BUKTI KORESPONDENSI

Judul : **Analytic Approach of Response Pattern of Diagnostic Test Items In Evaluating Students' Conceptual Understanding of Characteristics of Particle of Matter**

Jurnal : JBSE - **Journal of Baltic Science Education**, 19(5), 824-841. https://doi.org/10.33225/jbse/20.19.824

Gmail

Lukman Laliyo <lukman.laliyo019@gmail.com>

Re: Paper submission from

18 messages

Journal of Baltic Science Education <mail.jbse@gmail.com> Thu, Apr 9, 2020 at 10:30 PM
To: lukman.laliyo019@gmail.com

Dear author (s),

Your manuscript is received. Thank you very much that you are interested in publishing of your paper in the JBSE journal. You will be informed accordingly.

Sincerely yours,

Editor

2020-04-09, kt, 12:07 <scientia@scientiasocialis.lt> rašė:

> First name: Lukman
> Middle name: Abdul Rauf
> Last name: Laliyo
> Institutional affiliation: Univeristas Negeri Gorontalo
> E-mail: lukman.laliyo019@gmail.com
> Country: Indonesia
> Paper file:
> http://www.scientiasocialis.lt/jbse/sites/default/files/webform/papers/Template_JBSE_9.doc
> Declaration file:
> http://www.scientiasocialis.lt/jbse/sites/default/files/webform/declarations/JBSE_declaration_2020_1.pdf
> Additional authors: 4
> Additional author #1:
> Julhim S. Tangio
> E-mail: transbahasaid@gmail.com
> Master of Education, Assistant Professor, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Indonesia.
>
> Additional author #2:
> Bambang Sumintono
> Email: transbahasa.ilmiah@gmail.com
> PhD, Lecturer in Institute of Educational Leadership, Universiti Malaya, Kuala Lumpur, Malaysia,
>
> Additional author #3:
> Mohamad Jahja
> E-mail: transbahasatb@gmail.com
> Doctor, Associate Professor, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo, Indonesia.
>
> Additional author #4:
> Citra Panigooro
> E-mail: citrapanigoro25@gmail.com
> Master of Science, Assistant Professor, Faculty of Faculty of Fisheries and Marine Science, Universitas Negeri Gorontalo, Indonesia.
> Comments to Editor:
>
> Submitted on Thursday, April 9, 2020 - 05:06

--

Scientia Socialis Ltd., the Associated Member of the Lithuanian Publishers Association & Crossref, Lithuania.
In cooperation with SMC "Scientia Educologica", the Associated Member of Lithuanian Scientific Society, European Society for the History of Science (ESHS) and ICASE
Donelaicio Street 29,
LT-78115 Siauliai, Lithuania
Phone: +370 687 95668
Skype: scientia12
Home page: http://www.scientiasocialis.lt/jbse/
Facebook: https://www.facebook.com/ScientiaEducologica

Lukman Laliyo <lukman.laliyo019@gmail.com> Mon, Apr 13, 2020 at 1:28 PM
To: transbahasaid@gmail.com

Halo Admin,

This is Mimi. Kindly forward this e-mail to pka Lukman, ya.

Thank you!

---------- Forwarded message ---------
From: **Journal of Baltic Science Education** <mail.jbse@gmail.com>
Date: Thu, Apr 9, 2020 at 10:30 PM
Subject: Re: Paper submission from
To: <lukman.laliyo019@gmail.com>

Dear author (s),

Your manuscript is received. Thank you very much that you are interested in publishing of your paper in the JBSE journal. You will be informed accordingly.

Sincerely yours,

Editor

[Quoted text hidden]

Journal of Baltic Science Education <mail.jbse@gmail.com> Wed, May 20, 2020 at 3:55 PM
To: lukman.laliyo019@gmail.com

Dear author (s),

We send to you your manuscript with some remarks after the review process. Please make all corrections asap, but not later than **10 June 2020**. If you want to reject this paper please inform us asap. Your explanation / rebuttal letter should be added (Each comment from the referee, the author (s) should explicitly state whether they made a requested change or, if not, they must explain their reasons in detail).
Sincerely yours,

Editor
[Quoted text hidden]

---

3 attachments

200520.JBSE_review_Laliyo.pdf
154K

200520.jbse_JL-suo_reviewed_reevaluation.doc
1373K

ANALYTIC_APPROACH_OF_RESPONSE_PATTERN_OF_DIAGNOSTI.pdf
4894K

Lukman Laliyo <lukman.laliyo019@gmail.com> Wed, May 20, 2020 at 7:11 PM
To: nazmirifat@gmail.com

---------- Forwarded message ---------
From: **Journal of Baltic Science Education** <mail.jbse@gmail.com>
Date: Wed, May 20, 2020, 3:56 PM
Subject: Re: Paper submission from
To: <lukman.laliyo019@gmail.com>

[Quoted text hidden]

3 attachments

200520.JBSE_review_Laliyo.pdf
154K

200520.jbse_JL-suo_reviewed_reevaluation.doc
1373K

ANALYTIC_APPROACH_OF_RESPONSE_PATTERN_OF_DIAGNOSTI.pdf
4894K

Lukman Laliyo <lukman.laliyo019@gmail.com> Wed, Jun 10, 2020 at 4:04 PM
To: Journal of Baltic Science Education <mail.jbse@gmail.com>

Dear Editor(s),

Please kindly find the revised manuscript and the explanation/rebuttal letter attached.

Thank you and we look forward to hearing from you.

Sincerely yours,
Authors
[Quoted text hidden]

2 attachments

Explanation-Rebuttal Letter.docx
20K

[Revised] 200419.jbse_JL-suo_review.docx
1142K

Lukman Laliyo <lukman.laliyo019@gmail.com> Wed, Jun 10, 2020 at 4:06 PM
To: nazmirifat@gmail.com

---------- Forwarded message ---------
From: **Lukman Laliyo** <lukman.laliyo019@gmail.com>
Date: Wed, Jun 10, 2020 at 4:04 PM
Subject: Re: Paper submission from
[Quoted text hidden]
[Quoted text hidden]

2 attachments

Explanation-Rebuttal Letter.docx
20K

[Revised] 200419.jbse_JL-suo_review.docx
1142K

Lukman Laliyo <lukman.laliyo019@gmail.com> Wed, Jun 10, 2020 at 4:11 PM
To: lukmanarlaliyo@gmail.com

[Quoted text hidden]

2 attachments

Explanation-Rebuttal Letter.docx
20K

[Revised] 200419.jbse_JL-suo_review.docx
1142K

Journal of Baltic Science Education <mail.jbse@gmail.com> Wed, Jun 10, 2020 at 7:03 PM
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Dear author (s),

Your improved manuscript is received. Thank you very much for your efforts to improve the work. You will be informed accordingly.

Sincerely yours,

Editors
[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com> Thu, Jun 11, 2020 at 9:37 AM
To: nazmirifat@gmail.com

---------- Forwarded message ---------
From: **Journal of Baltic Science Education** <mail.jbse@gmail.com>
Date: Wed, Jun 10, 2020 at 7:04 PM
Subject: Re: Paper submission from
To: Lukman Laliyo <lukman.laliyo019@gmail.com>
[Quoted text hidden]

Journal of Baltic Science Education <mail.jbse@gmail.com> Wed, Jul 1, 2020 at 2:57 PM
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Dear author (s),

We send to you your manuscript with some remarks after the re-review process. Please make all corrections asap, but not later than **10 July 2020**. If you want to reject this paper please inform us asap. Your explanation / rebuttal letter should be added (Each comment from the referee, the author (s) should explicitly state whether they made a requested change or, if not, they must explain their reasons in detail).
Sincerely yours,

Editor
[Quoted text hidden]

200630.jbse_JL-suo_reevaluated.docx
1140K

Lukman Laliyo <lukman.laliyo019@gmail.com> Fri, Jul 10, 2020 at 12:34 PM
To: Journal of Baltic Science Education <mail.jbse@gmail.com>

Dear Editor(s),

Please kindly find the revised manuscript and the explanation/rebuttal letter attached.

Thank you and we look forward to hearing from you.

Sincerely yours,
Authors

[Quoted text hidden]

---

2 attachments

Explanation-Rebuttal Letter.docx
16K

[Revised] 200630.jbse_JL-suo_reevaluated.docx
1137K

Journal of Baltic Science Education <mail.jbse@gmail.com> Fri, Jul 10, 2020 at 2:17 PM
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Dear author (s),

Your improved manuscript is received. Thank you very much for your efforts to improve the work. You will be informed accordingly.

Sincerely yours,

Editors

[Quoted text hidden]

Journal of Baltic Science Education <mail.jbse@gmail.com> Tue, Jul 14, 2020 at 3:04 PM
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Dear author,

1. We want to inform you that your paper is preliminary accepted for publication in JBSE (Vol. 19, 2020).
2. According to JBSE requirements:

~~The author (authors) should confirm in writing (file in PDF or JPG format, including author`s signature), that the manuscript has not been published in another journal or handed over (transferred) to another journal for publication.~~
~~A template is attached. Please fill in, sign it and scan in PDF or JPG format and return back.~~ DONE

3. According to JBSE policy authors:

are required to pay the partial paper processing charges of **585** EUR (including postage & add.proofreading & 4 extra pages) in order to publish a paper in the journal (**excluding bank taxes/charges; it is a responsibility of payer. These charges are different from country to country**

and from bank to bank. Please ask at the bank how much the charge amount is in your case BEFORE to do the transfer).

The corresponding author will receive one hard copy of the journal via post.
P.S. If you want to get extra copy (-ies), please add **25** EUR for each copy to the indicated sum above + 6 EUR shipping expenses for each additional copy.

Options for Payment:

1 option – via bank transfer:

Payee (or beneficiary): Scientia Socialis, UAB
Identification number: **302614473**,
VAT: LT100006097614
Institution address: Donelaicio Street 29, Siauliai, Lithuania
Beneficiary Bank name: „SWEDBANK" AB
Beneficiary Bank address: Konstitucijos Street 20A, Vilnius, Lithuania.
SWIFT/BIC: HABALT22;
IBAN: **LT077300010126557188**

2 option – via PayPal system.

To: scientia@scientiasocialis.lt

https://www.paypal.com/lt/

10 EUR should be added to the total amount as a transfer fee (it is valid only in the case of PayPal).

*****

We kindly ask you to make a transfer not later than **30 July** 2020.

Sincerely yours,

Editorial Office

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com> Thu, Jul 16, 2020 at 9:44 AM
To: nazmirifat@gmail.com, Transbahasa <transbahasa.go@gmail.com>

---------- Forwarded message ---------
From: **Journal of Baltic Science Education** <mail.jbse@gmail.com>
Date: Tue, Jul 14, 2020 at 3:04 PM
Subject: Re: Paper submission from
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com> Fri, Jul 17, 2020 at 1:29 PM
To: Journal of Baltic Science Education <mail.jbse@gmail.com>

Dear Editorial Team of

Journal of Baltic Science Education

Please kindly find the attached file of publication payment receipt.

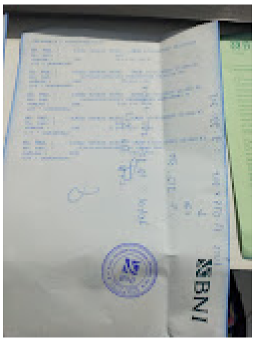
Thank you!

Best regards,

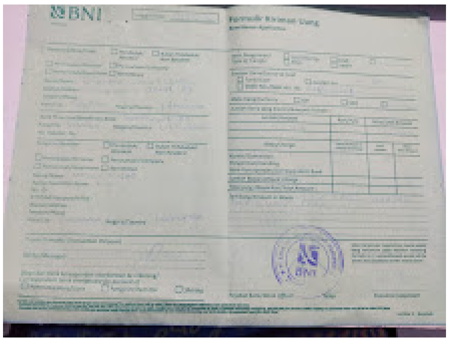
Authors

[Quoted text hidden]

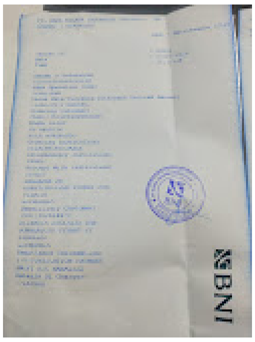
3 attachments



Payment Reciept 2.jpeg
82K



Payment Reciept 3.jpeg
107K



Payment Reciept 1.jpeg
86K

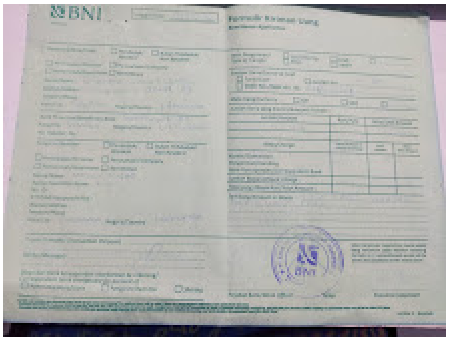
Lukman Laliyo <lukman.laliyo019@gmail.com> Mon, Jul 20, 2020 at 2:34 PM
To: Journal of Baltic Science Education <mail.jbse@gmail.com>, scientia@scientiasocialis.lt

[Quoted text hidden]

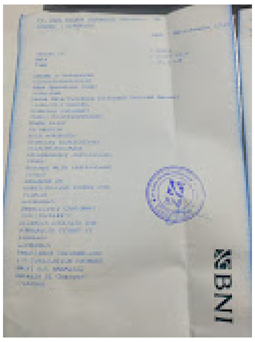
3 attachments

Payment Reciept 2.jpeg
82K

Payment Reciept 3.jpeg



107K



Payment Reciept 1.jpeg
86K

Journal of Baltic Science Education <mail.jbse@gmail.com> Mon, Jul 20, 2020 at 3:02 PM
To: Lukman Laliyo <lukman.laliyo019@gmail.com>
Cc: scientia@scientiasocialis.lt

Dear author,
Received, thank you.

Editor
[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com> Thu, Sep 3, 2020 at 6:04 PM
To: Journal of Baltic Science Education <mail.jbse@gmail.com>

Dear Editor,

We would like to informed regarding the article issued.
When would it be published?

Thank you!

Best Regards,
Authors

[Quoted text hidden]

JOURNAL OF BALTIC SCIENCE EDUCATION

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

ID: 200419.jbse_JL-suo_review

1.	Is the article original, and does it contribute something new to the field? (Importance of article / Relevance and Appeal to national / international scholarly community)	Excellent	<u>Good</u>	Moderate	Poor
2.	Do the title and abstract together give an adequate summary of the article / paper?	Excellent	<u>Good</u>	Moderate	Poor
3.	Statement of problem (s) / aim (s) / objective (s)	Excellent	Good	<u>Moderate</u>	Poor
4.	Theoretical basis / Theoretical framework / Literature review / Clarification of concepts	Excellent	Good	<u>Moderate</u>	Poor
5.	Appropriateness of the research plan and design (if applicable) /Appropriateness of data-collection and procedure /Data analysis /Trustworthiness/ reliability and validity	Excellent	<u>Good</u>	Moderate	Poor
6.	Steps taken to ensure that the research complies with standard ethical guidelines (if applicable)	Excellent	Good	<u>Moderate</u>	Poor
7.	Data presentation / Discussion (Are the results clearly and correctly presented? Are they consistent with the methodology?)	Excellent	<u>Good</u>	Moderate	Poor
8.	To what extent is the line of argumentation in the article clear, cohesive and logical?	Excellent	Good	<u>Moderate</u>	Poor
9.	Does the paper satisfy accepted criteria for academic writing in terms of coherence, grammar, layout and organisation?	Excellent	Good	<u>Moderate</u>	Poor
10.	Do the references adhere to APA?	Excellent	<u>Good</u>	Moderate	Poor
11.	Conclusions /Implications and/or recommendations are relevant and useful.	Excellent	Good	<u>Moderate</u>	Poor
12.	Is the language fluent and precise?	Excellent	Good	<u>Moderate</u>	Poor
13.	Is article significantly international in nature to be of value to global audience? /<u>underline</u>/ (Of Local Interest Only) (Of Regional Interest) <u>(Of International Interest)</u>	Excellent	Good	Moderate	Poor
14.	Overall assessment of content	Excellent	Good	<u>Moderate</u>	Poor
15.	Does the paper address relevant scientific questions within the scope of JBSE?	Excellent	Good	<u>Moderate</u>	Poor

16. Please write a brief narrative report on the article in which you provide a general or overall assessment of the manuscript and its suitability for publication.

The authors evaluated students' conceptual understanding and diagnosed the students' preconception related to particle characteristics of matter. This aim is realised through the development of diagnostic instrument and Rasch model response pattern analysis. The aim is good and method novel. However, the current report is more like a technical report on the Rasch analysis based on the data, collected with the developed instrument. More information about the questionnaire and its walidity are needed. Most serious weakness of the study is limited information about the validity of the instrument. The authors present several references to previous studies on particle model. However, there is limited information available, how the instrument was created based on the previpous literature. It is not clear what is the meaning and origin of "particle size", "particle mass", etc. A clear description of the items, what they measure and how the whole instrument cower the phenomena is needed. The instrument could be places as an attachment to the manuscript. Technical details related to Rush analysis could be reduced. Second there is not enough information available how the responses were coded and how the reliability of the coding was assured. How the item difficulty level was determined. What is the meaning of a sentence: "Research results, research results, research results, research results, research results, research results, research results."
17. Please indicate the strong aspects of the research that is reported.
The topic of the project is important and novel.
18. Please indicate the weak aspects of the research reported.
Validity and reliability issues should be clarified
19. Final recommendation:

Accept without revision	
Accept with minor revisions	
Accept: with moderate revisions	
Pre-accept: major revisions and re-evaluation	X
Reject: Rework and re-submit	
Reject: do NOT re-submit	

20. Comments: Please indicate in the space below any comments and suggestions for improving the article.

Date: April 21st 2020

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

200419.jbse_JL-suo_review

Abstract. *This study aimed to evaluate the students' conceptual understanding and to diagnose the students' preconception in elaborating the particle characteristics of matter by development of diagnostic instrument as well as Rasch model response pattern analysis approach. Data were acquired by 25 multiple-choice written test items distributed to 987 students in North Sulawesi, Indonesia. Analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding; and 3) diagnosis of students' preconception by estimation of item response pattern. The result generated information on diagnostic and summative measurement on students' conceptual understanding in elaborating the topic; information also acts as empirical evidence on the measurement's reliability and validity. Moreover, the result discovered significant disparity between students' conceptual understanding as based on their educational level. It was found that the distractor item response patterntended to be consistent, indicating a certain tendency of resistant preconception pattern. The findings are expected to be recommendation for future researchers and educational practitioners that integrates diagnostic and summative measurement with Rasch model item response pattern analysis approach in evaluating conceptual understanding and diagnosing misconception.*
Keywords: *item response pattern, conceptual understanding, Rasch model, particle characteristics of matter*.

