis (Appendix 2) informs that the observation mean starts from logit -1.73 for category 1 (LOK) to logit +1.76 (category 6, SK). This signifies that the category of students' understanding takes place consistently [27]. In addition, the high item separation index (logit 6.71) and the high item reliability (logit .98) (Table 3) indicate that the respondents (students) are sufficient to confirm the level of item difficulty, strengthening the instrument construct validity [27]. The higher the item separation and reliability index, the more confident the researchers are about replicating item placement in other suitable sample students [25, 27]. Person separation index and person reliability that reach logit 2.0 and logit .75 (Table 4), respectively, imply that the instrument is quite sensitive to differentiate the high and low abilities of the students [25, 27]. According to the Rasch model calculation, the coefficient of Cronbach Alpha of logit .81 (Table 4) reflects an interaction between 380 students and 15 items with an excellent category [24, 76]. In other words, the interaction between students and items is very significant. The instrument has an excellent internal psychometric consistency and is considered very reliable.

Table 3. Item statistics: misfit order.

Item	Difficult	Error	Outfit MNSQ	Outfit ZSTD	PTMEA CORR.
1	-.38	.05	1.36	2.87	.47
2	.20	.04	1.13	1.56	.49
3	-.36	.05	.91	-.79	.43
4	.33	.04	1.09	.77	.55
5	-.25	.05	.94	-.55	.56
6	.26	.04	1.20	2.44	.41
7	.15	.04	.91	-1.17	.54
8	.47	.04	.90	-1.45	.44
9	-.47	.05	1.19	1.49	.46
10	.08	.04	1.09	1.04	.55
11	-.34	.05	1.04	.42	.51
12	-.06	.04	.71	-3.50	.60
13	.46	.04	.74	-4.12	.55
14	-.36	.05	1.00	.77	.55
15	.26	.04	1.31	3.74	.47

3.2. The difference in students' understanding ability of hydrolysis concept

The result of the Mann-Whitney test (Table 5) brings out the fact that statistically, there is a significant difference in the results of pre-test ($U = 3459.000$), $p < 0.05$) and post-test ($U = 1723.000$, $p < 0.05$) among students in experimental and control groups. Further, the Wilcoxon test result (Table 6) shows that the results of pre-test and post-test of students in the experimental group ($Z = -8.076$) and the control group ($Z = -6.690$) at the significant level *(p)* < 0.05 are significantly different. This suggests that students' understanding of the hydrolysis concept after the intervention (post-test) is higher than before the intervention (pre-test), both in experimental and control groups. However, the abilities of students in the experimental group are better than those in the control group. Accordingly, the learning process with the POGIL in the SSI context is better than the conventional learning.

3.3. The changes in students' understanding ability of the hydrolysis concept

From the different changes in pre- and post-test (Table 7), students in the experimental and control groups have improved their understanding of the hydrolysis concept. The experimental group's mean of pre-test and post-test is logit .51 (S.E = logit .21) and logit 1.50 (S.E = logit .32), respectively, with the mean difference of both tests is (logit .99). In contrast, the mean of pre-test and post-test of the control group gets logit .26 (S.E = logit .20) and logit .87 (S.E = logit .26), respectively, with the mean difference of pre- and post-test is logit .61. Such differences indicate different effects of interventions in the experimental and control group.

If the pre-test and post-test results of the experimental group are plotted in pairs (Figure 1), so that the mean difference in the sample pre- and post-test (logit +.99) is displayed as an intercept on the horizontal axis with the plotted slope = 1, several facts obtained: First, two lines that form the upper and lower curves separate 66 students around the empirical plot line, in which the pre-test and post-test mean is not significantly different from the mean difference in the pre- and post-test in the experimental group. Second, above the curve, 23 students experience significant changes; the mean of pre- and post-test is greater than the mean difference in sample pre-test and post-test. Third, seven students do not change, and ten students have negative changes (under the curve), so that they are under the curve. Similarly, the results of pre- and post-test of the control group (Figure 2) show that 53 students are around the empirical plot line; the abilities of 25 students change significantly (greater than the mean of sample pre- and post-test (logit +.61); two students do not change; 13 students experience negative changes in abilities. The difference in the plotting of pre-test and post-test results signifies different effects of interventions in the experimental and control groups.

Table 5. The result of the Mann-Whitney U test based on students' pre-test and post-test abilities in experimental and control groups ($p < 0.05$).

Test	Experimental Group (N = 97)	Control Group (N = 93)	U	p
Pre-test	0.5026 (-0.57–1.26)[a]	0.3029 (-1.61–1.03)[a]	3459.000	0.005
Post-test	1.1722 (-0.09–3.00)[a]	0.7052 (-1.06–1.47)[a]	1723.000	0.000

Table 6. The result of the Wilcoxon test of students' pre-test and post-test in experimental and control groups ($p < 0.05$).