Commented [Reviewer1]: Up to 200 words
Commented [Reviewer2]: Research
Commented [Reviewer3]: In alphabetical order
Commented [Reviewer4]: Avoid "of"

Introduction

Central to the notion of learning about characteristics of particle of matter is the process of developing understanding on abstract concepts (Johnstone, 1991) without directly interacting with the object/fact (Stojanovska, Soptrajanov, & Petrusevski, 2012); therefore it is considered difficult subject for the students to learn. Echoing this, disparity in understanding is almost inevitable (Özgür Kapici & Akcay, 2016) since different student may develop one's own distinctive way of understanding a concept (Yildirir & Demirkol, 2018). The idea is also coined by experts as misconception (Johnstone, 2006, 2010; Taber, 2002, 2009), or alternative framework and preconception (Lu & Bi, 2016). The experts have discovered that students always have their own preconception that is not in line with scientific concepts (Alamina & Etokeren, 2018; Yaşar, İnce, & Kırbaşlar, 2014), therefore, one needs to conduct identification and improvement on the conceptual learning (Allen, 2014; Soeharto, Csapó, Sarimanah, Dewi, & Sabri, 2019).

In diagnosing preconception, several researchers have developed diagnostic instruments in different mechanisms (McClary & Bretz, 2012), i.e., conceptual map, essay test, interview, essay test with interview, or multiple choice test (Femintasari, 2015). Two-step multiple choice diagnostic test (Treagust, 1988; Chandrasegaran et al., 2007; Tüysüz, 2009; Adadan & Savasci, 2012) is preferred due to its ability to diagnose preconception and describe the underlying reasons. The instrument is indeed considered qualitatively effective in elaborating differences in students' thought process; however, it does not provide summative measurement features due to lack of internal consistency and the instrument's unidimensionality (Lu & Bi, 2016). In addition to that, the measurement conclusion generated is considered weak due to extracted from analysis on raw score (Sumintono & Widhiarso, 2014).

Studies on preconception have found that the concept is somewhat resistant. In early 2000s, it is discovered that students' preconception persisted even when they already undergo formal

Commented [Reviewer5]: According to APA Manual 7th ed., the in-text citation for works with three or more authors is now shortened right from the first citation. You only include the first author's name and "et al.": https://www.scribbr.com/apa-style/apa-seventh-edition-changes/
Commented [Reviewer6]: According to APA Manual 7th ed., the in-text citation for works with three or more authors is now shortened right from the first citation. You only include the first author's name and "et al.": https://www.scribbr.com/apa-style/apa-seventh-edition-changes/
Commented [Reviewer7]: **The APA Manual (6/7th ed.) says: "Order the citations of two or more works within the same parentheses alphabetically" (6.16 on page 177).**

education experience (Hoe & Subramaniam, 2016). Preconception can also change along with the development of students' conceptual understanding; it also varies in different level of understanding (Aktan, 2013). If one conducts two-step test and raw score analysis approach to diagnose resistant preconception, the result generated will only provide limited feedback information (Sumintono, 2018) due to the instrument's limitation in measuring students' conceptual understanding. Instead of supporting, the information will only make it harder for teachers to implement proper instructional decisions (Wilson, 2008).

During the middle of 2000s, the Rasch model analysis was commonly used in studies of chemistry education (Herrmann-Abell & DeBoer, 2011; Liu, 2012; Wei, Liu, Wang, & Wang, 2012). The approach provides testing apparatus that integrates diagnostic and summative measurement. Recently, this approach is used to develop formative assessment with the intention to conduct learning construction mapping, e.g., measuring the students' way of constructing their understanding process (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013). It is worth to note, however, that there are studies that integrate diagnostic and summative measurement with different approach (Hoe & Subramaniam, 2016); despite that, trends in chemistry education studies highlight that diagnostic-summative measurement by Rasch model analysis is more common to be carried out(Laliyo, Botutihe, & Panigoro, 2019; Lu & Bi, 2016).

Commented [Reviewer8]: interval

Research Problem

The characteristics of particle of matter is a fundamental concept in chemistry, usually taught in middle education level. Adequate comprehension regarding the particle characteristics of matter both in macroscopic and microscopic level is essential as the knowledge basis in understanding more advanced topics such as the concept of atoms and molecules as the submicroscopic component that is invisible to plain eyesight, but exists in all real-world phenomena (Cheng, 2018; Ozmen, 2011; Yildirir & Demirkol, 2018). The fact signifies the relevance and reasoning of complexity in chemistry learning that is considered difficult for both students and teachers to conduct (Alamina & Etokeren, 2018). In simpler terms, to ensure that the chemistry learning is conducted effectively, one requires to nurture students' comprehensive understanding regarding particle characteristics of matter and its change of state.

To evaluate the students' conceptual understanding on the aforementioned topic, one also needs to measure the students capability in interpreting particle state during change process of a matter's form (Alamina & Etokeren, 2018; Barbera, 2013; Boz, 2006; Cheng, 2018; Gabel, 1993; Hadenfeldt et al., 2013; Kind, 2004; Naah & Sanger, 2012; Ozalp & Kahvecib, 2015; Özgür Kapici & Akcay, 2016; Ozmen, 2011; Renström, Andersson, & Marton, 1990; Slapničar, Devetak, Glažar, & Pavlin, 2017; Stojanovska et al., 2012; Yildirir & Demirkol, 2018). Researches on particle characteristics and changes of matter generally employ diagnostic instruments in the form of essay test and/or essay followed by interview; the instruments are further analyzed based on raw score results. The approach is considered inefficient and somewhat lacked accuracy in measuring students conceptual understanding and misconception pattern. Despite its ineffectiveness, the conventional method is used by most teachers in Indonesia to measure and determine students' learning progress. The teachers argue that measuring the students' raw score is effective in determining how far the students have progressed in learning process. The students' raw score is regarded by many as early premature indication regarding the measured variable, and is not eligible to be the final measurement indicator due to its temporary nature. In addition to that, regarding decision-making process, the raw score contains only limited information for it to be treated as reference (He, Liu, Zheng, & Jia, 2016; Sumintono & Widhiarso, 2015).

Commented [Reviewer9]: rework in a whole text.

Research Focus

The research focuses on developing diagnostic instrument that integrates measurement of conceptual understanding and diagnosis of students' preconception regarding the aforementioned topic by approach of Rasch model item response pattern analysis. The analysis employs different test apparatuses to provide extensive information for practitioners and researchers in science education in evaluating students' learning progress in different topics.

Research Aim and Research Questions

This study aims to investigate the following questions: 1) How is the effectiveness of measurement instrument to evaluate the students' conceptual understanding and diagnose their preconception on characteristics of particle of matter? 2) Is there any significant difference between students in elaborating the aforementioned topic based on their educational level? 3) How is the pattern of students' conceptual understanding and preconception regarding the topic?

Research Methodology

Respondent

The research employed non-experimental quantitative descriptive approach, in ways that the researcher did not manipulate or regulate the learning process and materials. The study was conducted during the even semester in 2019-2020 academic year. The respondents were 987 people, comprised of students of eleventh grade from eight junior high schools as well as university students of chemistry department in Northern Sulawesi, Indonesia. The distribution of respondents is displayed in following Table 1.

Table 1

Demographic profile of respondents (N=947)

Demography	Code	Respondents	Percentage (%)
Gender			
Male	M	320	67.68
Female	F	667	32.42
Education level			
X Class students	M	168	17.02
XI Class students	N	473	47.92
XII Class students	O	186	18.84
University students from chemistry department	P	160	16.21

The respondents were chosen randomly and have voluntarily agreed to participate within the research. In addition, they received no learning treatment and other special treatments that allow them to complete the measurement instrument.

Instrument and Procedures Development

The design process refers to recommendation by Wilson (2005) that consists of four key steps: definition of construct map, item design, result blank, and measurement model.

Commented [Reviewer10]: Please consistently use these terms. If you select 'research' or "study", you should exploit only one of them rather than interplayed usage. It is preferable to use the term "research"

Commented [Reviewer11]: Past simple

Commented [Reviewer12]: **Investigate** means: to try to find facts about an incident or a situation in order to know the truth. **Investigate** is often related to police investigations.

Commented [Reviewer13]: I do not see any research aim

Commented [Reviewer14]: **Research Methodology**
· Past tense to describe what was done. It may be past active or past passive.
· Present tense for presenting diagrams and figures

Commented [Reviewer15]: General description of research is important in order to show the basis of the research. It is like a very brief introduction to the methodology section as a whole. Research design should be explained. Indicate also research type, time (when was it carried out?) and scope. Research approach, a theoretical framework, etc. should be shortly described.
It is not correct to start from the participants..

Commented [Reviewer16]: What ethical procedures were followed?

Phase 1: Definition of construct map The map offers substantive definition of measured constructs; the more constructs measured, the constructs' level will vary qualitatively (Wilson, 2009). In simpler words, it aims to develop the students' understanding map to measure the students' progress (Wilson, 2012). The instrument involved variables i.e., the students conceptual understanding and preconception in elaborating the characteristics of particle of matter; it is conducted in accordance with the Curriculum Standard of Chemistry Subject in Tenth Grade in Indonesia, as presented in Table 2. Instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures.

Table 2

Conceptual Understanding Level

Level 3 The students are able to connect between characteristics of particle of matter in macroscopic and submicroscopic level	
Phenomenon	Evaporation: item Q6/Bubble
10. Preconception	Air bubble consists of Hydrogen and Oxygen particles
9. Preconception	Air bubble is water-soluble
Phenomenon	Condensation: item Q5/Dew
8. Preconception	Water drops come from melting ice that penetrates the glass wall
7. Preconception	Water drops are the result of reaction between ice and air nearby the glass
Level 2 The students are able to determine SMRs diagram of particle structure during change of form: item Q11/SMRs/SL; Q12/SMRs/LG; Q13/SMRs/GS; Q25/SMRs/GG	
6. Preconception	The SMRs diagram of particle structure follows the physical form of matter
5. Preconception	The SMRs diagram of O_2 molecule shape undergoes change as a result of increase in the volume of container.
Level 1 The students are able to determine characteristics of particle of matter during change process of matter's form.	
4. Preconception	The particle size of matter changes into (large/small) as a result of change in matter form: item Q1/PS/SL; Q7/PS/LG; Q14/PS/LG; Q18/PS/SG; Q22/PS/GG
3. Preconception	The particle mass of matter changes into (large/small) due to change in matter form: item Q2/PM/SL; Q8/PM/LG; Q15/PM/LG; Q19/PM/SG; Q24/PM/GG
2. Preconception	Distance between matter particles changes into (faster/slower) due to change in matter form: item Q3/DP/SL; Q9/DP/SL; Q16/DP/LG; Q20/DP/SG; Q23/PM/GG
1. Preconception	Motion between matter particles changes into (dense/loose) due to change in matter form: item item Q4/PMo/SL; Q10/PMo/LG; Q17/PMo/LG; Q21/PMo/SG

Variation in conceptual understanding level illustrates the development process of the students' conceptual understanding. In the first level, the students were asked to determine particle characteristics (size, mass, motion, and distance) in change process of matter form. In the second level, the students were asked to determine submicroscopic representation diagram of particle structure. Further, in the third level, the students were asked to connect between characteristics of particle of matter in macroscopic and submicroscopic level. In each level, the construct map also features the students' tendency of preconception.

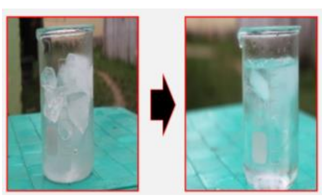
Phase 2: item design and evaluation The phase involves determination process of items to be used in acquiring evidences of students' construct understanding regarding the construct map (Wilson, 2005). Certain items may have different extent of effectiveness to measure students' conceptual understanding, (Mintzes, Wandersee, & Novak, 1999), however, multiple choices item is considered more practical and effective (Wilson, 2008). The instrument of concept understanding test of particle (or TPKP) is adapted from multiple choices instruments by Herrmann-Abell &

DeBoer (2011). Each item consists of two distractor answer choices and one open answer choice. The distractor answer choices are designed by referring to the common preconceptions by the students (see Table 2) as logical choices to distract the students from the correct one. The distractors function to emphasize the item diagnostic strength (Sadler, 1998). Some of the items are adopted from previous studies Osborne & Cosgrove (1983), Renström, Andersson, & Marton (1990); Devetak et al., (2004); Tóth & Kiss (2006); Davidowitz et al., (2010); Devetak & Glažar, (2010); Slapničar, Devetak, Glažar, & Pavlin (2017) and Yildirir & Demirkol (2018).

Figure 1

Sample of item Q1/PS/SL design

(a) (b)



Glass (a) contains ice chunks, glass (b) contains melting ice chunks. How is the size of water particle in solid form (ice) compared to that in liquid form?

a. Size of a water particle in solid form > a water particle in liquid form.
b. Size of a water particle in solid form < a water particle in liquid form.
c. Size of a water particle in solid form = a water particle in liquid form.
d. Other answers

The Picture 1 displays sample of item Q1/PS/SL design, in which Q1 is the number of item 1, PS is particle size, and SL is solid-liquid. The item measures student's capability in determining particle size in form change from solid to liquid. The choice A and B are distractors, the correct choice is C, and choice D is for other answers students may fill if the existing answer choices are not in accordance with their initial knowledge. Every correct answer is given mark 1, and wrong answers get 0 mark. Each student only have slight probability of 0.25 in choosing the right answer. The students will pick what they think the right answer based on their understanding. If the distractor item choice functions well, the students will not be able to predict the correct answer.

Phase 3: design of result blank, i.e., the correlation between construct map and items (Wilson, 2005). This phase aims to identify whether the answer the students pick correlates with their conceptual understanding; in simpler terms, it is intended to elaborate the conformity between the variable contents being measured. In order to elaborate the previous aspect, the TPKP instrument is validated by three independent experts and tested to the students to acquire their feedback. The process acquires 25 items of TPKP. Prior to the data collection process, it is ensured that all students have received formal education on the characteristics of particle of matter and its changes. The students' response towards the instrument is inputted manually by written answer sheet. The test were supervised by the teachers in school by referring to the agreed permission and duration. Each student is required to finish all test items within the allocated duration of 45 minutes. The instrument sheets are further collected, and checking process is conducted to ensure that the amount of instrument sheet is the same with participating students.

Phase 4: Rasch model analysis approach The analysis integrates algorithm as a result of probabilistic expectation of item 'i' and student 'n', as : The statement is the probability of student n in item i to result in correct answer (x = 1); with student ability, ßn, and item difficulty level (Bond & Fox, 2015). The above equation is simplified by inserting logarithm function, into , so that the probability of picking the right answer equals to student's ability subtracted by item difficulty

Commented [Reviewer17]: Figure?

level. The student (person) and item units are considered in the same interval scale and are independent to each other. The students' ability level and item difficulty level are measured in logarithm unit, namely odds or log that variates from -00 to +00 (Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015). The instrument efficiency, when compared to the item distribution towards item difficult level with distribution of student's ability level, is quantifiable in order to measure the students' conceptual understanding. In addition, the student's understanding level is differentiated based on the item size. The previous steps highlight the main difference of Rasch model analysis when compared to the raw score-based conventional one; the latter lacks accuracy in evaluating students' ability observed from different item difficulty level (Lu & Bi, 2016; Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015).

Data Analysis

The study employed WINSTEPS version 3.75 software to convert raw data into interval data (Linacre, 2012; Bond and Fox, 2015). The conversion result acts as the calibration of data on student's ability level and item difficulty level within the same interval measurement. Moreover, the analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding by Differential Item Functioning (DIF) item test; and 3) diagnosis of students' preconception by estimation of item response pattern through option probability curve test. Data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis.

Research Results

Effectiveness of Measuring Instruments

Person and Item Reliability. The first step to elaborate the effectiveness of measuring instruments is by measuring the person and item reliability. This is conducted to gather information to what extent the measurement produces consistent information in displaying latent trait or the unidimensionality of the measured variable (Sumintono & Widhiarso, 2015). The analysis result is presented in the form of statistical summary (Table 3). Research results, research results, research results, research results, research results, research results, research results.

Table 3

Summary of fit statistics

Parameter (N)	INFIT		OUTFIT		Separation	Reliability	Measure	KR-20	SD
	MNSQ	ZSTD	MNSQ	ZSTD					
Person (987)	1,00	-0,1	1,02	-0,1	1,55	0,71	-0,34	0,72	0,88
Items (25)	1,00	-0,8	1,02	-0,1	8,18	0,99	0,00		0,60

The above table indicates that the person reliability value of 0.71 is equivalent to the person separation index value of 1.55. This is to say that the consistency of students' response towards the test is deemed good. In addition to that, it is generated that the *Cronbach Alpha Coefficient* (KR-20) value is 0.72, signifying good interaction between students and the test. This further indicates strong correlation between the students' response towards the item, in the context that the students'

Commented [Reviewer18]: Research Results
· Past tense for results obtained
· Present tense to refer to figures, tables and graphs

Commented [Reviewer19]: italic

Commented [Reviewer20]: period

Commented [Reviewer21]: period

Commented [Reviewer22]: The APA Manual says: "Do not use a zero before a decimal fraction when the statistic cannot be greater than 1 (e.g., correlations, proportions, and levels of statistical significance, probability)" (4.35 on page 113).

knowledge tends to be non-fragmented, enabling it to be measured (Adams & Wieman, 2011). To the researchers and educational practitioners, such information is essential to prepare for follow-up plans and development of students' ability (Wei et al., 2012). Moreover, the result generates relatively high value of item separation index of 8.18 that is equivalent to the item reliability value of 0.99. This indicates very good item consistency, or the item is deemed capable to meet the unidimensionality criteria. In other words, the item performs very good in defining the measured variable. This is confirmed by the infit and outfit value result, in which most of the items are in the acceptable range for the multiple choice test (Herrmann-Abell & DeBoer, 2011; Trevor G. Bond and Christine M. Fox, 2015).