Group	Pre-test	Post-test	Z	p*
Experimental	0.5026 (-0.57–1.26)[a]	1.1722 (-0.09–3.00)[a]	-8.076	0.000
Control	0.3029 (-1.61–1.03)[a]	0.7052 (-1.06–1.47)[a]	-6.690	0.000

Table 7. Logit of mean of pre- and post-test items of experimental and control groups.

Group	Student	Item	Mean/SE (logit)		
			Pre-test	Post-test	Pre- and Post-test Difference
Experimental	97	15	.51/(.21)	1.50/(.32)	.99
Control	93	15	.26/(.20)	.87/(.24)	.61

Description: SE = Standard Error.

3.4. The changes in item difficulty level

Table 8 presents the results of the racking analysis in connection with the changes in item difficulty level in the pre- and post-test of experimental and control groups. It is shown that in terms of item difficulty level, the mean of pre-test of the experimental group is (logit .32), the mean of post-test is (logit -.34), and the mean difference of the pre- and post-test is (logit -.66). Moreover, the mean of pre-test of the control group is (logit .25), the mean of post-test is (logit -.25), and the mean difference of the pre- and post-test is (logit -.50). This research also finds out that seven items have significant changes in the item difficulty level in the experimental group, lower than the pre- and post-test mean difference of (logit -.66), namely item 1, 2, 5, 7, 9, and 11. Eight items with a difficulty level greater than the mean are item 3, 4, 6, 8, 12, 13, 14, and 15. Item 10 has the same difficulty level as the mean. In the control group, eight items change significantly or less than the pre- and post-test mean difference of (logit -.50), including item 2, 3, 4, 5, 9, 11, 12, and 14; five items (item 1, 6, 7, 8, 10, 13) are greater than the mean; one item

Table 4. Person separation and reliability statistics.

Parameter	Measure	SD	Separation	Reliability	INFIT		OUTFIT		KR-20
					MNSQ	ZSTD	MNSQ	ZSTD	
Person (N = 380)	.67	.52	1.72	.75	1.00	.04	1.02	.10	.81
Item (N = 15)	.00	.32	6.71	.98	1.07	.41	1.02	-.01	

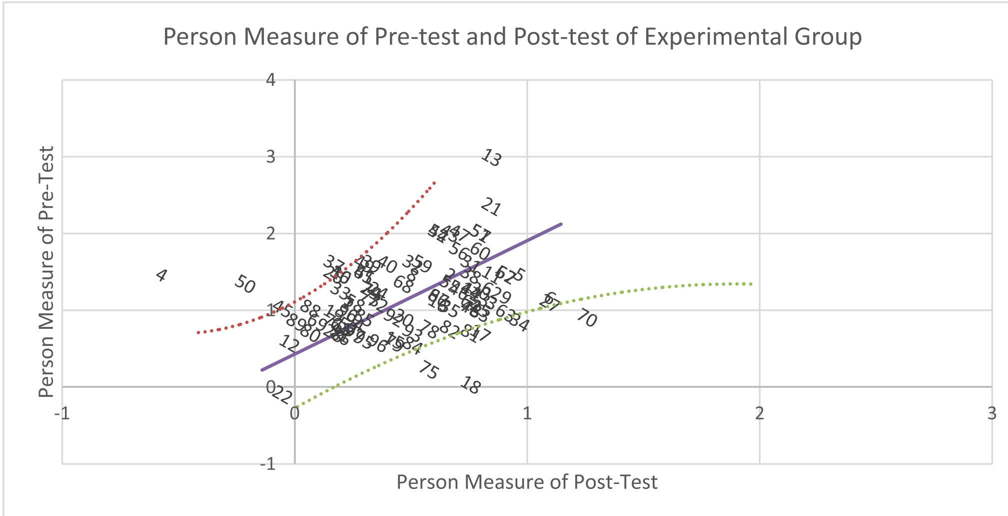

Figure 1. Scatter plots of person measures in pre- and post-test of the experimental group.