Figure 2

Function of Measurement Information

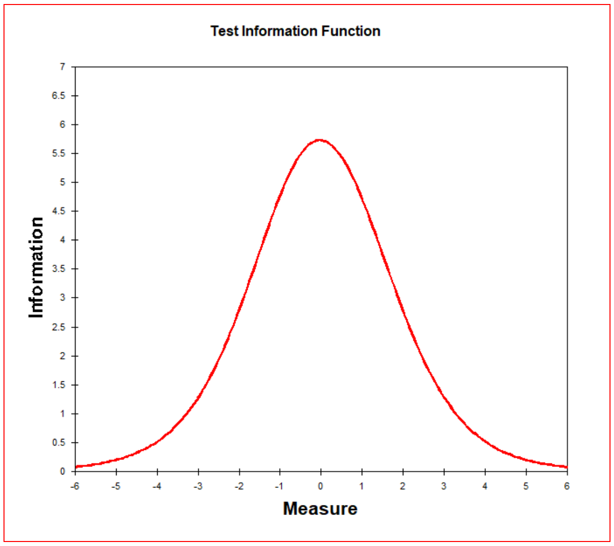


Figure 2 displays graph of measurement information in order to show the measurement reliability. The higher the tip of information function graph, the measurement reliability value is likely to increase. In the intermediate level of students' ability (-3.0 logit up to +3.0 logit), the measurement information is in very high spot. This indicates that the TPKP instrument is capable of producing optimal information to students with intermediate level of ability. This means that the instrument possess high measurement reliability (Misbach & Sumintono, 2014; Sumintono & Widhiarso, 2014; Trevor G. Bond and Christine M. Fox, 2015).

Validity. The next step is to measure the item validity by Fit item test to ensure that all items fit with the Rasch model. The process is aimed to identify whether or not the test item are able to measure the aspects that intended to be measured, or test validity (Linacre, 2012; Sumintono, 2018). The criteria used comprise outfit means-square (MNSQ): $0.5 < y < 1.5$; outfit z-standard: $-2.0 < Z < + 2.0$, as well as point measure correlation (PTMEA Corr). The PTMEA Corr is the correlation between score of item and person measure that is required to be positive value and not approaching zero (Trevor G. Bond & Christine M. Fox, 2015). The PTMEA Corr criteria: $0.4 < x < 0.8$. If all three criteria are not met, this signifies that the item is not good enough and need further elaboration

(Boone, Yale, & Staver, 2014). Both Outfiit MNSQ and Infit MNSQ are sensitive chi-squares in detecting outlier response pattern. There are two outlier response: the right response, guessed by the students with low ability in item with high difficulty level; or the wrong response due to the high-ability students' carelessness in items with low difficulty level. The expected ideal MNSQ value is 1.0. The analysis result on item appropriateness is displayed in Figure 3 as follows:

Figure 3

Item Statistics: Misfit Order

Commented [Reviewer23]: Poor quality

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
25	298	987	.60	.07	1.26	7.0	1.40	7.5	A .07	.37	65.6	73.6	Q6/Bubble
6	250	987	.88	.08	1.16	3.7	1.27	4.4	B .18	.36	74.5	77.3	Q2/PM/SL
8	287	987	.66	.08	1.12	3.3	1.20	3.8	C .22	.36	73.6	74.4	Q15/PM/LG
3	235	987	.97	.08	1.03	.8	1.18	2.8	D .30	.36	79.6	78.5	Q14/PS/LG
4	279	987	.71	.08	1.07	1.8	1.15	2.9	E .28	.36	76.2	75.0	Q18/PS/SG
24	549	987	-.63	.07	1.06	2.6	1.15	3.7	F .27	.34	65.8	64.7	Q5/Dew
2	287	987	.66	.08	1.06	1.6	1.14	2.8	G .29	.36	75.2	74.4	Q7/PS/LG
7	288	987	.65	.08	1.07	2.0	1.11	2.1	H .28	.36	73.9	74.3	Q8/PM/LG
1	264	987	.79	.08	1.00	-.1	1.06	1.1	I .35	.36	78.1	76.1	Q1/PS/SL
10	365	987	.25	.07	1.04	1.4	1.04	1.1	J .33	.37	67.1	69.1	Q24/PM/GG
9	268	987	.77	.08	.99	-.3	1.03	.6	K .36	.36	78.5	75.8	Q19/PM/SG
11	509	987	-.44	.07	1.01	.5	1.00	-.1	L .34	.35	63.4	64.4	Q3/DP/SL
17	430	987	-.07	.07	.98	-.8	.98	-.5	M .38	.36	68.8	66.0	Q10/PMo/LG
22	465	987	-.24	.07	.98	-1.1	.98	-.6	l .38	.36	67.6	64.9	Q13/SMRs/GS
12	482	987	-.32	.07	.97	-1.6	.95	-1.6	k .39	.36	66.0	64.6	Q9/DP/LG
16	557	987	-.66	.07	.96	-2.0	.93	-1.8	j .39	.34	67.5	65.0	Q4/PMo/SL
23	560	987	-.68	.07	.94	-2.9	.91	-2.4	i .41	.34	68.2	65.0	Q25/SMRs/GG
13	515	987	-.47	.07	.94	-3.1	.91	-2.8	h .42	.35	68.6	64.4	Q16/DP/LG
15	508	987	-.44	.07	.92	-3.7	.93	-2.1	g .43	.35	69.6	64.4	Q23/DP/GG
21	549	987	-.63	.07	.92	-3.8	.87	-3.5	f .44	.34	68.6	64.7	Q12/SMRs/LG
19	557	987	-.66	.07	.92	-4.0	.89	-2.9	e .43	.34	70.1	65.0	Q21/PMo/SG
18	566	987	-.71	.07	.91	-4.4	.87	-3.5	d .44	.34	71.4	65.2	Q17/PMo/LG
5	430	987	-.07	.07	.90	-4.4	.87	-3.9	c .47	.36	70.4	66.0	Q22/PS/GG
20	473	987	-.27	.07	.90	-4.9	.87	-4.0	b .47	.36	69.9	64.7	Q11/SMRs/SL
14	554	987	-.65	.07	.86	-6.6	.83	-4.6	a .49	.34	75.3	64.9	Q20/DP/SG
MEAN	421.0	987.0	.00	.07	1.00	-.8	1.02	-.1			70.9	68.9	
S.D.	120.7	.0	.60	.00	.09	3.2	.14	3.1			4.4	5.1	

From the previous Item Statistics, it is generated that all items meet the Outfit MNSA criteria and no negative PTMEA Corr occurs. This means that all items are not deviant, appropriate, and valid. Despite some items do not meet one of the criteria, this by no means decrease the quality of the items. For instance, item (Q6/Bubble, Q2/PM/SL, and Q15/PM/LG) do not meet the criteria of Outfit Z Standard and PTMEA Corr; item (Q1/PS/SL, Q24/PM/GG and Q19/PM/SG) do not meet the criteria of PTMEA Corr; and item (Q25/SMRs/GG, Q16/DP/LG, and Q23/DP/GG) do not meet the criteria of Outfit ZSTD; this is supposedly caused by large size of sample, or N > 500 (Boone et al., 2014).

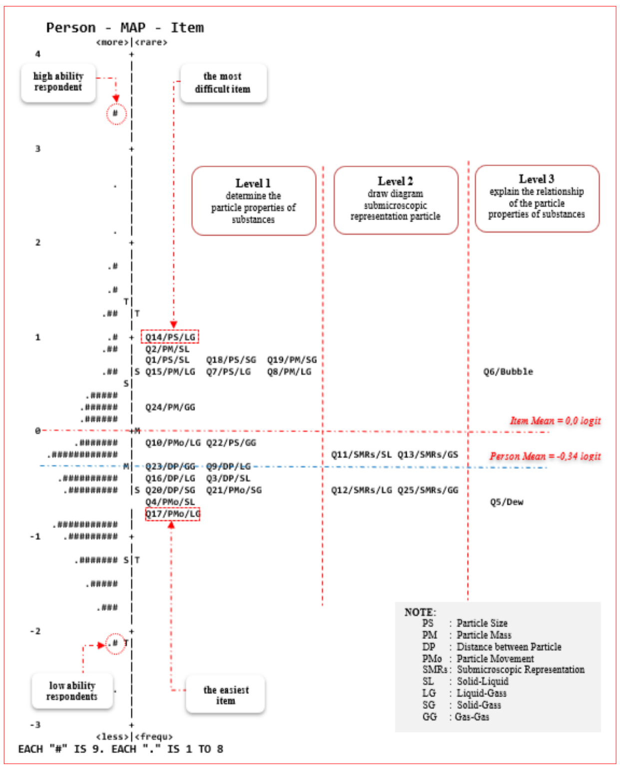
Wright Map: Person-Map-Item. The third step is to measure the consistency of item difficulty level and student's ability test constructed in Table 2. The higher the item difficulty level, the higher also the student's ability level will result. Information of Wright Map: Person-Map-Item is displayed in Figure 4. The previous Wright map generates that all instrument items encompass almost all the students' ability. The map generates variance from students with very high ability (> 3.0 logit), to those with very low ability (< -2.0 logit) as well. In addition to that, disparity (in which there is no item that is appropriate with the student's ability) is observed within the interval of -3.0

logit up to -0.5 logit and in the interval of +1.0 logit up to +3.7 logit. This signifies that the information generated within the interval range is somewhat limited and needs further elaboration. On the other hand, the item difficulty level is mostly located in the interval of -1.0 logit up to +1.0 logit; what is more, the items tend to occur in the same difficulty level. The item Q14/PS/LG is the most difficult item with logit of +0.97, while item Q17/Pmo/LG is the easiest item with logit of -0.71.

Figure 4

Wright Map: Person-Map-Item

Commented [Reviewer24]: Poor quality

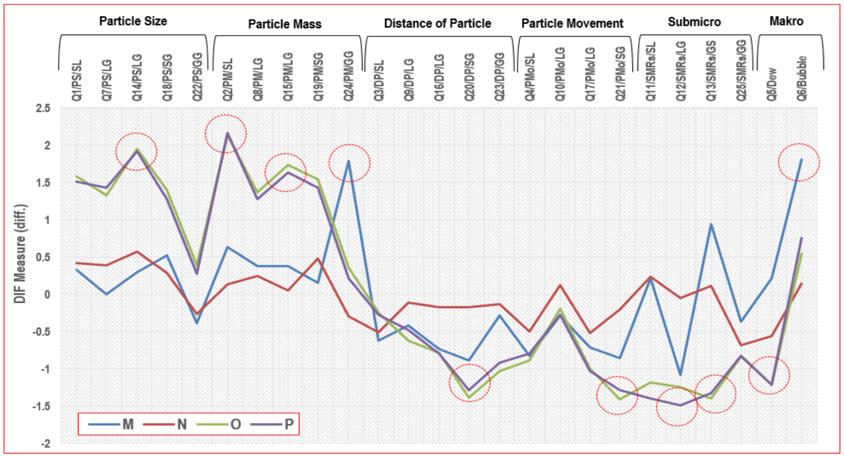


As observed from the differences in item size, some interesting cases are explained as follows: **Firstly**, the items in level 1: Q14/PS/LG (0.97) > Q1/PS/SL (0.79) > Q18/PS/PG (0.71) > Q7/PS/GG (0.66) are instead assumed by the students to possess different difficulty level. The items above, however, are more difficult than item Q6/Bubble in level 3 (0.60). In other words, determining a particle size is more difficult than explaining the particle characteristics of matter in evaporation phenomenon. **Secondly**, the size of item Q5/Dew (-0.63) < item Q6/Bubble; this indicates that it is harder for the students to elaborate the particle characteristics of matter in evaporation phenomenon than in condensation phenomenon, despite that both items are in the same level. **Thirdly**, the size of following items: Q2/PM/SL (0.88) > Q19/PM/SG (0.77) > Q15/PM/LG (0.66) > Q8/PM/LG (0.65) > Q24/PM/GG in level 1 is larger compared to that of items Q13/SMRs/GS (-0.24) > Q11/SMRs/SL (-0.27) > Q12/SMRs/LG (-0.63) > Q25/SMRs/GG (-0.68) in level 2. The finding illustrates that it is harder for the students in determining the particle mass than determining submicrorepresentation (SMRs) diagram in different form changes of matter. The previous cases identifies disparity in students' conceptual understanding, signifying that the level of understanding in particle characteristics of matter is relatively low. Overall, 80% of test item difficulty level is relatively parallel with the measured constructs. By that, the test possess good construct validity (Blanc & Rojas, 2018; Lu & Bi, 2016; Neumann, Neumann, & Nehm, 2011).

Disparity in Conceptual Understanding Level

The next step is the measurement of disparity of students' conceptual understanding in the focused topic based on educational level by Differential Item Functioning (DIF).

Figure 5
Person DIF plot based on educational level



Description: M = X Class students,
N = XI Class students,
O = XII Class students, and
P = University students from chemistry department

Figure 5 of DIF plot based on students' educational level depicts that ten items are identified to possess significant disparity. **Firstly**, five curves approaching the upper limit are items with high difficulty level (Q14/PS/LG, Q2/PM/SL, Q15/PM/SG, Q24/PM/GG and Q6/Bubble); while five curves approaching the lower limit are items with low difficulty level (Q20/DP/SG, Q21/PMo/SG, Q12/SMRs/LG, Q13/SMRs/GS and Q5/Dew). **Secondly**, the item Q14/PS/LG (particle size in form change of liquid-gas), Q2/PM/SL (particle mass in form change of solid-liquid), and Q15/PM/SG

Commented [Reviewer25]: In the hard copy of the journal, colours will become gray, and the graphs will be indistinguishable. Mark them explicitly and reliably for printing in black-and-white, add the explanation.

(particle mass in change form of solid-gas) are deemed very hard by the students of XII class and the university students compared to students in X and XI class. **Thirdly**, the study discovers different result for item Q24/PM/GG and Q6/Bubble. The item Q24/PM/GG (particle mass of O2 in larger volume) and Q6/Bubble (constructing elements of air bubbles during boiling process of water) are deemed very hard for X class students compared to students in XI and XII classes, as well as university students. **Fourthly**, the items Q20/DP/SG (distance between particles in form change of solid-gas), Q21/PMo/SG (motion between particles in form change of solid-gas), Q12/SMRs/LG (SMRs diagram of particle in form change of liquid-gas), Q13/SMRs/GS (SMRs diagram of particle in change form of gas-liquid), and Q5/Dew (condensation) are deemed too easy for students in XII class and university students compared to the students in X and XI classes.

Pattern of Conceptual Understanding and Preconception

The analysis on pattern of conceptual understanding and preconception employs option probability curve test (Linacre, 2012). Option probability curve aims to display probability of picking every answer choice to elaborate the performance level of all students in the measured items (Abell & DeBoer, 2011). The test relies on the principle that the curve of correct answer will rise along with the decrease of the curve of distractor choices (Haladyna, 2004; Trevor G. Bond and Christine M. Fox, 2015). For items that are influenced by distractor options, the curve produced tends to be non-parallel with the traditional monotonous item behavior (Sadler, 1998), by this reason, each answer choice is analyzed separately.

The instrument provides four answer choices, therefore, four curves are formed. Each curve displays the students' comprehension. Students with low ability tend to pick distractor choice, while students whose high ability are more likely to prefer other preconceptions (Abell and DeBoer, 2011; Perera, Sumintono, & Jiang, 2018). Below is the elaboration of pattern of students' conceptual understanding and preconception based on four option probability curves.

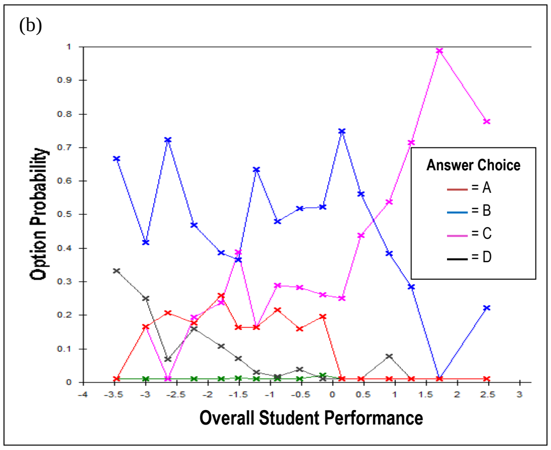
Commented [Reviewer26]: An ampersand

Figure 6

(a) sample of item Q2/PM/SL, (b) option probability curve

(a)
When some ice cubes in the glass melts, some other ice cubes are seen floating on the water surface. How is the comparison between mass of one particle of ice and one particle of water?

a. Mass of one particle of ice is bigger than mass of one particle of water.
b. mass of one particle of ice is smaller than mass of one particle of water.
c. mass of one particle of ice is similar with mass of one particle of water.
d Other answer...



First example, the item Q2/PM/SL (0.88) is shown un Figure 6(a). The item measures students' capability in determining particle size in form change from solid to liquid. The option probability curve is displayed in Figure 6(b). Students with low ability (< 0.5 logit) tend to pick distractor choice B (mass of one particle of ice is smaller than mass of one particle of water) or A (mass of one particle of ice is bigger than mass of one particle of water). In addition, students with very low ability (< -1.0 logit) tend to pick D (other answer). Some students with relatively low ability (> -2.5 logit), however, pick the right answer C (mass of one particle of ice is similar with mass of one particle of water). One can predict the response pattern of students with low ability, as the distractors A, B, and D contain third preconception in level 1 (see Table 2). The students possess the knowledge that mass of particle of matter can change into larger or smaller size by observing the matter's change of form. It is interesting to note that there are students with high ability (>2.0) that pick B; this indicates the presence of resistant preconception.

Commented [Reviewer27]: In the hard copy of the journal, colours will become gray, and the graphs will be indistinguishable. Mark them explicitly and reliably for printing in black-and-white, add the explanation.

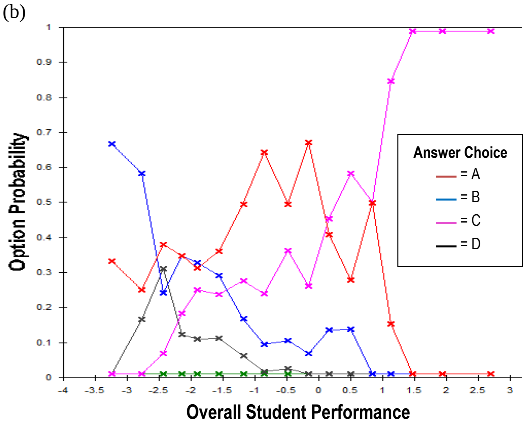
Figure 7
(a) item Q8/PM/LG; (b) option probability curve

(a)

During boiling process of water, a change of form of water from liquid to gas occurs. How is the comparison between mass of one particle of water and one particle of water vapor?

a. Mass of one water particle is bigger than mass of one vapor particle
b. mass of one water particle is smaller than mass of one vapor particle.
c. mass of one water particle is similar with mass of one vapor particle.
d Other answer...

(b)



Second sample, item Q8/PM/LG (0.65) is shown in figure 7(a) as the item to measure students' ability in determining mass of particle in form change of liquid-gas. The option probability curve is displayed in Figure 7(b). The curve of distractor B (mass of one water particle is smaller than mass of one vapor particle) is chosen by students with low ability (< -2.0 logit), while curve of choice A (mass of one water particle is bigger than mass of one vapor particle) is chosen by students with ability in range of -3.5 to 1.5 logit. The correct answer, option C (mass of one water particle is similar with mass of one vapor particle), is chosen by students with ability in > -2.5 logit. As highlighted in the table, the decline of curve of distractor A is followed by the increase of curve of right answer C; both curves intersect in level of 1.0 logit. The shape of curve A indicates the presence of resistant preconception type-three in level 1.

It depicts that the particular item response pattern that signifies students' conceptual understanding pattern in the given level. Moreover, the curve shape of distractors A and B in the items Q2/PM/SL and Q8/PM/SL tend to have identical pattern. The finding indicates that students with either low or high ability have consistent preconception that mass of particle can change into larger or smaller in size along with the change in matter form.

Third sample, item Q5/Dew (-0.63) as shown in Figure 8(a) measures the students' ability in elaborating characteristics of particle in condensation phenomenon. The option probability curve is displayed in Figure 8(b). Students with low ability (< 1.0 logit) tend to pick distractor A (water drops come from liquid of melting ice that breaks through the glass wall) and option D (other answer). Some students with high ability (> 1.0 logit) also pick distractor B (water drops are the result of reaction between ice and air nearby the glass). The shape of curve B is wavy and non-linear, even in the interval of 2.0 to 4.0 logit, it can reach option probability value up to 1.0 logit. This is regarded as deviation from the right answer C (water drops come from condensing water vapor nearby the glass). A worth note, however, is to consider in the unstable, wavy shape of curve C; this indicates the students' inconsistency (particularly those with high ability) in comprehending concept of condensation. This shows that students have their own preconception regarding concept of condensation.

Figure 8

(a) item Q5/Dew; (b) option probability curve

(a)

In a glass filled with ice, you can see water drops at the glass. According to you, where do the water drops come from?

a. Water drops come from melting ice that penetrates the glass wall
b. water drops are the result of reaction between ice and air nearby the glass
c. Water drops come from condensing water vapor nearby the glass
d Other answer...

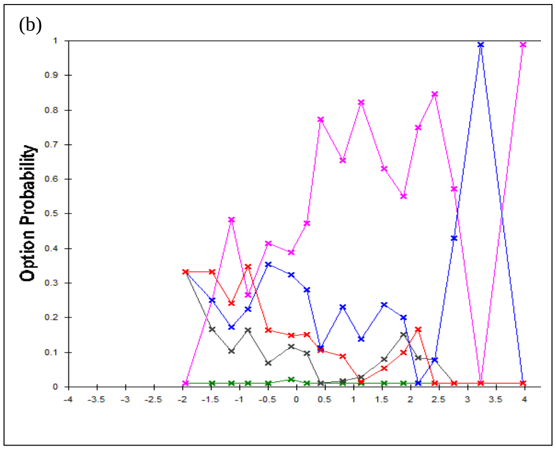


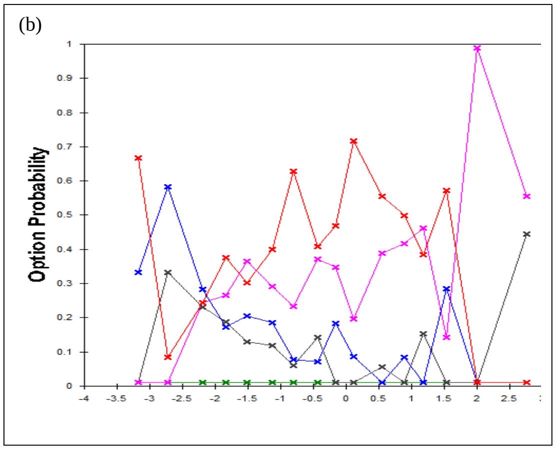
Figure 9

(a) item Q6/Bubble; (b) option probability curve

(a)

In a container filled with boiling water, you can see air bubbles on the top of it. According to you, what are the composing elements of the air bubbles?

a. Hydrogen and Oxygen particles
b. Air that is dissolved in water
c. Water molecules
d. Other answer...



Fourth sample, item Q6/Bubble as shown in Figure 9(a) measures the students' ability in elaborating characteristics of particle in evaporation phenomenon. The option probability curve is displayed in Figure 9(b). The distractor A (air bubbles are Hydrogen and Oxygen particles) is dominantly chosen by students whose ability in range between -3.0 to 2.0 logit. Moreover, the distractor B (air bubbles are Hydrogen and Oxygen particles) is dominantly chosen by students whose ability in range between -3.0 to 0.5 logit. The form of curve A and B picked by students with low ability is predictable. The curve of right answer C (air bubbles are water molecules), however, shows interesting hint; in the interval range of -2.5 to 3.0 logit, the tip of curve shows up-and-down pattern. Moreover, in the level of 1.5 logit, the curve shape of distractors A and B shows decline pattern, while that of curve C tends to increase. Another finding worth noting is that the curve D (other answers) is picked by some students with high ability (> 2.0 logit). This indicates that the particular students have their own preconception regarding evaporation concept.

Discussion

The research findings indicate that the instruments have good effectiveness, meet the requisites of person and item reliability, and show good construct validity. When applied in evaluating students' conceptual understanding, it is found that: Firstly, almost all students with high ability face difficulty in understanding concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, or determining concept of particle regarding evaporation and condensation phenomena in level 3. Secondly, the information of response pattern of students with high ability is quite consistent, repetitive, and systematic in particular items. This indicates the presence of permanent and latent preconception. The analysis of option probability curve of item Q2/PM/SL (0.88), Q8/PM/LG (0.65), Q5/Dew (-0.63) and Q6/Bubble (0.60) indicates that the approach of item response pattern is able to explore in detail and comprehensively regarding students' conceptual understanding and preconception.

Sequences of verification conducted that involves Rasch model approach shows detailed, accurate, and quantifiable results since the approach integrates development procedure of diagnostic and summative instruments. Several samples of preconception, e.g., item Q2/PM/SL (0.88) and Q8/PM/LG (0.65) indicate that distractor options are potential to be elaborated further in order to investigate tendency of preconception by the students. In addition, it also provides information regarding main idea unknown to the students and their degree of misunderstanding.

The approach employed in this study is an effective illustration to help teacher in evaluating the learning process as well as the students' learning progress. This is due to the integration of qualitative item development procedure and quantitative data analysis, allowing the teachers to explore in-depth on the students' understanding, concepts the students understand and/or do not understand, and misconception.

Commented [Reviewer28]: Discussion
· Present tense to explain significance of results
· Past tense to summarize findings, with present tense to interpret results

Commented [Reviewer29]: There is an imitation of discussion

APA 6/7th. Ed. Manual reads: "Similarities and differences between your results and the work of others should be used to contextualize, confirm, and clarify your conclusions." It means you are expected to cite other authors' papers in this section.

Conclusions and Implications

The measuring instrument developed performed well in its validity and reliability, thus, is deemed applicable in measuring students' conceptual understanding and preconception in elaborating particle characteristics of matter. During implementation of the instruments, the study finds out that: 1) almost all students with high ability face difficulty in understanding concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, as well as determining concept of particle regarding evaporation and condensation phenomena in level 3.

2) There is significant disparity between students' conceptual understanding as based on their educational level. 3) On certain cases, it is found that the distractor item response pattern by high-ability students tend to be consistent, indicating a certain tendency of resistant preconception pattern.

The development of diagnostic instrument with Rasch model approach is deemed as literacy process for practitioners and researchers in Indonesia. The result indicates that there is no single item that is parallel with both highest ability and lowest ability students. This calls for further elaboration in order to improve the instrument items quality. Moreover, an anomaly is found that students with high ability (> 1.0 logit) tend to pick distractor choices. This urges further studies to investigate structured comprehension problems.

Commented [Reviewer30]: It must be clear what is the added value of this research to previous reports. It is obligatory to make sure that reported new information is significant for international level readers of JBSE.

Acknowledgements

This research is funded by the Directorate of Research and Community Outreach/ (Direktorat Riset dan Pengabdian Masyarakat or DRPM), Directorate General of Research and Development (Dirjen Riset dan Pengembangan) of Ministry of Research and Technology of the Republic of Indonesia through Penugasan Penelitian Dasar Unggulan Perguruan Tinggi Grant Program, Research and Community Outreach Institute (Lembaga Penelitian dan Pengabdian Masyarakat) of Universitas Negeri Gorontalo in 2020.

References

Adadan, E., & Savasci, F. (2012). An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, *34*(4), 513–544. https://doi.org/10.1080/09500693.2011.636084

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

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

Alamina, J. I., & Etokeren, I. S. (2018). Effectiveness of Imagination Stretch Teaching Strategy in Correcting Misconceptions of Students about Particulate Nature of Matter. *Journal of Education, Society and Behavioural Science*, *27*(1), 1–11. https://doi.org/10.9734/jesbs/2018/43063

Allen, M. (2014). Misconceptions in Primary Science. In *McGraw-Hill Education (UK).* *revise*

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

Blanc, A., & Rojas, A. J. (2018). Use of rasch person-item maps to validate a theoretical model for measuring attitudes toward sexual behaviors. *PLoS ONE*, *13*(8), 1–17. https://doi.org/10.1371/journal.pone.0202551

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch Analysis in the Human Sciences. In *Rasch Analysis in the Human Sciences*. https://doi.org/10.1007/978-94-007-6857-4

The reference is incomplete. What is the journal issue and/or pages? Or volume? Publisher /if needed/Editor(s) etc. Check format for chapters from books, if needed.

Boz, Y. (2006). Turkish pupils' conceptions of the particulate nature of matter. *Journal of Science Education and Technology*, *15*(2), 203–213. https://doi.org/10.1007/s10956-006-9003-9

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The Development of Two Tier Multiple-

Choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Levels of Representation. *Chemistry Education Research and Practice*, *8*(3), 293–307.

Cheng, M. M. W. (2018). Students' visualisation of chemical reactions-insights into the particle model and the atomic model. *Chemistry Education Research and Practice*, *19*(1), 227–239. https://doi.org/10.1039/c6rp00235h

Davidowitz, B., Chittleborough, G., & Murray, E. (2010). Student-generated submicro diagrams: A useful tool for teaching and learning chemical equations and stoichiometry. *Chemistry Education Research and Practice*, *11*(3), 154–164. https://doi.org/10.1039/c005464j

Devetak, I., & Glažar, S. A. (2010). The influence of 16-year-old Students' Gender, Mental Abilities, and motivation on their Reading and Drawing Submicrorepresentations Achievements. *International Journal of Science Education*, *32*(12), 1561–1593. https://doi.org/10.1080/09500690903150609

Devetak, I., Urbancic, M., Grm, K. S. W., Krnel, D., & Glazar, S. A. (2004). Submicroscopic Representations As a Tool for Evaluating Students' Chemical Misconception. *Acta Chim. Slov, 51*(December 2004), 799–814.

Journal`s title should be written in full

Femintasari, V. (2015). the Effectiveness of Two-Tier Multiple Choice Test and Multiple Choice Test Followed With Interview in Identifying Misconception of Students With Different Scientific Reasoning Skills. *Jurnal Ilmu Pendidikan*, *21*(2), 192–197.

Gabel, D. L. (1993). Use of the particle nature of matter in developing conceptual understanding. *Journal of Chemical Education*, *70*(3), 193. https://doi.org/10.1021/ed070p193

Hadenfeldt, J. C., Bernholt, S., Liu, X., Neumann, K., & Parchmann, I. (2013). Using Ordered Multiple-Choice Items To Assess Students' Understanding of the Structure and Composition of Matter. *Journal of Chemical Education*, *90*(12), 1602–1608. https://doi.org/10.1021/ed3006192

Haladyna, T. M. (2004). *Developing and Validating Multiple-choice Test Items*. https://doi.org/10.4324/9780203825945 incomplete

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

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

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

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

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

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

Kind, V. (2004). Beyond Appearances: Students' misconceptions about basic chemical ideas 2 nd Edition. *School of Education*, 1–84. revise

Laliyo, Botutihe, & Panigoro. (2019). The development of two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12

Linacre, J. M. (2012). *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 3.75.0*. https://doi.org/ISBN 0-941938-03-4

Liu, X. (2012). Developing Measurement Instruments for Science Education Research. In C. J. Fraser, Barry, Tobin, Kenneth, McRobbie (Ed.), *Second International Handbook of Science Education* (pp. 651–665). Springer Netherlands,.

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

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, *34*(15), 2317–2341. https://doi.org/10.1080/09500693.2012.684433

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (1999). Assessing Science Understanding: A Human Constructivist View. In Joel J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Academic Press, San Diego， California ： Elsevier Academic Press.* Retrieved from https://zodml.org/sites/default/files/%5BJoel_J._Mintzes%2C_James_H._Wandersee%2C_Joseph_D._No_0.pdf

Misbach, I. H., & Sumintono, B. (2014). Pengembangan dan validasi instrumen "Persepsi Siswa tehadap Karakter Moral Guru" di Indonesia dengan Model Rasch [Instrument Development and Validation of "Students' Perception of Teacher's Moral Characters in Indonesia" by Rasch Model]. *Seminar Nasional Psikometri*, (May).

Naah, B. M., & Sanger, M. J. (2012). Student misconceptions in writing balanced equations for dissolving ionic compounds in water. *Chem. Educ. Res. Pract.*, *13*(3), 186–194. https://doi.org/10.1039/C2RP00015F

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*(9), 825–838. https://doi.org/10.1002/tea.3660200905

Ozalp, D., & Kahvecib, A. (2015). Diagnostic assessment of student misconceptions about the particulate nature of matter from ontological perspective. *Chemical Education Research and Practice*, *16*, 619–639. https://doi.org/10.1039/C5RP00096C

Özgür Kapici, H., & Akcay, H. (2016). Particulate Nature of Matter Misconceptions Held By Middle and High School Students in Turkey. *European Journal of Education Studies*, *2*(8), 43–58. https://doi.org/10.5281/zenodo.163547

Ozmen, H. (2011). Turkish Primary Students´ Conceptions about the Particulate Nature of Matter. *International Journal of Environment & Science Education*, *6*(1), 99–121.

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

Renström, L., Andersson, B., & Marton, F. (1990). Students' Conceptions of Matter. *Journal of Educational Psychology*, *82*(3), 555–569. https://doi.org/10.1037/0022-0663.82.3.555

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

Slapničar, M., Devetak, I., Glažar, A. S., & Pavlin, J. (2017). Identification of the understanding of the states of matter of water and air among Slovenian students aged 12, 14 and 16 years through solving authentic tasks. *Journal of Baltic Science Education*, *16*(3), 308–323. Retrieved from https://www.researchgate.net/publication/318215148

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

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

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

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi MOdel Rasch untuk Penelitian Ilmu-Ilmu Sosial [Application of Rasch Model in Social Sciences Research]* (B. Trim, ed.). Retrieved from https://www.researchgate.net/publication/268688933%0AAplikasi incomplete

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan [Application of Rasch Model in Education Assessment]*. Retrieved from https://www.researchgate.net/publication/282673464%0AAplikasi

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

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

Tóth, Z., & Kiss, E. (2006). Using particulate drawings to study 13-17 year olds' understanding of physical and chemical composition of matter as well as the state of matter. *Practice and Theory in Systems of Education*, *1*(1), 109–125. Retrieved from http://eduscience.fw.hu/

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169. https://doi.org/10.1080/0950069880100204

Trevor G. Bond and Christine M. Fox. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences. In *Routledge Taylor & Francis Group* (Third Edit, Vol. 44). https://doi.org/10.1088/1751-8113/44/8/085201

Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, *4*(6), 626–631. Retrieved from http://www.academicjournals.org/SRE

URLs are no longer preceded by "Retrieved from," unless a retrieval date is needed. The website name is included (unless it's the same as the author), and web page titles are italicized (APA 7th.ed)

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335–345. https://doi.org/10.1021/ed100852t

Wilson, M. (2005). Constructing Measures: An Item Response Modeling Approach. In *Lawrence Erlbaum Associates: Mahwah, NJ, 2005*. https://doi.org/10.4324/9781410611697

Wilson, M. (2008). Cognitive Diagnosis Using Item Response Models. *Zeitschrift Für Psychologie / Journal of Psychology*, *216*(2), 74–88. https://doi.org/10.1027/0044-3409.216.2.74

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

Wilson, M. (2012). Responding to a Challenge that Learning Progressions Pose to Measurement Practice. In Alicia C. Alonzo and & Amelia Wenk Gotwals (Eds) (Eds.), *Learning Progression in Science* (pp. 317–344). https://doi.org/10.1007/978-94-6091-824-7

Yaşar, I. Z., İnce, E., & Kırbaşlar, F. G. (2014). 7. Class Science and Technology Course "Structure of Atom" Subject Readiness Improvement Test. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2014.09.259

incomplete

Yildirir, H. E., & Demirkol, H. (2018). Identifying Mental Models of Students for Physical and Chemical Change. *Journal of Baltic Science Education*, *17*(6), 986–1004. https://doi.org/10.33225/jbse/18.17.986

Quick Reference Guide: https://apastyle.apa.org/instructional-aids/reference-guide.pdf

The list of references should be carefully checked for an appropriate use of dots, commas, intervals/periods, italic, capital letters etc. A special attention must be paid to punctuation.

Brushing the References section: the author(s) should: (a) make sure that sources cited in the narrative are included in the References; (b) employ the APA style (italicization, capitalization; translation, etc);

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

By Lukman Abdul Rauf Laliyo

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

200419,jbse_JL-suo_review

Abstract. *This study aimed to evaluate the students' conceptual understanding and to diagnose the students' preconception in elaborating the particle characteristics of matter by development of diagnostic instrument as well as Rasch model response pattern analysis approach. Data were acquired by 25 multiple-choice written test items distributed to 987 students in North Sulawesi, Indonesia. Analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding; and 3) diagnosis of students' preconception by estimation of item response pattern. The result generated information on diagnostic and summative measurement on students' conceptual understanding in elaborating the topic; information also acts as empirical evidence on the measurement's reliability and validity. Moreover, the result discovered significant disparity between students' conceptual understanding as based on their educational level. It was found that the distractor item response patterntended to be consistent, indicating a certain tendency of resistant preconception pattern. The findings are expected to be recommendation for future researchers and educational practitioners that integrates diagnostic and summative measurement with Rasch model item response pattern analysis approach in evaluating conceptual understanding and diagnosing misconception.*
Keywords: *item response pattern, conceptual understanding, Rasch model, particle characteristics of matter.*

Introduction

Central to the notion of learning about characteristics of particle of matter is the process of developing understanding on abstract concepts (Johnstone, 1991) without directly interacting with the object/fact (Stojanovska, Soptrajanov, & Petrusevski, 2012); therefore it is considered difficult subject for the students to learn. Echoing this, disparity in understanding is almost inevitable (Özgür Kapici & Akcay, 2016) since different student may develop one's own distinctive way of understanding a concept (Yildirir & Demirkol, 2018). The idea is also coined by experts as misconception (Johnstone, 2006, 2010; Taber, 2002, 2009), or alternative framework and preconception (Lu & Bi, 2016). The experts have discovered that students always have their own preconception that is not in line with scientific concepts (Alamina & Etokeren, 2018; Yaşar, İnce, & Kırbaşlar, 2014), therefore, one needs to conduct identification and improvement on the conceptual learning (Allen, 2014; Soeharto, Csapó, Sarimanah, Dewi, & Sabri, 2019).