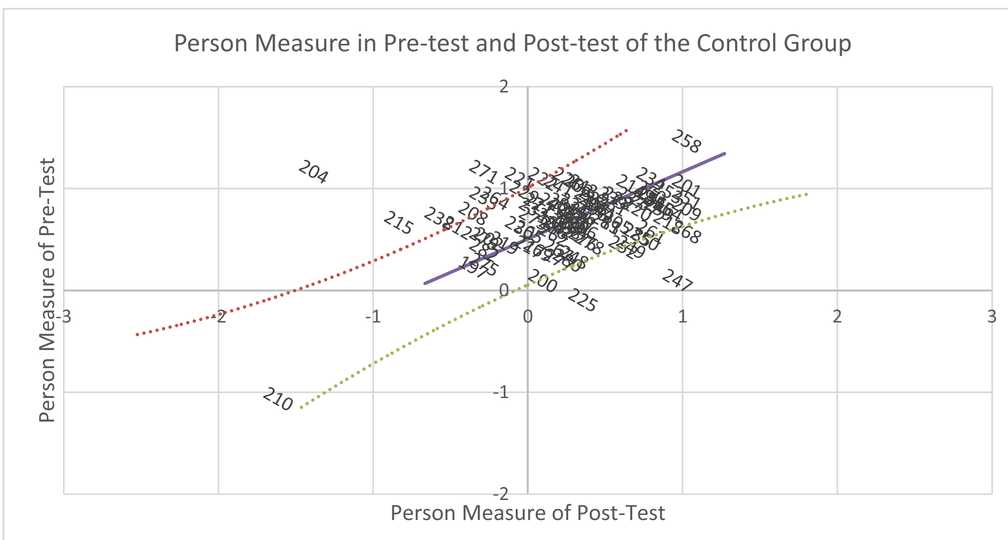

Figure 2. Scatter plots of person measures in pre- and post-test of the control group.

(item 15) has negative changes or becomes more difficult. The most difficult item in the experimental group is item 1 (.80 logit) and the easiest one is item 14 (logit -.10). Meanwhile, the most difficult item in the control group is item 13 (logit .64), and item 3 (logit -.15) is the easiest one. These findings indicate differences in the item difficulty level changes between students taught by the POGIL in the SSI context and the conventional model.

3.5. Conceptual changes in students' ability and item difficulty levels

Apart from the effect of learning interventions, there are three other factors that tend to influence the changes in students' ability and item difficulty levels, as follows: 1) guessing which happened to be correct or (lucky guess), 2) cheating, 3) carelessness. These factors can be identified from the student's item response pattern using a scalogram. For instance, the response pattern of post-test item 7 for student 353, 375, and 170 (Table 9). These three students, in the seventh and eighth row from the left, cannot understand item 12 (logit -.06) and item 10 (logit .08). Meanwhile, they can correctly explain the more difficult item, i.e., item 7 (logit .15). This situation implies a lucky guess, which in fact, these students have higher post-test abilities than the item 7 logit. Next is a cheating indication in the response pattern of student 128, 129, 134, 137, and 146. Such an indication is initially detected from the same post-test mean (logit 1.61) and item response pattern. The last one is carelessness, e.g., student 110, 118, and 139 are considered to be careless as they cannot correctly explain the easy item 4 (logit .33), yet can accurately understand item 13 (logit .46), which is harder than item 4. Moreover, they get very high post-test abilities.

3.6. Negative changes

Negative changes in conceptual understanding are detected from the changes in students' post-test logit less than the pre-test logit. For example, two students from the experimental group (E18 and E75) and the control group (C225 and C247) are taken; they have negative changes (Table 10). This means that these four students experience decreased abilities after the intervention. The pre-test item mean and the post-item mean of student E18 are (logit .76) and (logit .04), sequentially, with the mean difference of pre- and post-test arriving at (logit -.72). Moreover, the pre- and post-test item standard errors of student E18 are (logit .22) and (logit .18), respectively, with the combined standard error of logit .40. On account of the higher combined

Table 8. Data of item measures of pre- and post-test of experimental and control groups.

Item	Experimental (Mean)			Control (Mean)		
	Pre-test	Post-test	Difference Pre- and Post-test	Pre-test	Post-test	Difference Pre- and Post-test
Item1	.16	-1.00	-1.16	-.06	-.76	-.7
Item2	.80	.01	-.79	.39	-.40	-.79
Item3	.20	-.63	-.43	-.15	-.83	-.68
Item4	.62	.25	-.37	.54	.02	-.52
Item5	.14	-.78	-.92	.10	-.49	-.59
Item6	.26	.22	-.04	.41	.30	-.11
Item7	.66	-.33	-.99	.33	-.06	-.39
Item8	.59	.45	-.14	.49	.47	-.02
Item9	-.04	-.85	-.81	-.08	-.93	-.85
Item10	.40	-.26	-.66	.32	-.01	-.33
Item11	.13	-.91	-1.04	.05	-.78	-.83
Item12	.33	-.23	-.56	.25	-.51	-.76
Item13	.77	.16	-.61	.64	.33	-.31
Item14	-.10	-.80	-.7	.15	-.83	-.98
Item15	.25	-.40	-.65	.39	.72	.33
Mean	.32	-.34	-.66	.25	-.25	-.50

Table 9. Scalogram.

```
GUTTMAN SCALE OF RESPONSES:
Person |Item
       |  11 11   1 1   ID       Pre- Post-test  Pre Post   Item
       |913415207265438 Person   Mean    Mean   Difference  Response Pattern
       |---------------
   353 +666555536665554 353MFCB   .8    .97     .17 ┐ Lucky Guess
   375 +166566516133664 375MMCB  -.28   .40     .68 ├ Guessing answer
   170 +664666446566556 170NFEB   .33  1.17     .84 ┘ accidentaly correct

   128 +666666666646555 128DFEB   .76  1.61     .85 ┐
   129 +666666666646555 129DFEB   .51  1.61    1.10 │ Same response pattern
   134 +666666666646555 134JFEB   .17  1.61    1.44 ├ Cheating indication
   137 +666666666646555 137MMEB   .04  1.61    1.21 │
   146 +666666666646555 146NFEB   .30  1.61    1.31 ┘

   110 +666666666666566 110NFEB   .85  3.00    2.15 ┐ Response pattern
   118 +666666666666565 118RFEB   .85  2.36    1.51 ├ "Careless"
   139 +666666665666565 139MFEB   .62  2.01    1.39 ┘
```

Table 10. Scalogram results of student E18, E75, C225, and 247.