In diagnosing preconception, several researchers have developed diagnostic instruments in different mechanisms (McClary & Bretz, 2012), i.e., conceptual map, essay test, interview, essay test with interview, or multiple choice test (Femintasari, 2015). Two-step multiple choice diagnostic test (Treagust, 1988; Chandrasegaran et al., 2007; Tüysüz, 2009; Adadan & Savasci, 2012) is preferred due to its ability to diagnose preconception and describe the underlying reasons. The instrument is indeed considered qualitatively effective in elaborating differences in students' thought process; however, it does not provide summative measurement features due to lack of internal consistency and the instrument's unidimensionality (Lu & Bi, 2016). In addition to that, the measurement conclusion generated is considered weak due to extracted from analysis on raw score (Sumintono & Widhiarso, 2014).

Studies on preconception have found that the concept is somewhat resistant. In early 2000s, it is discovered that students' preconception persisted even when they already undergo formal education experience (Hoe & Subramaniam, 2016). Preconception can also change along with the development of students' conceptual understanding; it also varies in different level of understanding (Aktan, 2013). If one conducts two-step test and raw score analysis approach to diagnose resistant preconception, the result generated will only provide limited feedback information (Sumintono, 2018) due to the instrument's limitation in measuring students' conceptual understanding. Instead of supporting, the information will only make it harder for teachers to implement proper instructional decisions (Wilson, 2008).

During the middle of 2000s, the Rasch model analysis was commonly used in studies of chemistry education (Herrmann-Abell & DeBoer, 2011; Liu, 2012; Wei, Liu, Wang, & Wang, 2012). The approach provides testing apparatus that integrates diagnostic and summative measurement. Recently, this approach is used to develop formative assessment with the intention to conduct learning construction mapping, e.g., measuring the students' way of constructing their understanding process (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013). It is worth to note, however, that there are studies that integrate diagnostic and summative measurement with different approach (Hoe & Subramaniam, 2016); despite that, trends in chemistry education studies highlight that diagnostic-summative measurement by Rasch model analysis is more common to be carried out(Laliyo, Botutihe, & Panigoro, 2019; Lu & Bi, 2016).

Research Problem

The characteristics of particle of matter is a fundamental concept in chemistry, usually taught in middle education level. Adequate comprehension regarding the particle characteristics of matter both in macroscopic and microscopic level is essential as the knowledge basis in understanding more advanced topics such as the concept of atoms and molecules as the submicroscopic component that is invisible to plain eyesight, but exists in all real-world phenomena (Cheng, 2018; Ozmen, 2011; Yildirir & Demirkol, 2018). The fact signifies the relevance and reasoning of complexity in chemistry learning that is considered difficult for both students and teachers to conduct (Alamina & Etokeren, 2018). In simpler terms, to ensure that the chemistry learning is conducted effectively, one requires to nurture students' comprehensive understanding regarding particle characteristics of matter and its change of state.

To evaluate the students' conceptual understanding on the aforementioned topic, one also needs to measure the students capability in interpreting particle state during change process of a matter's form (Alamina & Etokeren, 2018; Barbera, 2013; Boz, 2006; Cheng, 2018; Gabel, 1993; Hadenfeldt et al., 2013; Kind, 2004; Naah & Sanger, 2012; Ozalp & Kahvecib, 2015; Özgür Kapici & Akcay, 2016; Ozmen, 2011; Renström, Andersson, & Marton, 1990; Slapničar, Devetak, Glažar, & Pavlin, 2017; Stojanovska et al., 2012; Yildirir & Demirkol, 2018). Researches on particle characteristics and changes of matter generally employ diagnostic instruments in the form of essay test and/or essay followed by interview; the instruments are further analyzed based on raw score results. The approach is considered inefficient and somewhat lacked accuracy in measuring students conceptual understanding and misconception pattern. Despite its ineffectiveness, the conventional method is used by most teachers in Indonesia to measure and determine students' learning progress. The teachers argue that measuring the students' raw score is effective in determining how far the students have progressed in learning process. The students' raw score is regarded by many as early premature indication regarding the measured variable, and is not eligible to be the final measurement indicator due to its temporary nature. In addition to that, regarding decision-making process, the raw score contains only limited information for it to be treated as reference (He, Liu, Zheng, & Jia, 2016; Sumintono & Widhiarso, 2015).

Research Focus

The research focuses on developing diagnostic instrument that integrates measurement of conceptual understanding and diagnosis of students' preconception regarding the aforementioned topic by approach of Rasch model item response pattern analysis. The analysis employs different test apparatuses to provide extensive information for practitioners and researchers in science education in evaluating students' learning progress in different topics.

Research Aim and Research Questions

This study aims to investigate the following questions: 1) How is the effectiveness of measurement instrument to evaluate the students' conceptual understanding and diagnose their preconception on characteristics of particle of matter? 2) Is there any significant difference between students in elaborating the aforementioned topic based on their educational level? 3) How is the pattern of students' conceptual understanding and preconception regarding the topic?

Research Methodology

Respondent

The research employed non-experimental quantitative descriptive approach, in ways that the researcher did not manipulate or regulate the learning process and materials. The study was conducted during the even semester in 2019-2020 academic year. The respondents were 987 people, comprised of students of eleventh grade from eight junior high schools as well as university students of chemistry department in Northern Sulawesi, Indonesia. The distribution of respondents is displayed in following Table 1.

Table 1

Demographic profile of respondents (N=947)

Demography	Code	Respondents	Percentage (%)
Gender			
Male	M	320	67.68
Female	F	667	32.42
Education level			
X Class students	M	168	17.02
XI Class students	N	473	47.92
XII Class students	O	186	18.84
University students from chemistry department	P	160	16.21

The respondents were chosen randomly and have voluntarily agreed to participate within the research. In addition, they received no learning treatment and other special treatments that allow them to complete the measurement instrument.

Instrument and Procedures Development

The design process refers to recommendation by Wilson (2005) that consists of four key steps: definition of construct map, item design, result blank, and measurement model.

Phase 1: Definition of construct map The map offers substantive definition of measured constructs; the more constructs measured, the constructs' level will vary qualitatively (Wilson, 2009). In simpler words, it aims to develop the students' understanding map to measure the students' progress (Wilson, 2012). The instrument involved variables i.e., the students conceptual understanding and preconception in elaborating the characteristics of particle of matter; it is conducted in accordance with the Curriculum Standard of Chemistry Subject in Tenth Grade in Indonesia, as presented in Table 2. Instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures, instrument and procedures.

Table 2

Conceptual Understanding Level

Level 3 The students are able to connect between characteristics of particle of matter in macroscopic and submicroscopic level	
Phenomenon	Evaporation: item Q6/Bubble
10. Preconception	Air bubble consists of Hydrogen and Oxygen particles
9. Preconception	Air bubble is water-soluble
Phenomenon	Condensation: item Q5/Dew
8. Preconception	Water drops come from melting ice that penetrates the glass wall
7. Preconception	Water drops are the result of reaction between ice and air nearby the glass
Level 2 The students are able to determine SMRs diagram of particle structure during change of form: item Q11/SMRs/SL; Q12/SMRs/LG; Q13/SMRs/GS; Q25/SMRs/GG	
6. Preconception	The SMRs diagram of particle structure follows the physical form of matter
5. Preconception	The SMRs diagram of O_2 molecule shape undergoes change as a result of increase in the volume of container.
Level 1 The students are able to determine characteristics of particle of matter during change process of matter's form.	
4. Preconception	The particle size of matter changes into (large/small) as a result of change in matter form: item Q1/PS/SL; Q7/PS/LG; Q14/PS/LG; Q18/PS/SG; Q22/PS/GG
3. Preconception	The particle mass of matter changes into (large/small) due to change in matter form: item Q2/PM/SL; Q8/PM/LG; Q15/PM/LG; Q19/PM/SG; Q24/PM/GG
2. Preconception	Distance between matter particles changes into (faster/slower) due to change in matter form: item Q3/DP/SL; Q9/DP/SL; Q16/DP/LG; Q20/DP/SG; Q23/PM/GG
1. Preconception	Motion between matter particles changes into (dense/loose) due to change in matter form: item item Q4/PMo/SL; Q10/PMo/LG; Q17/PMo/LG; Q21/PMo/SG

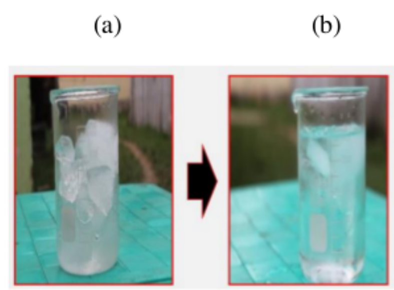
Variation in conceptual understanding level illustrates the development process of the students' conceptual understanding. In the first level, the students were asked to determine particle characteristics (size, mass, motion, and distance) in change process of matter form. In the second level, the students were asked to determine submicroscopic representation diagram of particle structure. Further, in the third level, the students were asked to connect between characteristics of particle of matter in macroscopic and submicroscopic level. In each level, the construct map also features the students' tendency of preconception.

Phase 2: item design and evaluation The phase involves determination process of items to be used in acquiring evidences of students' construct understanding regarding the construct map (Wilson, 2005). Certain items may have different extent of effectiveness to measure students' conceptual understanding, (Mintzes, Wandersee, & Novak, 1999), however, multiple choices item is considered more practical and effective (Wilson, 2008). The instrument of concept understanding test of particle (or TPKP) is adapted from multiple choices instruments by Herrmann-Abell &

DeBoer (2011). Each item consists of two distractor answer choices and one open answer choice. The distractor answer choices are designed by referring to the common preconceptions by the students (see Table 2) as logical choices to distract the students from the correct one. The distractors function to emphasize the item diagnostic strength (Sadler, 1998). Some of the items are adopted from previous studies Osborne & Cosgrove (1983), Renström, Andersson, & Marton (1990); Devetak et al., (2004); Tóth & Kiss (2006); Davidowitz et al., (2010); Devetak & Glažar, (2010); Slapničar, Devetak, Glažar, & Pavlin (2017) and Yildirir & Demirkol (2018).

Figure 1

Sample of item Q1/PS/SL design



Glass (a) contains ice chunks, glass (b) contains melting ice chunks. How is the size of water particle in solid form (ice) compared to that in liquid form?

a. Size of a water particle in solid form > a water particle in liquid form.
b. Size of a water particle in solid form < a water particle in liquid form.
c. Size of a water particle in solid form = a water particle in liquid form.
d. Other answers

The Picture 1 displays sample of item Q1/PS/SL design, in which Q1 is the number of item 1, PS is particle size, and SL is solid-liquid. The item measures student's capability in determining particle size in form change from solid to liquid. The choice A and B are distractors, the correct choice is C, and choice D is for other answers students may fill if the existing answer choices are not in accordance with their initial knowledge. Every correct answer is given mark 1, and wrong answers get 0 mark. Each student only have slight probability of 0.25 in choosing the right answer. The students will pick what they think the right answer based on their understanding. If the distractor item choice functions well, the students will not be able to predict the correct answer.

Phase 3: design of result blank, i.e., the correlation between construct map and items (Wilson, 2005). This phase aims to identify whether the answer the students pick correlates with their conceptual understanding; in simpler terms, it is intended to elaborate the conformity between the variable contents being measured. In order to elaborate the previous aspect, the TPKP instrument is validated by three independent experts and tested to the students to acquire their feedback. The process acquires 25 items of TPKP. Prior to the data collection process, it is ensured that all students have received formal education on the characteristics of particle of matter and its changes. The students' response towards the instrument is inputted manually by written answer sheet. The test were supervised by the teachers in school by referring to the agreed permission and duration. Each student is required to finish all test items within the allocated duration of 45 minutes. The instrument sheets are further collected, and checking process is conducted to ensure that the amount of instrument sheet is the same with participating students.

Phase 4: Rasch model analysis approach The analysis integrat algorithm as a result of probabilistic expectation of item 'i' and student 'n', as : The statement is the probability of student n in item i to result in correct answer (x = 1); with student ability, ßn, and item difficulty level (Bond & Fox, 2015). The above equation is simplified by inserting logarithm function, into , so that the probability of picking the right answer equals to student's ability subtracted by item difficulty level. The student (person) and item units are considered in the same interval scale and are independent to each other. The students' ability level and item difficulty level are measured in logarithm unit, namely odds or log that variates from -00 to +00 (Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015). The instrument efficiency, when compared to the item distribution towards item difficult level with distribution of student's ability level, is quantifiable in order to measure the students' conceptual understanding. In addition, the student's understanding level is differentiated based on the item size. The previous steps highlight the main difference of Rasch model analysis when compared to the raw score-based conventional one; the latter lacks accuracy in evaluating students' ability observed from different item difficulty level (Lu & Bi, 2016; Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015).

Data Analysis

The study employed WINSTEPS version 3.75 software to convert raw data into interval data nacre, 2012; Bond and Fox, 2015). The conversion result acts as the calibration of data on student's ability level and item difficulty level within the same interval measurement. Moreover, the analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding by Differential Item Functioning (DIF) item test; and 3) diagnosis of students' preconception by estimation of item response pattern through option probability curve test. Data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis, data analysis.

Research Results

Effectiveness of Measuring Instruments

Person and Item Reliability. The first step to elaborate the effectiveness of measuring instruments is by measuring the person and item reliability. This is conducted to gather information to what extent the measurement produces consistent information in displaying latent trait or the unidimensionality of the measured variable (Sumintono & Widhiarso, 2015). The analysis result is presented in the form of statistical summary (Table 3). Research results, research results, research results, research results, research results, research results, research results.

Table 3

Summary of fit statistics

Parameter (N)	INFIT		OUTFIT		Separation	Reliability	Measure	KR-20	SD
	MNSQ	ZSTD	MNSQ	ZSTD					
Person (987)	1,00	-0,1	1,02	-0,1	1,55	0,71	-0,34	0,72	0,88
Items (25)	1,00	-0,8	1,02	-0,1	8,18	0,99	0,00		0,60

The above table indicates that the person reliability value of 0.71 is equivalent to the person separation index value of 1.55. This is to say that the consistency of students' response towards the test is deemed good. In addition to that, it is generated that the *Cronbach Alpha Coefficient* (KR-20) value is 0.72, signifying good interaction between students and the test. This further indicates strong correlation between the students' response towards the item, in the context that the students' knowledge tends to be non-fragmented, enabling it to be measured (Adams & Wieman, 2011). To the researchers and educational practitioners, such information is essential to prepare for follow-up plans and development of students' ability (Wei et al., 2012). Moreover, the result generates relatively high value of item separation index of 8.18 that is equivalent to the item reliability value of 0.99. This indicates very good item consistency, or the item is deemed capable to meet the unidimensionality criteria. In other words, the item performs very good in defining the measured variable. This is confirmed by the infit and outfit value result, in which most of the items are in the acceptable range for the multiple choice test (Herrmann-Abell & DeBoer, 2011; Trevor G. Bond and Christine M. Fox, 2015).

Figure 2

Function of Measurement Information

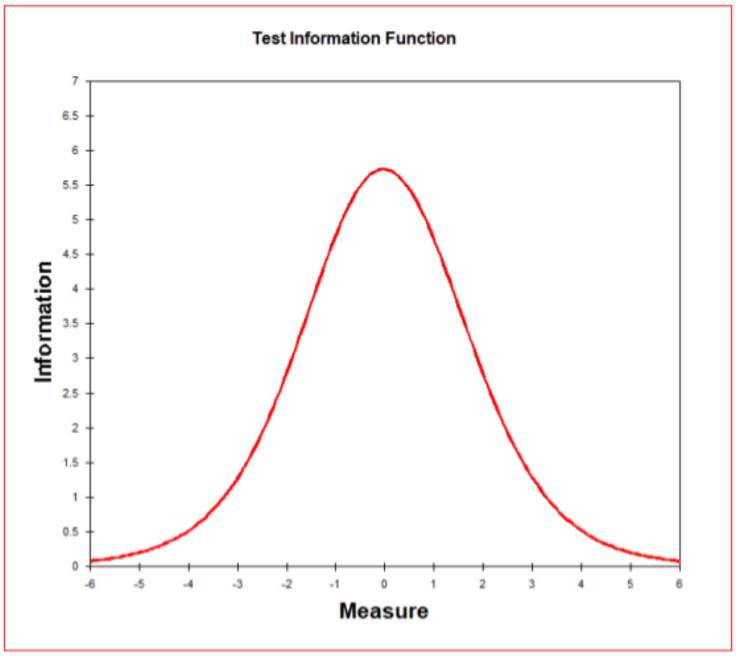


Figure 2 displays graph of measurement information in order to show the measurement reliability. The higher the tip of information function graph, the measurement reliability value is likely to increase. In the intermediate level of students' ability (-3.0 logit up to +3.0 logit), the measurement information is in very high spot. This indicates that the TPKP instrument is capable of producing optimal information to students with intermediate level of ability. This means that the instrument possesses high measurement reliability (Misbach & Sumintono, 2014; Sumintono & Widhiarso, 2014; Trevor G. Bond and Christine M. Fox, 2015).