ID Person	Test	Item Response Pattern	Mean			
		| 11 11 1 1 |913415207265438 |---------------	Item Logit	S.E* Logit	Pre- test and post-test difference	Combined S.E
E18	Pre-test	+665666636366333	.76	.22	-.72	.40
	Post-test	+666661322521161	.04	.18		
E75	Pre-test	+562664552566426	.58	.20	-.35	.38
	Post-test	+655664322323463	.23	.18		
C225	Pre-test	+616665663261613	.36	.19	-.45	.37
	Post-test	+611622566131613	-.09	.18		
C247	Pre-test	+663636666666435	.97	.25	-.87	.43
	Post-test	+563345555314133	.10	.18		

Description: S.E = Standar Error

standard error than the pre- and post-test measures, the ability of student E18 in both tests is not significantly different. This also applied to student E75, C225, and C247.

4. Discussion and conclusion

The findings show changes in students' understanding abilities of the hydrolysis concept and items after the intervention. From the pre- and post-test mean difference, the experimental group has better positive changes than the control group [58]. In addition to the effect of the intervention, there is another factor contributing to the positive conceptual changes mentioned above, in terms of students' ability and item difficulty levels [24, 58]. The factor refers to some students who "accidently" give a correct response pattern (in the post-test). Even so, both groups have also experienced negative changes, implying that the intervention is specifically responded by students on account of the carelessness factor or a misconception-comprising response pattern [56, 58, 77]. Regarding this, not all learning objectives of the hydrolysis concept match the approach of POGIL in the SSI context. Negative changes of the students are because they are not epistemologically involved in the learning process, particularly in the observing, measuring, and calculating stages. These activities are interrelated up to group discussions as part of the stages of conceptual formation based on empirical facts [78]. Students are expected to explain and link the concepts they have learned following their epistemological reasoning [16, 79].

Furthermore, the interpretation of changes due to pedagogical interventions is exemplified by four students (Table 8) in item 5. In the pre-test, the ability of student E18 (logit .76), student E75 (logit .58), student C225 (logit .36), and student C247 (logit .96) is greater. They also respond to item 5 (-.25 logit) accurately. However, in the post-test item 5, the response of student E18, E75, C225, and C247 is incorrect due to their decreased post-test abilities. Therefore, the pre- and post-test mean difference is lower than item 5. Why do these changes occur? Such changes are exemplified by the response pattern of student E18 in item 5. This item measures students' ability in determining the reaction of NaOCl reaction: $OCl^- + H_2O \rightarrow HOCl + OH^-$, with the estimated pH = 7 and is alkaline. The question (Q1) of this item is, "is it correct that NaOCl is alkaline?". E18 answers "correct" in the pre-test, yet responds to "incorrect" in the post-test. The question (Q2) of this item is "what is your consideration for your answer in the Q1?". Four options are provided: (a) because NaOCl is formed of strong acids and weak bases; (b) because NaOCl is formed of weak acids and strong bases; (c) because NaOCl is formed of weak acids and weak bases; (d) because NaOCl is formed of strong acids and strong bases. In the pre-test, E18 chooses the correct answer (b), yet selects the incorrect answer (a) in the post-test that comprises misconception. Next, in the Q3 of this item, E18 chooses "very sure" in the pre-test and "not sure" in the post-test. The item 5 response pattern of E18 becomes CCC (category of scientific knowledge - SK) in the pre-test and IIU (category of lack of knowledge - LOK) in the post-test. Accordingly, the response pattern changes from CCC to IIU. The pre- and post-test mean difference of E18 (logit -.72) lower than item 5 (-.25) signifies that the error of response pattern results from misconception. This also applies to the response pattern of E75 (logit -.35), C225 (logit -.45), and C247 (logit -.87).

The misconception refers to the inability to identify the NaOCl salt hydrolysis that is formed of weak acids and strong bases. In short, the four students tend to not understand the concept of acid and base and acid-base reaction. These findings strengthen several previous studies [51, 53, 54, 80]. A study on the understanding of the acid-base concept of senior high school students in Malaysia concludes that some students have little understanding of the function of detergents as the cleaning agent, the difference between strong acids and strong bases, and the treatment for soil acidity using fertilizers [53]. In the same tune, such little understanding is because they do not conceptualize acid-base strength as a property that arises from the interaction of many reaction factors [51]. Additionally, research on an alternative conception of salt hydrolysis among senior high school students contends that the concept of hydrolysis is challenging for the students [54]. They are usually able to state the acidity of a salt solution correctly, yet writing a chemical equation to explain such a phenomenon is a great challenge. Most of the alternative conceptions are identifiable, rooted in the misunderstanding of equilibrium process, acid and base, material structure and other basic problems, student tendency to use a wrong analogy, and the lack of laboratory practice.

This research findings and elaboration of negative changes (case E18) prove the advantages of the Rasch model, specifically its potential in linking the result of changes (pre- and post-test), the item difficulty level, and the content being measured [18]. Such information solely comes from the Rasch model-based stacking and racking analysis techniques. The stacking technique provides information regarding "who has changed"; in contrast, the racking technique offers information of "what has changed" [56, 58], allowing the researchers to spell out the effect of the applied pedagogical innovation [18, 33, 34]. Although the instrument measurement result of this work is not data-rich, the analysis strength of the Rasch model can describe in detail the conceptual changes, both in the students' ability and item difficulty levels.

4.1. Limitations and further studies

The primary limitation of this research is that it did not take into account the aspects of learning style, culture, and motivation that can change due to learning interventions. Future studies, therefore, can address these aspects. The present study can be continued by considering the context of a problem that closely connects with the parameter of item difficulty level. The analysis will be more interesting if it can prove that different item difficulty levels are influenced by problem contexts in each item [81]. Further studies are also expected to find an analysis technique that can integrate problem contexts, item characteristics, and item difficulty levels in a measurement model. It is assumed that different problem contexts in each item will be more likely to affect measurement results because problem contexts have conceptual linkage with items and student activities in doing experiments, measuring, interpreting data/-graphs, and others. Thus, the linkages between the learning process during the intervention and conceptual changes in students' ability and item difficulty levels can be explained in detail; which part of the process leads the students to change their understanding related to specific ideas taught to them.

Declarations

Author contribution statement

Lukman Abdul Rauf Laliyo: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Bambang Sumintono: Analyzed and interpreted the data; Wrote the paper.

Citra Panigoro: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

We would like to express our deep gratitude to all parties for their assistance in data processing, especially the Head of Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo.

Appendix

Appendix 1. Standardized residual variance in eigen value units.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	22.7067	100.0%		100.0%
Raw variance explained by measures	=	7.7067	33.9%		35.9%
Raw variance explained by persons	=	2.7733	12.2%		12.9%
Raw Variance explained by items	=	4.9334	21.7%		23.0%
Raw unexplained variance (total)	=	15.0000	66.1%	100.0%	64.1%
Unexplned variance in 1st contrast	=	2.0698	9.1%	13.8%	
Unexplned variance in 2nd contrast	=	1.5312	6.7%	10.2%	
Unexplned variance in 3rd contrast	=	1.3696	6.0%	9.1%	
Unexplned variance in 4th contrast	=	1.3124	5.8%	8.7%	
Unexplned variance in 5th contrast	=	1.1945	5.3%	8.0%	

Appendix 2. Summary of category structure.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	317	6	-.18	-.20	1.06	1.08	NONE	(-1.73)	1
2	2	190	3	.10	.09	1.03	1.18	.46	-.77	2
3	3	963	17	.33	.31	1.02	.93	-1.43	-.22	3
4	4	542	10	.56	.52	1.02	.97	.98	.21	4
5	5	1262	22	.62	.73	1.27	.99	-.22	.74	5
6	6	2425	43	1.02	.98	.97	1.04	.20	(1.76)	6
MISSING		1	0	-.30						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

References

[1] S. Lu, H. Bi, Development of a measurement instrument to assess students' electrolyte conceptual understanding, Chem. Educ. Res. Pract. 17 (4) (2016) 1030–1040.

[2] H.E. Yildirir, H. Demirkol, Identifying mental models of students for physical and chemical change, J. Baltic Sci. Educ. 17 (6) (2018) 986–1004.

[3] C.R. Gette, M. Kryjevskaia, M.R. Stetzer, P.R.L. Heron, Probing student reasoning approaches through the lens of dual-process theories: a case study in buoyancy, Phys. Rev. Phys. Educ. Res. 14 (1) (2018) 10113.

[4] A.H. Johnstone, Chemical education research in Glasgow in perspective, Chem. Educ. Res. Pract. 7 (2) (2006) 49–63.

[5] A.H. Johnstone, You can't get there from here, J. Chem. Educ. 87 (1) (2010) 22–29.

[6] K.S. Taber, Chemical misconceptions—prevention, diagnosis, and cure, J. Chem. Educ. 80 (5) (2003) 491.

[7] K.S. Taber, Challenging misconceptions in the chemistry classroom: resources to support teachers, Educació Química EduQ. 4 (2009) 13–20.

[8] J.I. Alamina, I.S. Etokeren, Effectiveness of imagination stretch teaching strategy in correcting misconceptions of students about particulate nature of matter, J. Educ. Soc. Behav. Sci. 27 (1) (2018) 1–11.