Validity. The next step is to measure the item validity by Fit item test to ensure that all items fit with the Rasch model. The process is aimed to identify whether or not the test item are able to measure the aspects that intended to be measured, or test validity (Linacre, 2012; Sumintono, 2018). The criteria used comprise outfit means-square (MNSQ): $0.5 < y < 1.5$; outfit z-standard: $-2.0 < Z < +2.0$, as well as point measure correlation (PTMEA Corr). The PTMEA Corr is the correlation between score of item and person measure that is required to be positive value and not approaching zero (Trevor G. Bond & Christine M. Fox, 2015). The PTMEA Corr criteria: $0.4 < x < 0.8$. If all three criteria are not met, this signifies that the item is not good enough and need further elaboration

(Boone, Yale, & Staver, 2014). Both Outfiit MNSQ and Infit MNSQ are sensitive chi-squares in detecting outlier response pattern. There are two outlier response: the right response, guessed by the students with low ability in item with high difficulty level; or the wrong response due to the high-ability students' carelessness in items with low difficulty level. The expected ideal MNSQ value is 1.0. The analysis result on item appropriateness is displayed in Figure 3 as follows:

Figure 3

Item Statistics: Misfit Order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
25	298	987	.60	.07	1.26	7.0	1.40	7.5	A .07	.37	65.6	73.6	Q6/Bubble
6	250	987	.88	.08	1.16	3.7	1.27	4.4	B .18	.36	74.5	77.3	Q2/PM/SL
8	287	987	.66	.08	1.12	3.3	1.20	3.8	C .22	.36	73.6	74.4	Q15/PM/LG
3	235	987	.97	.08	1.03	.8	1.18	2.8	D .30	.36	79.6	78.5	Q14/PS/LG
4	279	987	.71	.08	1.07	1.8	1.15	2.9	E .28	.36	76.2	75.0	Q18/PS/SG
24	549	987	-.63	.07	1.06	2.6	1.15	3.7	F .27	.34	65.8	64.7	Q5/Dew
2	287	987	.66	.08	1.06	1.6	1.14	2.8	G .29	.36	75.2	74.4	Q7/PS/LG
7	288	987	.65	.08	1.07	2.0	1.11	2.1	H .28	.36	73.9	74.3	Q8/PM/LG
1	264	987	.79	.08	1.00	-.1	1.06	1.1	I .35	.36	78.1	76.1	Q1/PS/SL
10	365	987	.25	.07	1.04	1.4	1.04	1.1	J .33	.37	67.1	69.1	Q24/PM/GG
9	268	987	.77	.08	.99	-.3	1.03	.6	K .36	.36	78.5	75.8	Q19/PM/SG
11	509	987	-.44	.07	1.01	.5	1.00	-.1	L .34	.35	63.4	64.4	Q3/DP/SL
17	430	987	-.07	.07	.98	-.8	.98	-.5	M .38	.36	68.8	66.0	Q10/PMo/LG
22	465	987	-.24	.07	.98	-1.1	.98	-.6	l .38	.36	67.6	64.9	Q13/SMRs/GS
12	482	987	-.32	.07	.97	-1.6	.95	-1.6	k .39	.36	66.0	64.6	Q9/DP/LG
16	557	987	-.66	.07	.96	-2.0	.93	-1.8	j .39	.34	67.5	65.0	Q4/PMo/SL
23	560	987	-.68	.07	.94	-2.9	.91	-2.4	i .41	.34	68.2	65.0	Q25/SMRs/GG
13	515	987	-.47	.07	.94	-3.1	.91	-2.8	h .42	.35	68.6	64.4	Q16/DP/LG
15	508	987	-.44	.07	.92	-3.7	.93	-2.1	g .43	.35	69.6	64.4	Q23/DP/GG
21	549	987	-.63	.07	.92	-3.8	.87	-3.5	f .44	.34	68.6	64.7	Q12/SMRs/LG
19	557	987	-.66	.07	.92	-4.0	.89	-2.9	e .43	.34	70.1	65.0	Q21/PMo/SG
18	566	987	-.71	.07	.91	-4.4	.87	-3.5	d .44	.34	71.4	65.2	Q17/PMo/LG
5	430	987	-.07	.07	.90	-4.4	.87	-3.9	c .47	.36	70.4	66.0	Q22/PS/GG
20	473	987	-.27	.07	.90	-4.9	.87	-4.0	b .47	.36	69.9	64.7	Q11/SMRs/SL
14	554	987	-.65	.07	.86	-6.6	.83	-4.6	a .49	.34	75.3	64.9	Q20/DP/SG
MEAN	421.0	987.0	.00	.07	1.00	-.8	1.02	-.1			70.9	68.9	
S.D.	120.7	.0	.60	.00	.09	3.2	.14	3.1			4.4	5.1	

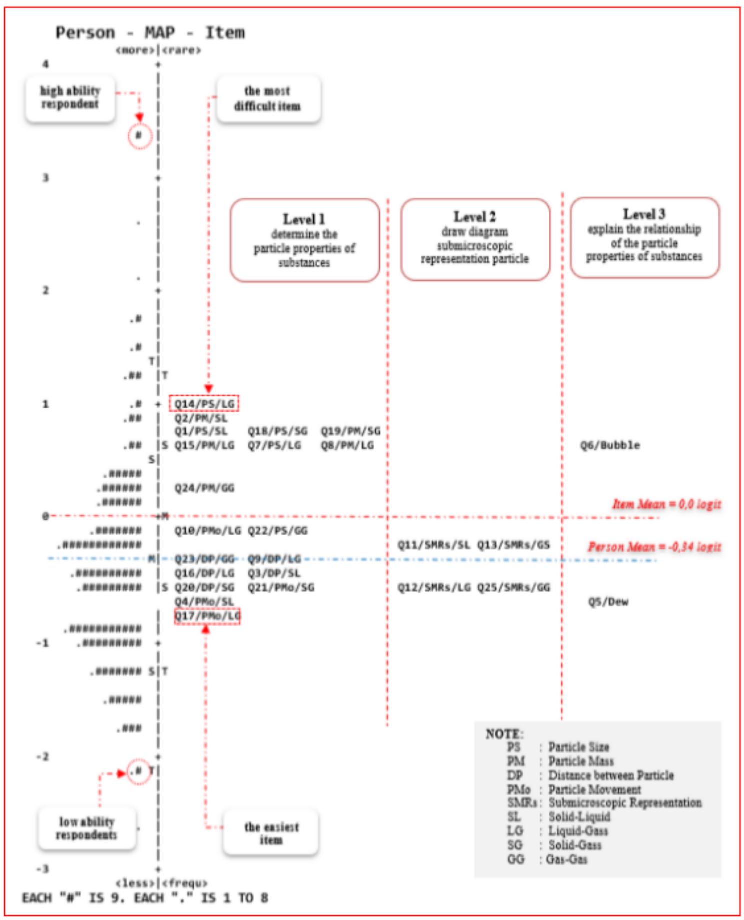
From the previous Item Statistics, it is generated that all items meet the Outfit MNSA criteria and no negative PTMEA Corr occurs. This means that all items are not deviant, appropriate, and valid. Despite some items do not meet one of the criteria, this by no means decrease the quality of the items. For instance, item (Q6/Bubble, Q2/PM/SL, and Q15/PM/LG) do not meet the criteria of Outfit Z Standard and PTMEA Corr; item (Q1/PS/SL, Q24/PM/GG and Q19/PM/SG) do not meet the criteria of PTMEA Corr; and item (Q25/SMRs/GG, Q16/DP/LG, and Q23/DP/GG) do not meet the criteria of Outfit ZSTD; this is supposedly caused by large size of sample, or N > 500 (Boone et al., 2014).

Wright Map: Person-Map-Item. The third step is to measure the consistency of item difficulty level and student's ability test constructed in Table 2. The higher the item difficulty level, the higher also the student's ability level will result. Information of Wright Map: Person-Map-Item is displayed in Figure 4. The previous Wright map generates that all instrument items encompass almost all the students' ability. The map generates variance from students with very high ability (> 3.0 logit), to those with very low ability (< -2.0 logit) as well. In addition to that, disparity (in which there is no item that is appropriate with the student's ability) is observed within the interval of -3.0

logit up to -0.5 logit and in the interval of +1.0 logit up to +3.7 logit. This signifies that the information generated within the interval range is somewhat limited and needs further elaboration. On the other hand, the item difficulty level is mostly located in the interval of -1.0 logit up to +1.0 logit; what is more, the items tend to occur in the same difficulty level. The item Q14/PS/LG is the most difficult item with logit of +0.97, while item Q17/Pmo/LG is the easiest item with logit of -0.71.

Figure 4

Wright Map: Person-Map-Item

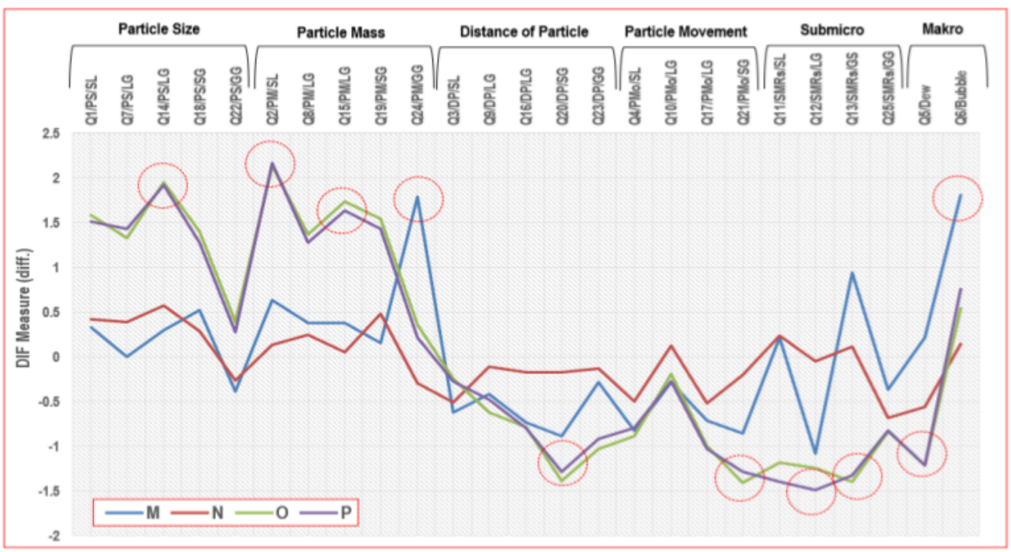


As observed from the differences in item size, some interesting cases are explained as follows: **Firstly**, the items in level 1: Q14/PS/LG (0.97) > Q1/PS/SL (0.79) > Q18/PS/PG (0.71) > Q7/PS/GG (0.66) are instead assumed by the students to possess different difficulty level. The items above, however, are more difficult than item Q6/Bubble in level 3 (0.60). In other words, determining a particle size is more difficult than explaining the particle characteristics of matter in evaporation phenomenon. **Secondly**, the size of item Q5/Dew (-0.63) < item Q6/Bubble; this indicates that it is harder for the students to elaborate the particle characteristics of matter in evaporation phenomenon than in condensation phenomenon, despite that both items are in the same level. **Thirdly**, the size of following items: Q2/PM/SL (0.88) > Q19/PM/SG (0.77) > Q15/PM/LG (0.66) > Q8/PM/LG (0.65) > Q24/PM/GG in level 1 is larger compared to that of items Q13/SMRs/GS (-0.24) > Q11/SMRs/SL (-0.27) > Q12/SMRs/LG (-0.63) > Q25/SMRs/GG (-0.68) in level 2. The finding illustrates that it is harder for the students in determining the particle mass than determining submicrorepresentation (SMRs) diagram in different form changes of matter. The previous cases identifies disparity in students' conceptual understanding, signifying that the level of understanding in particle characteristics of matter is relatively low. Overall, 80% of test item difficulty level is relatively parallel with the measured constructs. By that, the test possess good construct validity (Blanc & Rojas, 2018; Lu & Bi, 2016; Neumann, Neumann, & Nehm, 2011).

Disparity in Conceptual Understanding Level

The next step is the measurement of disparity of students' conceptual understanding in the focused topic based on educational level by Differential Item Functioning (DIF).

Figure 5
Person DIF plot based on educational level



Description: M = X Class students,
N = XI Class students,
O = XII Class students, and
P = University students from chemistry department

Figure 5 of DIF plot based on students' educational level depicts that ten items are identified to possess significant disparity. **Firstly**, five curves approaching the upper limit are items with high difficulty level (Q14/PS/LG, Q2/PM/SL, Q15/PM/SG, Q24/PM/GG and Q6/Bubble); while five curves approaching the lower limit are items with low difficulty level (Q20/DP/SG, Q21/PMo/SG, Q12/SMRs/LG, Q13/SMRs/GS and Q5/Dew). **Secondly**, the item Q14/PS/LG (particle size in form change of liquid-gas), Q2/PM/SL (particle mass in form change of solid-liquid), and Q15/PM/SG (particle mass in change form of solid-gas) are deemed very hard by the students of XII class and the university students compared to students in X and XI class. **Thirdly**, the study discovers different result for item Q24/PM/GG and Q6/Bubble. The item Q24/PM/GG (particle mass of O2 in larger volume) and Q6/Bubble (constructing elements of air bubbles during boiling process of water) are deemed very hard for X class students compared to students in XI and XII classes, as well as university students. **Fourthly**, the items Q20/DP/SG (distance between particles in form change of solid-gas), Q21/PMo/SG (motion between particles in form change of solid-gas), Q12/SMRs/LG (SMRs diagram of particle in form change of liquid-gas), Q13/SMRs/GS (SMRs diagram of particle in change form of gas-liquid), and Q5/Dew (condensation) are deemed too easy for students in XII class and university students compared to the students in X and XI classes.

Pattern of Conceptual Understanding and Preconception

The analysis on pattern of conceptual understanding and preconception employs option probability curve test (Linacre, 2012). Option probability curve aims to display probability of picking every answer choice to elaborate the performance level of all students in the measured items (Abell & DeBoer, 2011). The test relies on the principle that the curve of correct answer will rise along with the decrease of the curve of distractor choices (Haladyna, 2004; Trevor G. Bond and Christine M. Fox, 2015). For items that are influenced by distractor options, the curve produced tends to be non-parallel with the traditional monotonous item behavior (Sadler, 1998), by this reason, each answer choice is analyzed separately.

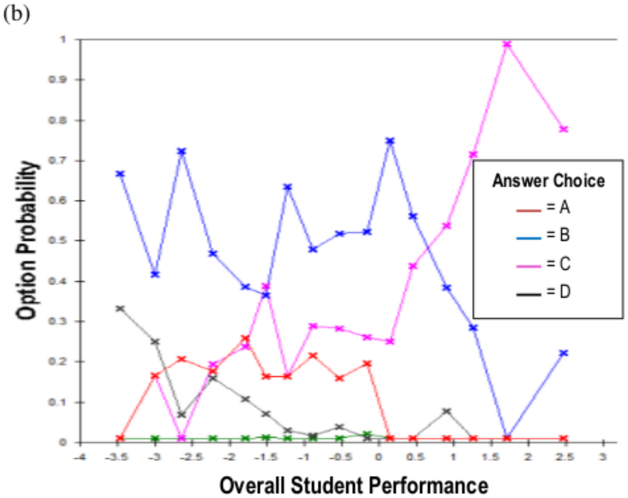
The instrument provides four answer choices, therefore, four curves are formed. Each curve displays the students' comprehension. Students with low ability tend to pick distractor choice, while students whose high ability are more likely to prefer other preconceptions (Abell and DeBoer, 2011; Perera, Sumintono, & Jiang, 2018). Below is the elaboration of pattern of students' conceptual understanding and preconception based on four option probability curves.

Figure 6

(a) sample of item Q2/PM/SL, (b) option probability curve

(a)
When some ice cubes in the glass melts, some other ice cubes are seen floating on the water surface. How is the comparison between mass of one particle of ice and one particle of water?

a. Mass of one particle of ice is bigger than mass of one particle of water.
b. mass of one particle of ice is smaller than mass of one particle of water.
c. mass of one particle of ice is similar with mass of one particle of water.
d Other answer...

(b)


First example, the item Q2/PM/SL (0.88) is shown un Figure 6(a). The item measures students' capability in determining particle size in form change from solid to liquid. The option probability curve is displayed in Figure 6(b). Students with low ability (< 0.5 logit) tend to pick distractor choice B (mass of one particle of ice is smaller than mass of one particle of water) or A (mass of one particle of ice is bigger than mass of one particle of water). In addition, students with very low ability (< -1.0 logit) tend to pick D (other answer). Some students with relatively low ability (> -2.5 logit), however, pick the right answer C (mass of one particle of ice is similar with mass of one particle of water). One can predict the response pattern of students with low ability, as the distractors A, B, and D contain third preconception in level 1 (see Table 2). The students possess the knowledge that mass of particle of matter can change into larger or smaller size by observing the matter's change of form. It is interesting to note that there are students with high ability (>2.0) that pick B; this indicates the presence of resistant preconception.

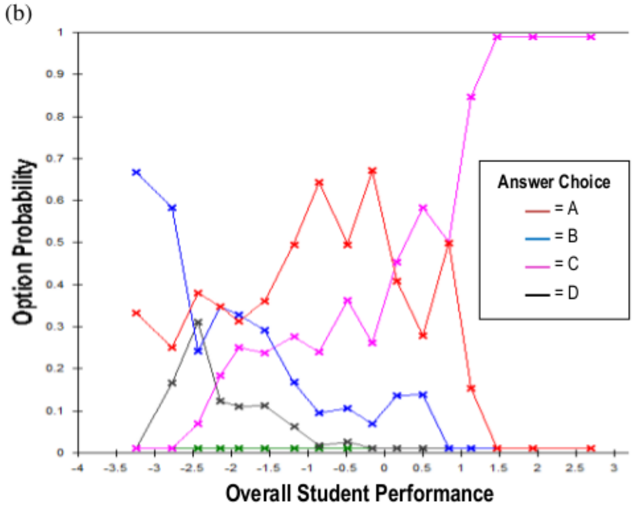
Figure 7
(a) item Q8/PM/LG; (b) option probability curve

(a)

During boiling process of water, a change of form of water from liquid to gas occurs. How is the comparison between mass of one particle of water and one particle of water vapor?

a. Mass of one water particle is bigger than mass of one vapor particle
b. mass of one water particle is smaller than mass of one vapor particle.
c. mass of one water particle is similar with mass of one vapor particle.
d Other answer...

(b)



Second sample, item Q8/PM/LG (0.65) is shown in figure 7(a) as the item to measure students' ability in determining mass of particle in form change of liquid-gas. The option probability curve is displayed in Figure 7(b). The curve of distractor B (mass of one water particle is smaller than mass of one vapor particle) is chosen by students with low ability (< -2.0 logit), while curve of choice A (mass of one water particle is bigger than mass of one vapor particle) is chosen by students with ability in range of -3.5 to 1.5 logit. The correct answer, option C (mass of one water particle is similar with mass of one vapor particle), is chosen by students with ability in > -2.5 logit. As highlighted in the table, the decline of curve of distractor A is followed by the increase of curve of right answer C; both curves intersect in level of 1.0 logit. The shape of curve A indicates the presence of resistant preconception type-three in level 1.