[9] I.Z. Yaşar, E. İnce, F.G. Kırbaşlar, 7. Class Science and Technology Course "Structure of Atom" Subject Readiness Improvement Test, 152, Procedia Soc. Behav. Sci., 2014, pp. 662–667.

[10] K.Y. Hoe, R. Subramaniam, On the prevalence of alternative conceptions on acid-base chemistry among secondary students: insights from cognitive and confidence measures, Chem. Educ. Res. Pract. 17 (2) (2016) 263–282.

[11] D.C. Aktan, Investigation of students' intermediate conceptual understanding levels: the case of direct current electricity concepts, Eur. J. Phys. 34 (1) (2013) 33–43.

[12] M. Allen, Misconceptions in Primary Science, Open University Press, Maidenhead, 2014.

[13] Soeharto, B. Csapó, E. Sarimanah, F.I. Dewi, T. Sabri, A review of students' common misconceptions in science and their diagnostic assessment tools, J. Pendidik. IPA Indones. 8 (2) (2019) 247–266.

[14] S. Almuntasheri, R.M. Gillies, T. Wright, The effectiveness of a guided inquiry-based, teachers' professional development programme on saudi students' understanding of density, Sci. Educ. Int. 27 (1) (2016) 16–39.

[15] M.Z. Hashweh, The complexity of teaching density in middle school, Res. Sci. Technol. Educ. 34 (1) (2016) 1–24.

[16] K. Zvoch, S. Holveck, L. Porter, Teaching for conceptual change in a density unit provided to seventh graders: a comparison of teacher- and student-centered approaches, Res. Sci. Educ. 51 (2019) 1395–1421.

[17] R. Grob, M. Holmeier, P. Labudde, Formative assessment to support students' competences in inquiry-based science education, Interdiscip. J. Probl-based. Learn. 11 (2) (2017).

[18] T.C. Pentecost, J. Barbera, Measuring learning gains in chemical education: a comparison of two methods, J. Chem. Educ. 90 (7) (2013) 839–845.

[19] B. Sumintono, Rasch model measurements as tools in assesment for learning, in: Proceedings of the 1st International Conference on Education Innovation (ICEI 2017); 2017 October 14; Surabaya, Atlantis Press, Paris, 2018.

[20] R. Micelli, M. Settanni, G. Vidotto, Measuring change in training programs : an empirical illustration Measurement of change and Item Response Theory, Psychol. Sci. Q. 50 (3) (2008) 433–447.

[21] S.D. Willoughby, A. Metz, Exploring gender differences with different gain calculations in astronomy and biology, Am. J. Phys. 77 (7) (2009) 651–657.

[22] R.L. Linn, J.A. Slinde, The determination of the significance of change between pre- and posttesting periods, Rev. Educ. Res. 47 (1) (1977) 121–150.

[23] P. He, X. Liu, C. Zheng, M. Jia, Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons, Chem. Educ. Res. Pract. 17 (2) (2016) 381–393.
[24] B. Sumintono, W. Widhiarso, Aplikasi Pemodelan Rasch Pada Assessment Pendidikan [Application of Rasch Modeling in Educational Assessment], Bandung: Penerbit Trim Komunikata, 2015. https://www.researchgate.net/publication/282673464%0AAplikasi.
[25] W.J. Boone, M.S. Yale, J.R. Staver, Rasch Analysis in the Human Sciences, Springer, Dordrecht, 2014.
[26] X. Liu, Developing measurement instruments for science education research, in: B. Fraser, K.G. Tobin, C.J. McRobbie (Eds.), Second International Handbook of Science Education, Springer, Netherlands, 2012, pp. 651–665.
[27] J.M. Linacre, 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. US, Winsteps, 2020.
[28] S. Sabah, H. Hammouri, M. Akour, Validation of a scale of attitudes toward science across countries using Rasch model: findings from TIMSS, J. Baltic Sci. Educ. 12 (5) (2013) 692–703.
[29] J.C. Hadenfeldt, S. Bernholt, X. Liu, K. Neumann, I. Parchmann, Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter, J Chem. Educ. 90 (12) (2013) 1602–1608.
[30] C.F. Herrmann-Abell, G.E. DeBoer, Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items, Chem. Educ. Res. Pract. 2 (2) (2011) 184–192.
[31] C.F. Herrmann-Abell, G.E. Deboer, Using Rasch modeling and option probability curves to diagnose students' misconceptions, in: The 2016 AERA Annual Meeting; 2016 April 8-12, American Educational Research Assossiation, Washington DC. USA, 2016. Available from: https://www.aaas.org/sites/default/files/Project2061_CHA-AERA%20energy%20paper%204-7-16.pdf.
[32] L.A.R. Laliyo, J.S. Tangio, B. Sumintono, M. Jahja, C. Panigoro, Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter, J. Baltic Sci. Educ. 19 (5) (2020).
[33] B. Davidowitz, M. Potgieter, Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry, Int. J. Sci. Educ. 38 (9) (2016) 1483–1503.
[34] M. Park, X. Liu, An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics, Res. Sci. Educ. (2019).
[35] D.F. Treagust, S.S. Qureshi, V.R. Vishnumolakala, J. Ojeil, M. Mocerino, D.C. Southam, Process-oriented guided inquiry learning (POGIL) as a culturally relevant pedagogy (CRP) in Qatar: a perspective from grade 10 chemistry classes, Res. Sci. Educ. 1–19 (2018).
[36] A.T. Kinslow, T.D. Sadler, H.T. Nguyen, Socio-scientific reasoning and environmental literacy in a field-based ecology class, Environ. Educ. Res. 4622 (2018) 1–23.
[37] D.C. Owens, T.D. Sadler, P. Friedrichsen, Teaching practices for enactment of socio-scientific issues instruction: an instrumental case study of an experienced biology teacher, Res. Sci. Educ. (2019).
[38] R. Bruder, A. Prescott, Research evidence on the benefits of IBL, ZDM Int. J. Math. Educ. 45 (6) (2013) 811–822.
[39] D.L. Zeidler, Socioscientific issues as a curriculum emphasis: theory, research and practice, in: N.G. Lederman, S.K. Abell (Eds.), Handbook of Research on Science Education, Routledge, New York, 2014, pp. 697–726.
[40] A.G. Espeja, D.C. Lagarón, Socio-scientific Issues (SSI) in initial training of primary school teachers: pre-service teachers' conceptualization of SSI and appreciation of the value of teaching SSI, Procedia Soc. Behav. Sci. 196 (2015) 80–88.
[41] National Research Council, A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas, The National Academies Press, Washington DC, 2012.
[42] T.D. Sadler, D.L. Zeidler, Scientific literacy, PISA, and socioscientific discourse: assessment for progressive aims of science education, J. Res. Sci. Teach. 46 (8) (2009) 909–921.
[43] N.G. Lederman, J.S. Lederman, A. Antink, Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy, Int. J. Educ. Math. Sci. Technol. 1 (3) (2013) 138–147.
[44] A.N. Cooke, K.S. Fielding, W.R. Louis, Environmentally active people: the role of autonomy, relatedness, competence and self-determined motivation, Environ. Educ. Res. 22 (5) (2016) 631–657.
[45] T.D. Sadler, Moral and ethical dimensions of socioscientific decision-making as integral components of scientific literacy, Sci. Educ. 13 (1) (2004) 39–48.
[46] G.E. DeBoer, Scientific literacy: another look at its historical and contemporary meanings and its relationship to science education reform, J. Res. Sci. Teach. 37 (6) (2000) 582–601.
[47] T.D. Sadler, Socio-scientific issues-based education: what we know about science education in the context of SSI, in: T.D. Sadler (Ed.), Socio-Scientific Issues in the Classroom: Teaching, Learning and Research, Springer, Dordrecht, 2011, pp. 355–369.
[48] W. Gräber, P. Nentwig, H.J. Becker, E. Sumfleth, A. Pitton, K. Wollweber, D. Jorde, Scientific literacy: from theory to practice, Res. Sci. Educ. Past. Present. Future. 1996 (2005) 61–70.
[49] J. Grooms, A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction, J. Chem. Educ. (2020).
[50] M.L. Presley, A.J. Sickel, N. Muslu, D. Merle, A framework for socio-scientific issues based education, Sci. Educ. 22 (2013) 26–32.
[51] H. Tümay, Emergence, learning difficulties, and misconceptions in chemistry undergraduate students' conceptualizations of acid strength, Sci. Educ. 25 (1–2) (2016) 21–46.
[52] N. Seçken, Identifying student's misconceptions about SALT, Procedia Soc. Behav. Sci. 2 (2) (2010) 234–245.