It depicts that the particular item response pattern that signifies students' conceptual understanding pattern in the given level. Moreover, the curve shape of distractors A and B in the items Q2/PM/SL and Q8/PM/SL tend to have identical pattern. The finding indicates that students with either low or high ability have consistent preconception that mass of particle can change into larger or smaller in size along with the change in matter form.

Third sample, item Q5/Dew (-0.63) as shown in Figure 8(a) measures the students' ability in elaborating characteristics of particle in condensation phenomenon. The option probability curve is displayed in Figure 8(b). Students with low ability (< 1.0 logit) tend to pick distractor A (water drops come from liquid of melting ice that breaks through the glass wall) and option D (other answer). Some students with high ability (> 1.0 logit) also pick distractor B (water drops are the result of reaction between ice and air nearby the glass). The shape of curve B is wavy and non-linear, even in the interval of 2.0 to 4.0 logit, it can reach option probability value up to 1.0 logit. This is regarded as deviation from the right answer C (water drops come from condensing water vapor nearby the glass). A worth note, however, is to consider in the unstable, wavy shape of curve C; this indicates the students' inconsistency (particularly those with high ability) in comprehending concept of condensation. This shows that students have their own preconception regarding concept of condensation.

Figure 8

(a) item Q5/Dew; (b) option probability curve

(a)

In a glass filled with ice, you can see water drops at the glass. According to you, where do the water drops come from?

a. Water drops come from melting ice that penetrates the glass wall
b. water drops are the result of reaction between ice and air nearby the glass
c. Water drops come from condensing water vapor nearby the glass
d Other answer...

(b)

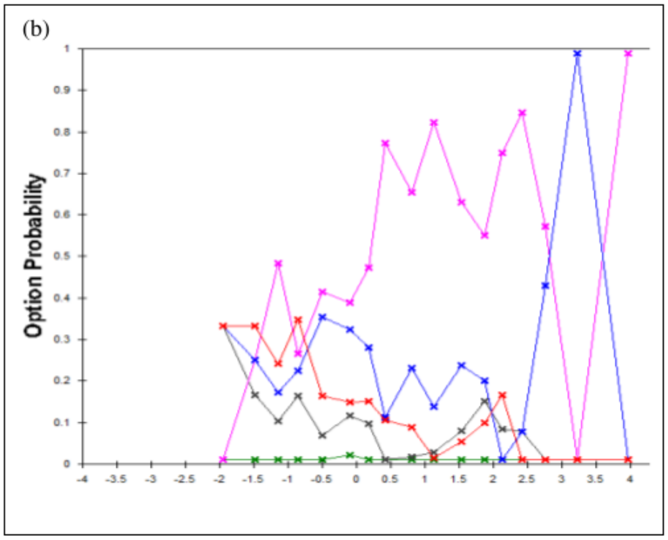


Figure 9

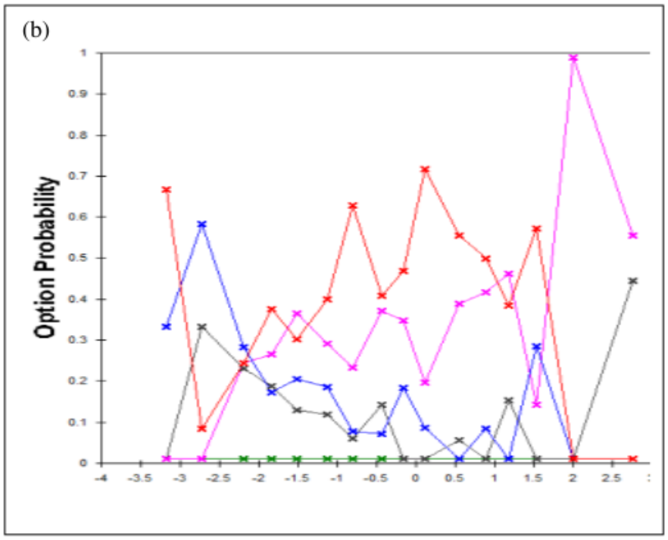
(a) item Q6/Bubble; (b) option probability curve

(a)

In a container filled with boiling water, you can see air bubbles on the top of it. According to you, what are the composing elements of the air bubbles?

a. Hydrogen and Oxygen particles
b. Air that is dissolved in water
c. Water molecules
d. Other answer...

(b)



Fourth sample, item Q6/Bubble as shown in Figure 9(a) measures the students' ability in elaborating characteristics of particle in evaporation phenomenon. The option probability curve is displayed in Figure 9(b). The distractor A (air bubbles are Hydrogen and Oxygen particles) is dominantly chosen by students whose ability in range between -3.0 to 2.0 logit. Moreover, the distractor B (air bubbles are Hydrogen and Oxygen particles) is dominantly chosen by students whose ability in range between -3.0 to 0.5 logit. The form of curve A and B picked by students with low ability is predictable. The curve of right answer C (air bubbles are water molecules), however, shows interesting hint; in the interval range of -2.5 to 3.0 logit, the tip of curve shows up-and-down pattern. Moreover, in the level of 1.5 logit, the curve shape of distractors A and B shows decline pattern, while that of curve C tends to increase. Another finding worth noting is that the curve D (other answers) is picked by some students with high ability (> 2.0 logit). This indicates that the particular students have their own preconception regarding evaporation concept.

Discussion

The research findings indicate that the instruments have good effectiveness, meet the requisites of person and item reliability, and show good construct validity. When applied in evaluating students' conceptual understanding, it is found that: Firstly, almost all students with high ability face difficulty in understanding concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, or determining concept of particle regarding evaporation and condensation phenomena in level 3. Secondly, the information of response pattern of students with high ability is quite consistent, repetitive, and systematic in particular items. This indicates the presence of permanent and latent preconception. The analysis of option probability curve of item Q2/PM/SL (0.88), Q8/PM/LG (0.65), Q5/Dew (-0.63) and Q6/Bubble (0.60) indicates that the approach of item response pattern is able to explore in detail and comprehensively regarding students' conceptual understanding and preconception.

Sequences of verification conducted that involves Rasch model approach shows detailed, accurate, and quantifiable results since the approach integrates development procedure of diagnostic and summative instruments. Several samples of preconception, e.g., item Q2/PM/SL (0.88) and Q8/PM/LG (0.65) indicate that distractor options are potential to be elaborated further in order to investigate tendency of preconception by the students. In addition, it also provides information regarding main idea unknown to the students and their degree of misunderstanding.

The approach employed in this study is an effective illustration to help teacher in evaluating the learning process as well as the students' learning progress. This is due to the integration of qualitative item development procedure and quantitative data analysis, allowing the teachers to explore in-depth on the students' understanding, concepts the students understand and/or do not understand, and misconception.

Conclusions and Implications

The measuring instrument developed performed well in its validity and reliability, thus, is deemed applicable in measuring students' conceptual understanding and preconception in elaborating particle characteristics of matter. During implementation of the instruments, the study finds out that: 1) almost all students with high ability face difficulty in understanding concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, as well as determining concept of particle regarding evaporation and condensation phenomena in level 3.

2) There is significant disparity between students' conceptual understanding as based on their educational level. 3) On certain cases, it is found that the distractor item response pattern by high-ability students tend to be consistent, indicating a certain tendency of resistant preconception pattern.

The development of diagnostic instrument with Rasch model approach is deemed as literacy process for practitioners and researchers in Indonesia. The result indicates that there is no single item that is parallel with both highest ability and lowest ability students. This calls for further elaboration in order to improve the instrument items quality. Moreover, an anomaly is found that students with high ability (> 1.0 logit) tend to pick distractor choices. This urges further studies to investigate structured comprehension problems.

Acknowledgements

This research is funded by the Directorate of Research and Community Outreach/ (Direktorat Riset dan Pengabdian Masyarakat or DRPM), Directorate General of Research and Development (Dirjen Riset dan Pengembangan) of Ministry of Research and Technology of the Republic of Indonesia through Penugasan Penelitian Dasar Unggulan Perguruan Tinggi Grant Program, Research and Community Outreach Institute (Lembaga Penelitian dan Pengabdian Masyarakat) of Universitas Negeri Gorontalo in 2020.

References

Adadan, E., & Savasci, F. (2012). An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, *34*(4), 513–544. https://doi.org/10.1080/09500693.2011.636084

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

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

Alamina, J. I., & Etokeren, I. S. (2018). Effectiveness of Imagination Stretch Teaching Strategy in Correcting Misconceptions of Students about Particulate Nature of Matter. *Journal of Education, Society and Behavioural Science*, *27*(1), 1–11. https://doi.org/10.9734/jesbs/2018/43063

Allen, M. (2014). Misconceptions in Primary Science. In *McGraw-Hill Education (UK)*.

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

Blanc, A., & Rojas, A. J. (2018). Use of rasch person-item maps to validate a theoretical model for measuring attitudes toward sexual behaviors. *PLoS ONE*, *13*(8), 1–17. https://doi.org/10.1371/journal.pone.0202551

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch Analysis in the Human Sciences. In *Rasch Analysis in the Human Sciences*. https://doi.org/10.1007/978-94-007-6857-4

Boz, Y. (2006). Turkish pupils' conceptions of the particulate nature of matter. *Journal of Science Education and Technology*, *15*(2), 203–213. https://doi.org/10.1007/s10956-006-9003-9

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The Development of Two Tier Multiple-Choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Levels of Representation. *Chemistry Education Research and Practice*, *8*(3), 293–307.

Cheng, M. M. W. (2018). Students' visualisation of chemical reactions-insights into the particle model and the atomic model. *Chemistry Education Research and Practice*, *19*(1), 227–239.

https://doi.org/10.1039/c6rp00235h

Davidowitz, B., Chittleborough, G., & Murray, E. (2010). Student-generated submicro diagrams: A useful tool for teaching and learning chemical equations and stoichiometry. *Chemistry Education Research and Practice*, *11*(3), 154–164. https://doi.org/10.1039/c005464j

Devetak, I., & Glažar, S. A. (2010). The influence of 16-year-old Students' Gender, Mental Abilities, and motivation on their Reading and Drawing Submicrorepresentations Achievements. *International Journal of Science Education*, *32*(12), 1561–1593. https://doi.org/10.1080/09500690903150609

Devetak, I., Urbancic, M., Grm, K. S. W., Krnel, D., & Glazar, S. A. (2004). Submicroscopic Representations As a Tool for Evaluating Students' Chemical Misconception. *Acta Chim. Slov*, *51*(December 2004), 799–814.

Femintasari, V. (2015). the Effectiveness of Two-Tier Multiple Choice Test and Multiple Choice Test Followed With Interview in Identifying Misconception of Students With Different Scientific Reasoning Skills. *Jurnal Ilmu Pendidikan*, *21*(2), 192–197.

Gabel, D. L. (1993). Use of the particle nature of matter in developing conceptual understanding. *Journal of Chemical Education*, *70*(3), 193. https://doi.org/10.1021/ed070p193

Hadenfeldt, J. C., Bernholt, S., Liu, X., Neumann, K., & Parchmann, I. (2013). Using Ordered Multiple-Choice Items To Assess Students' Understanding of the Structure and Composition of Matter. *Journal of Chemical Education*, *90*(12), 1602–1608. https://doi.org/10.1021/ed3006192

Haladyna, T. M. (2004). *Developing and Validating Multiple-choice Test Items*. https://doi.org/10.4324/9780203825945

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

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

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

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

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

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

Kind, V. (2004). Beyond Appearances: Students' misconceptions about basic chemical ideas 2 nd Edition. *School of Education*, 1–84.

Laliyo, Botutihe, & Panigoro. (2019). The development of two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237.

https://doi.org/10.26803/ijlter.18.9.12

Linacre, J. M. (2012). *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 3.75.0*. https://doi.org/ISBN 0-941938-03-4

Liu, X. (2012). Developing Measurement Instruments for Science Education Research. In C. J. Fraser, Barry, Tobin, Kenneth, McRobbie (Ed.), *Second International Handbook of Science Education* (pp. 651–665). Springer Netherlands,.

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

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, *34*(15), 2317–2341. https://doi.org/10.1080/09500693.2012.684433

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (1999). Assessing Science Understanding: A Human Constructivist View. In Joel J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Academic Press, San Diego, California : Elsevier Academic Press*. Retrieved from https://zodml.org/sites/default/files/%5BJoel_J._Mintzes%2C_James_H._Wandersee%2C_Joseph_D._No_0.pdf

Misbach, I. H., & Sumintono, B. (2014). Pengembangan dan validasi instrumen "Persepsi Siswa tehadap Karakter Moral Guru" di Indonesia dengan Model Rasch [Instrument Development and Validation of "Students' Perception of Teacher's Moral Characters in Indonesia" by Rasch Model]. *Seminar Nasional Psikometri*, (May).

Naah, B. M., & Sanger, M. J. (2012). Student misconceptions in writing balanced equations for dissolving ionic compounds in water. *Chem. Educ. Res. Pract.*, *13*(3), 186–194. https://doi.org/10.1039/C2RP00015F

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*(9), 825–838. https://doi.org/10.1002/tea.3660200905

Ozalp, D., & Kahvecib, A. (2015). Diagnostic assessment of student misconceptions about the particulate nature of matter from ontological perspective. *Chemical Education Research and Practice*, *16*, 619–639. https://doi.org/10.1039/C5RP00096C

Özgür Kapici, H., & Akcay, H. (2016). Particulate Nature of Matter Misconceptions Held By Middle and High School Students in Turkey. *European Journal of Education Studies*, *2*(8), 43–58. https://doi.org/10.5281/zenodo.163547

Ozmen, H. (2011). Turkish Primary Students´ Conceptions about the Particulate Nature of Matter. *International Journal of Environment & Science Education*, *6*(1), 99–121.

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

Renström, L., Andersson, B., & Marton, F. (1990). Students' Conceptions of Matter. *Journal of Educational Psychology*, *82*(3), 555–569. https://doi.org/10.1037/0022-0663.82.3.555

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

Slapničar, M., Devetak, I., Glažar, A. S., & Pavlin, J. (2017). Identification of the understanding of the states of matter of water and air among Slovenian students aged 12, 14 and 16 years through solving authentic tasks. *Journal of Baltic Science Education*, *16*(3), 308–323. Retrieved from https://www.researchgate.net/publication/318215148

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

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

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

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi MOdel Rasch untuk Penelitian Ilmu-Ilmu Sosial [Application of Rasch Model in Social Sciences Research]* (B. Trim, ed.). Retrieved from https://www.researchgate.net/publication/268688933%0AAplikasi

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan [Application of Rasch Model in Education Assessment]*. Retrieved from https://www.researchgate.net/publication/282673464%0AAplikasi

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

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

Tóth, Z., & Kiss, E. (2006). Using particulate drawings to study 13-17 year olds' understanding of physical and chemical composition of matter as well as the state of matter. *Practice and Theory in Systems of Education*, *1*(1), 109–125. Retrieved from http://eduscience.fw.hu/

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169. https://doi.org/10.1080/0950069880100204

Trevor G. Bond and Christine M. Fox. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences. In *Routledge Taylor & Francis Group* (Third Edit, Vol. 44). https://doi.org/10.1088/1751-8113/44/8/085201

Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, *4*(6), 626–631. Retrieved from http://www.academicjournals.org/SRE

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335–345. https://doi.org/10.1021/ed100852t

Wilson, M. (2005). Constructing Measures: An Item Response Modeling Approach. In *Lawrence Erlbaum Associates: Mahwah, NJ, 2005*. https://doi.org/10.4324/9781410611697

Wilson, M. (2008). Cognitive Diagnosis Using Item Response Models. *Zeitschrift Für Psychologie / Journal of Psychology*, *216*(2), 74–88. https://doi.org/10.1027/0044-3409.216.2.74

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

Wilson, M. (2012). Responding to a Challenge that Learning Progressions Pose to Measurement Practice. In Alicia C. Alonzo and & Amelia Wenk Gotwals (Eds) (Eds.), *Learning Progression in Science* (pp. 317–344). https://doi.org/10.1007/978-94-6091-824-7

Yaşar, I. Z., İnce, E., & Kırbaşlar, F. G. (2014). 7. Class Science and Technology Course “Structure of Atom” Subject Readiness Improvement Test. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2014.09.259

Yildirir, H. E., & Demirkol, H. (2018). Identifying Mental Models of Students for Physical and Chemical Change. *Journal of Baltic Science Education*, *17*(6), 986–1004. https://doi.org/10.33225/jbse/18.17.986

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

ORIGINALITY REPORT

5%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.jbse.webinfo.lt Internet	61 words — 1%
2	www.ijicc.net Internet	23 words — < 1%
3	Shanshan Lu, Hualin Bi. "Development of a measurement instrument to assess students' electrolyte conceptual understanding", Chemistry Education Research and Practice, 2016 Crossref	18 words — < 1%
4	www.project2061.org Internet	17 words — < 1%
5	Prasanna Ramakrisnan, Azizah Jaafar, Noor Faezah Mohd Yatim, Mohd Noor Mamat. "Validating Instrument Quality for Measuring Students' Acceptance of an Online Discussion Site (ODS)", 2013 International Conference on Advanced Computer Science Applications and Technologies, 2013 Crossref	16 words — < 1%
6	www.2014.icemst.com Internet	11 words — < 1%
7	Huanhuan Lu, Yanxia Jiang, hualin Bi. "Development of a measurement instrument to assess students' proficiency levels regarding galvanic cells", Chemistry Education Research and Practice, 2020	10 words — < 1%

Crossref

8	etd.lib.metu.edu.tr Internet	9 words — < 1%
9	www.tandfonline.com Internet	9 words — < 1%
10	hrcak.srce.hr Internet	9 words — < 1%
11	repository.ung.ac.id Internet	8 words — < 1%
12	eprints.uad.ac.id Internet	8 words — < 1%
13	Vassilia Hatzinikita, Vasilis Koulaidis. "Pupils' Ideas on Conservation during Changes in the State of Water", Research in Science & Technological Education, 1997 Crossref	8 words — < 1%
14	researchcommons.waikato.ac.nz Internet	8 words — < 1%
15	Internet	8 words — < 1%
16	onlinelibrary.wiley.com Internet	8 words — < 1%
17	www.sociologicalscience.com Internet	8 words — < 1%
18	file.scirp.org Internet	8 words — < 1%
19	library.oum.edu.my Internet	8 words — < 1%
20	Grazia Messineo, Salvatore Vassallo. "A Rasch analysis of Math tests using M.In.E.R.Va. platform",	8 words — < 1%