[53] M.I.M. Damanhuri, D.F. Treagust, M. Won, A.L. Chandrasegaran, High school students' understanding of acid-base concepts: an ongoing challenge for teachers, Int. J. Environ. Sci. Educ. 11 (1) (2016) 9–27.
[54] K. Orwat, P. Bernard, A. Migdał-Mikuli, Alternative conceptions of common salt hydrolysis among upper-secondary school students, J. Baltic Sci. Educ. 16 (1) (2017) 64–76.
[55] J.W. Creswell, Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, fourth ed., Pearson Education, Inc, New York, 2012.
[56] B.D. Wright, Rack and stack: time 1 vs. time 2 or pre-test vs. post-test, Rasch Meas. Trans. 17 (1) (2003) 905–906.
[57] P. Anselmi, G. Vidotto, O. Bettinardi, G. Bertolotti, Measurement of change in health status with Rasch models, Health Qual. Life Outcome 13 (1) (2015) 1–7.
[58] M. Ling, V. Pang, C.C. Ompok, Measuring change in early mathematics ability of children who learn using games: stacked analysis in rasch measurement, in: Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings, 2016 July 30-August 3; Singapore, Springer, Singapore, 2018.
[59] K.S. Taber, Ethical considerations of chemistry education research involving "human subjects", Chem. Educ. Res. Pract. 15 (2) (2014) 109–113.
[60] F. Aldresti, S. Rahayu, F. Fajaroh, The influence of inquiry-based chemistry learning with the contex of socio-scientific issues on high school students' scientific explanation skills, J. Pendidik. IPA. 23 (2) (2019).
[61] Ibnu S. Mitarlis, S. Rahayu, Sutrisno, The effectiveness of new inquiry-based learning (NIBL) for improving multiple higher-order thinking skills (M-HOTS) of prospective chemistry teachers, Eur. J. Educ. Res. 9 (3) (2020) 1309–1325.
[62] M. Pedaste, M. Mäeots, L.A. Siiman, T. de-Jong, S.A.N. Van-Riesen, E.T. Kamp, C.C. Manoli, Z.C. Zacharia, E. Tsourlidaki, Phases of inquiry-based learning: definitions and the inquiry cycle, Educ. Res. Rev. 14 (2015) 47–61.
[63] W.L. Romine, T.D. Sadler, Measuring changes in interest in science and technology at the college level in response to two instructional interventions, Res. Sci. Educ. 46 (3) (2016) 309–327.
[64] M. Wilson, Constructing Measures: an Item Response Modeling Approach, Lawrence Erlbaum Associates, USA, 2005.
[65] M. Wilson, Cognitive diagnosis using item response models, Zeits Für Psych/J. Psych. 216 (2) (2008) 74–88.
[66] M. Wilson, Measuring progressions: assessment structures underlying a learning progression, J. Res. Sci. Teach. 46 (6) (2009) 716–730.
[67] L.A.R. Laliyo, Mendiagnosis Sifat Perubahan Konseptual Siswa: Penerapan Teknik Analisis Stacking Dan Racking Rasch Model [Diagnosing the Nature of Students' Conceptual Change: Application of Stacking and Racking Analysis Techniques Rasch Model], Penerbit Deepublish, Yogyakarta, 2021.
[68] G. Chittleborough, D. Treagust, Correct interpretation of chemical diagrams requires transforming from one level of representation to another, Res. Sci. Educ. 38 (4) (2008) 463–482.
[69] D.F. Treagust, Development and use of diagnostic tests to evaluate students' misconceptions in science, Int. J. Sci. Educ. 10 (2) (1988) 159–169.
[70] H.O. Arslan, C. Cigdemoglu, C. Moseley, A Three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain, Int. J. Sci. Educ. 34 (11) (2012) 1667–1686.
[71] S. Hasan, D. Bagayoko, E.L. Kelley, Misconceptions and the certainty of response index (CRI), Phys. Educ. 34 (5) (1999) 294–299.
[72] H. Habiddin, E.M. Page, Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK), Indones. J. Chem. 19 (3) (2019) 720–736.
[73] T.G. Bond, C.M. Fox, Applying the Rasch Model: Fundamental Measurement in the Human Sciences, third ed., Routledge Taylor & Francis Group, New York, 2015.
[74] C.F. Herrmann-abell, J.C. Flanagan, J.E. Roseman, Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: using student measures to evaluate the promise of the intervention, in: Proceeding of the 2013 NARST Annual International Conference. 2013 April 6-9; Puerto Rico, NARST, Puerto Rico, 2013. Available from: https://www.aaas.org/sites/default/files/project2061/files/NARST2013-Flanagan-Paper5-Usingteachermeasurestoe.pdf.
[75] W.P. Fisher, Rating scale instrument quality criteria, Rasch Meas. Trans. 21 (1) (2007) 1095. www.rasch.org/rmt/rmt211m.htm.
[76] W.K. Adams, C.E. Wieman, Development and validation of instruments to measure learning of expert-like thinking, Int. J. Sci. Educ. 33 (9) (2011) 1289–1312.
[77] B. Sumintono, W. Widhiarso, Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [application of rasch model for social science studies], in: B. Trim (Ed.), Bandung, Trim Komunikata Publishing House, 2014. https://www.researchgate.net/publication/268688933%0AAplikasi.
[78] J.G. Rodriguez, K.H. Hunter, L.J. Scharlott, N.M. Becker, A review of research on process oriented guided inquiry learning: implications for research and practice, J. Chem. Educ. 97 (10) (2020) 3506–3520.
[79] T.S. Hancock, P.J. Friedrichsen, A.T. Kinslow, T.D. Sadler, Selecting socio-scientific issues for teaching: a grounded theory study of how science teachers collaboratively design SSI-based curricula, Sci. Educ. 28 (6–7) (2019) 639–667.
[80] N. Ültay, M. Çalik, A comparison of different teaching designs of "acids and bases" subject, Eurasia J. Math. Sci. Technol. Educ. 12 (1) (2016) 57–86.
[81] J. Hartig, A. Frey, G. Nold, E. Klieme, An application of explanatory item response modeling for model-based proficiency scaling, Educ. Psychol. Meas. 72 (4) (2012) 665–686.