2014 International Conference on Education Technologies and Computers (ICETC), 2014
Crossref

21 www.emeraldinsight.com
Internet
8 words — < 1%

22 researchdirect.uws.edu.au
Internet
8 words — < 1%

23 www.aensiweb.com
Internet
8 words — < 1%

24 www.scribd.com
Internet
8 words — < 1%

25 Cooper, Melanie M., Leah M. Corley, and Sonia M. Underwood. "An investigation of college chemistry students' understanding of structure-property relationships : STUDENTS' UNDERSTANDING OF STRUCTURE-PROPERTY", Journal of Research in Science Teaching, 2013.
Crossref
8 words — < 1%

26 etd.lsu.edu
Internet
8 words — < 1%

27 my.unisa.ac.za
Internet
8 words — < 1%

28 Herrmann-Abell, Cari F., and George 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", Chemistry Education Research and Practice, 2011.
Crossref
7 words — < 1%

29 "Overcoming Students' Misconceptions in Science", Springer Science and Business Media LLC, 2017
Crossref
7 words — < 1%

30 "Rasch Measurement", Springer Science and Business Media

LLC, 2020
Crossref

6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		

Explanation / rebuttal letter

No.	Comments from Reviewers	Section & Page	Explanation/Rebuttal
1	Up to 200 words	Abstract: Page 1	The abstract has been reduced to no more than 200 words. (DONE)
2	Change the word "study" to "research"	Abstract: Page 1	It has been changed according to the reviewer's request. (DONE)
3	In alphabetical order	Keywords: Page 1	It has been changed according to the reviewer's request. (DONE)
4	Avoid "of"	Keywords: Page 1	It has been changed according to the reviewer's request. (DONE)
5	According to APA Manual 7th ed., the in-text citation for works with three or more authors is now shortened right from the first citation. You only include the first author's name and "et al.": https://www.scribbr.com/apa-style/apa-seventh-edition-changes/	Introduction: Page 1	It has been changed according to the reviewer's request. (DONE)
6	According to APA Manual 7th ed., the in-text citation for works with three or more authors is now shortened right from the first citation. You only include the first author's name and "et al.": https://www.scribbr.com/apa-style/apa-seventh-edition-changes/	Introduction: Page 1	It has been changed according to the reviewer's request. (DONE)
7	The APA Manual (6/7th ed.) says: "Order the citations of two or more works within the same parentheses alphabetically" (6.16 on page 177).	Introduction: Page 1	It has been changed according to the reviewer's request. (DONE)
8	Interval	Introduction: Page 2	It has been changed according to the reviewer's request. (DONE)
9	Rework in a whole text.	Research Problem: Page 2	It has been changed according to the reviewer's request. (DONE)
10	Please consistently use these terms. If you select 'research' or "study", you should exploit only one of them rather than interplayed usage. It is preferable to use the term "research"	Research Questions: Page 3	It has been changed according to the reviewer's request. (DONE)
11	Past simple	Research Questions: Page 3	It has been changed according to the reviewer's request. (DONE)
12	Investigate means: to try to find facts about an incident or a situation in order to know the truth. Investigate is often	Research Questions: Page 3	It has been changed according to the reviewer's request. (DONE)

	related to police investigations.		
13	I do not see any research aim	Research Questions: Page 3	It has been elaborated in the research questions. (DONE)
14	Research Methodology • Past tense to describe what was done. It may be past active or past passive. • Present tense for presenting diagrams and figures	Research Methodology : Page 3	It has been changed according to the reviewer's request. (DONE)
15	General description of research is important in order to show the basis of the research. It is like a very brief introduction to the methodology section as a whole. Research design should be explained. Indicate also research type, time (when was it carried out?) and scope. Research approach, a theoretical framework, etc. should be shortly described. It is not correct to start from the participants.	Research Methodology : Page 3	It has been changed according to the reviewer's request. (DONE)
16	What ethical procedures were followed?	Research Methodology : Page 4	It has been explained according to the reviewer's question. (DONE)
17	Figure?	Figure 1: Page 6	Yes, Figure. Previously it was written “picture 1” and has been changed. (DONE)
18	Research Results • Past tense for results obtained • Present tense to refer to figures, tables and graphs	Research Results: Page 7	It has been revised according to the reviewer's request. (DONE)
19	Italic	Table 3: Page 7	It has been revised according to the reviewer's request. (DONE)
20	Period	Table 3: Page 7	It has been revised according to the reviewer's request. (DONE)
21	Period	Table 3: Page 7	It has been revised according to the reviewer's request. (DONE)
22	The APA Manual says: “Do not use a zero before a decimal fraction when the statistic cannot be greater than 1 (e.g., correlations, proportions, and levels of statistical significance, probability)” (4.35 on page 113).	Table 3: Page 7	It has been revised according to the reviewer's request. (DONE)
23	Poor quality	Table 4: Page 9	In the previous manuscript table was figure 3. It has been changed to Table 4

			for the revision. (DONE)
24	Poor quality	Figure 3: Page 10	In the previous manuscript table was figure 4. It has been changed to Figure 3 for the revision. (DONE)
25	In the hard copy of the journal, colours will become gray, and the graphs will be indistinguishable. Mark them explicitly and reliably for printing in black-and-white, add the explanation.	Figure 4: Page 11	In the previous manuscript table was figure 5. It has been changed to Figure 4 for the revision. (DONE)
26	An ampersand	Pattern of Conceptual Understandin g and Preconceptio n: Page 12	It has been changed according to the reviewer's request. (DONE)
27	In the hard copy of the journal, colours will become gray, and the graphs will be indistinguishable. Mark them explicitly and reliably for printing in black-and-white, add the explanation.	Figure 5: Page 12	In the previous manuscript table was figure 6. It has been changed to Figure 5 for the revision (and sequentially the figure names after also change). and revised as the reviewer's suggestion. (DONE)
28	Discussion • Present tense to explain significance of results • Past tense to summarize findings, with present tense to interpret results	Discussion: Page 15	It has been changed according to the reviewer's request. (DONE)
29	There is an imitation of discussion APA 6/7th. Ed. Manual reads: "Similarities and differences between your results and the work of others should be used to contextualize, confirm, and clarify your conclusions." It means you are expected to cite other authors' papers in this section.	Discussion: Page 16	It has been added according to the reviewer's request. (DONE)
30	It must be clear what is the added value of this research to previous reports. It is obligatory to make sure that reported new information is significant for international level readers of JBSE.	Conclusions: Page 17	It has been added according to the reviewer's request. (DONE)
31	The reference is incomplete. What is the journal issue and/or pages? Or volume? Publisher /if needed/Editor(s) etc. Check format for chapters from books, if needed.	References: Page 17-21	The references have been corrected according to the reviewer's request

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

200419.jbse_JL-suo_review

Abstract. *This research aimed to evaluate the students' conceptual understanding and to diagnose the students' preconceptions in elaborating the particle characteristics of matter by development of diagnostic instrument as well as Rasch model response pattern analysis approach. Data were acquired by 25 multiple-choice written test items distributed to 987 students in North Sulawesi, Indonesia. Analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding; and 3) diagnosis of students' preconception by estimation of item response pattern. The result generated information on the diagnostic and summative measurement on students' conceptual understanding in elaborating the topic; information also acts as empirical evidence on the measurement's reliability and validity. Moreover, the result discovered a significant disparity between students' conceptual understanding based on their educational level. It was found that the distractor item response pattern tended to be consistent, indicating a certain tendency of resistant preconception pattern. The findings are expected to be a recommendation for future researchers and educational practitioners that integrates diagnostic and summative measurement with Rasch model in evaluating conceptual understanding and diagnosing misconception.*
Keywords: *conceptual understanding, item response, particle, Rasch model*.

Introduction

Central to the notion of learning about characteristics of a particle of matter is the process of developing an understanding on abstract concepts (Johnstone, 1991) without directly interacting with the object/fact (Stojanovska et al., 2012); therefore it is considered a difficult subject for the students to learn. Echoing this, the disparity in understanding is almost inevitable (Kapici & Akcay, 2016) since different students may develop ones' own distinctive way of understanding a concept (Yildirir & Demirkol, 2018). The idea is also coined by experts as misconception (Johnstone, 2006, 2010; Taber, 2015), or alternative framework and preconception (Lu & Bi, 2016). The experts have discovered that students always have their own preconception that is not in line with scientific concepts (Alamina & Etokeren, 2018; Yaşar et al., 2014); therefore, one needs to conduct identification and improvement on the conceptual learning (Allen, 2014; Soeharto et al., 2019).

In diagnosing preconceptions, several researchers have developed diagnostic instruments in different mechanisms (McClary & Bretz, 2012), i.e., conceptual map, essay test, interview, essay test with interview, or multiple-choice test (Femintasari et al., 2015). Two-step multiple choice diagnostic test (Adadan & Savasci, 2012; Chandrasegaran et al., 2007; Treagust, 1988; Tüysüz, 2009) is preferred due to its ability to diagnose preconception and describe the underlying reasons. The instrument is indeed considered qualitatively effective in elaborating differences in students' thought processes; however, it does not provide summative measurement features due to lack of internal consistency and the instrument's unidimensionality (Lu & Bi, 2016). In addition to that, the measurement conclusion generated is considered weak due to extracted from analysis on the raw score (Sumintono, 2018)

Studies on preconception have found that the concept is somewhat resistant. In the early 2000s, it is discovered that students' preconceptions persisted even when they already undergo formal education experience (Hoe & Subramaniam, 2016). Preconception can also change along with the development of students' conceptual understanding; it also varies in different levels of understanding (Aktan, 2013). If one conducts a two-step test and raw score analysis approach to diagnose resistant preconception, the result generated will only provide limited feedback information (Sumintono, 2018) due to the instrument's limitation in measuring students' conceptual understanding. Instead of supporting, the information will only make it harder for teachers to implement proper instructional decisions (Wilson, 2008).

During the middle of the 2000s, the Rasch model analysis was commonly used in studies of chemistry education (Herrmann-Abell & DeBoer, 2011; Liu, 2012; Wein et al., 2012). The approach provides a testing apparatus that integrates diagnostic and summative measurement. Recently, this approach is used to develop formative assessment with the intention to conduct learning construction mapping, e.g., measuring the students' way of constructing their understanding process (Hadenfeldt et al., 2013). It is worth to note, however, that there are studies that integrate diagnostic and summative measurement with a different approach (Hoe & Subramaniam, 2016); despite that, trends in chemistry education studies highlight that diagnostic-summative measurement by Rasch model analysis is more common to be carried out (Laliyo et al., 2019; Lu & Bi, 2016).

Research Problem

The characteristics of a particle of matter is a fundamental concept in chemistry, usually taught in middle education level. Adequate comprehension regarding the particle characteristics of matter both in macroscopic and microscopic level is essential as the knowledge basis in understanding more advanced topics such as the concept of atoms and molecules as the submicroscopic component that is invisible to plain eyesight but exists in all real-world phenomena (Cheng, 2018; Ozmen, 2011; Yildirir & Demirkol, 2018). The fact signifies the relevance and reasoning of complexity in chemistry learning that is considered difficult for both students and teachers to conduct (Alamina & Etokeren, 2018). In simpler terms, to ensure that the chemistry learning is conducted effectively, one requires to nurture students' comprehensive understanding regarding particle characteristics of matter and its change of state.

To evaluate the students' conceptual understanding on the aforementioned topic, one also needs to measure the students' capability in interpreting particle state during change process of a matter's form (Alamina & Etokeren, 2018; Barbera, 2013; Boz, 2006; Cheng, 2018; Gabel, 1993; Hadenfeldt et al., 2013; Kapici & Akcay, 2016; Kind, 2004; Naah & Sanger, 2012; Ozalp & Kahvecib, 2015; Ozmen, 2011; Renström et al., 1990; Slapničar et al., 2017; Stojanovska et al., 2012; Yildirir & Demirkol, 2018). Researches on particle characteristics and changes of matter generally employ diagnostic instruments in the form of essay tests and/or essays followed by interview; the instruments are further analyzed based on raw score results. The approach is considered inefficient and somewhat lacked accuracy in measuring students' conceptual understanding and misconception pattern. Despite its ineffectiveness, the conventional method is used by most teachers in Indonesia to measure and determine students' learning progress. The teachers argue that measuring the students' raw score is effective in determining how far the students have progressed in the learning process. The students' raw score is regarded by many as an early premature indication regarding the measured variable and is not eligible to be the final measurement indicator due to its temporary nature. In addition to that, regarding the decision-making process, the raw score contains only limited information for it to be treated as reference (He et al., 2016; Sumintono & Widhiarso, 2015)

Research Focus

The research focuses on developing a diagnostic instrument that integrates measurement of conceptual understanding and diagnosis of students' preconceptions regarding the aforementioned topic by the approach of Rasch model item response pattern analysis. The analysis employs different test apparatuses to provide extensive information for practitioners and researchers in science education in evaluating students' learning progress in different topics.

Research Questions

This research aimed to figure out the following questions: 1) How is the effectiveness of measurement instrument to evaluate the students' conceptual understanding and diagnose their preconceptions on the characteristics of a particle of matter? 2) Is there any significant difference between students in elaborating on the aforementioned topic based on their educational level? 3) How is the pattern of students' conceptual understanding and preconception regarding the topic?

Research Methodology

General Background

The descriptive-quantitative research employed a non-experimental approach, in which the students' conceptual understanding in explaining the characteristics of a particle of matter was treated as the measurable variable. Prior to conducting the research, it was ensured that the students already experience formal learning of the aforementioned topic. The researchers did not conduct any intervention on the learning process or the learning material. In other words, no treatment was implemented to the students for them to be able to answer all test items in the measurement instrument.

The data collection step was implemented for four months in the even semester of the 2019-2020 academic year; the process was conducted after obtaining approval from the Government of Province of Gorontalo and heads of universities in the Northern part of Sulawesi, Indonesia. Moreover, the schools' and parents' approval was obtained in cooperation with the school committee. The school administrators were willing to facilitate the data collection process that adjusted with the schedule.

Respondent

The respondents were 987 people consisting of students of eleventh grade from eight junior high schools as well as university students of the chemistry department in Northern Sulawesi, Indonesia. The distribution of respondents is displayed in Table 1 below.

Table 1
Demographic profile of respondents (N=947)

Demography	Code	Respondents	Percentage (%)
Gender			
Male	M	320	67.68
Female	F	667	32.42
Education level			
X Class students	M	168	17.02
XI Class students	N	473	47.92
XII Class students	O	186	18.84
University students from the chemistry department	P	160	16.21

The respondents were chosen randomly and have voluntarily agreed to participate in the research. In addition, they received no learning treatment and other special treatments that allow them to complete the measurement instrument. Students were asked to write down their responses in the answer sheet; the process was supervised by teachers in the respective schools and lecturers in the respective university. All students were instructed to answer all questions in the instruments within 45 minutes. All instrument sheets and answer sheets were collected by the researchers shortly after the session ended; it was ensured that the numbers of instruments matched the numbers of participants.

Instrument and Procedures Development

The design process refers to a recommendation by Wilson (2005), which consists of four key steps: definition of construct map, item design, result blank, and measurement model.

Phase 1: Definition of construct map. The map offers a substantive definition of measured constructs; the more constructs measured, the constructs' level will vary qualitatively (Wilson, 2009). In simpler words, it aims to develop the students' understanding map to measure the students' progress (Wilson, 2012). The instrument involved variables, i.e., the students conceptual understanding and preconception in elaborating the characteristics of a particle of matter; it was conducted in accordance with the Curriculum Standard of Chemistry Subject in Tenth Grade in Indonesia, as presented in Table 2.

Table 2
Conceptual Understanding Level

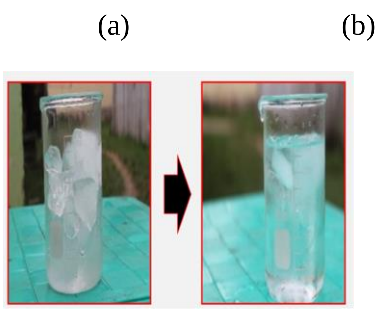
Level 3 The students are able to connect between characteristics of a particle of matter in macroscopic and submicroscopic level	
Phenomenon	Evaporation: item Q6/Bubble
10. Preconception	Air bubble consists of Hydrogen and Oxygen particles
9. Preconception	Air bubble is water-soluble
Phenomenon	Condensation: item Q5/Dew
8. Preconception	Water drops come from melting ice that penetrates the glass wall
7. Preconception	Water drops are the result of the reaction between ice and air nearby the glass
Level 2 The students are able to determine SMRs diagram of particle structure during a change of form: item Q11/SMRs/SL; Q12/SMRs/LG; Q13/SMRs/GS; Q25/SMRs/GG	
6. Preconception	The SMRs diagram of particle structure follows the physical form of matter
5. Preconception	The SMRs diagram of O_2 molecule shape undergoes change as a result of an increase in the volume of the container.

Level 1 The students are able to determine the characteristics of a particle of matter during the change process of matter’s form.	
4. Preconception	The particle size of matter changes into (large/small) as a result of change in matter form: item Q1/PS/SL; Q7/PS/LG; Q14/PS/LG; Q18/PS/SG; Q22/PS/GG
3. Preconception	The particle mass of matter changes into (large/small) due to change in matter form: item Q2/PM/SL; Q8/PM/LG; Q15/PM/LG; Q19/PM/SG; Q24/PM/GG
2. Preconception	Distance between matter particles changes into (faster/slower) due to change in matter form: item Q3/DP/SL; Q9/DP/SL; Q16/DP/LG; Q20/DP/SG; Q23/PM/GG
1. Preconception	Motion between matter particles changes into (dense/loose) due to change in matter form: item item Q4/PMo/SL; Q10/PMo/LG; Q17/PMo/LG; Q21/PMo/SG

Variation in conceptual understanding level illustrates the development process of the students’ conceptual understanding. In the first level, the students were asked to determine particle characteristics (size, mass, motion, and distance) in the change process of matter form. In the second level, the students were asked to determine the submicroscopic representation diagram of particle structure. Further, in the third level, the students were asked to connect between characteristics of a particle of matter at the macroscopic and submicroscopic level. In each level, the construct map also features the students’ tendency of preconception.

Phase 2: item design and evaluation The phase involved the determination process of items to be used in acquiring evidence of students’ construct understanding regarding the construct map (Wilson, 2005). Certain items may have a different extent of effectiveness to measure students’ conceptual understanding (Sadler, 1999); however, multiple choices item is considered more practical and effective (Wilson, 2008). The instrument of concept understanding test of the particle (or TPKP) is adapted from multiple-choice instruments by (Herrmann-Abell & DeBoer, 2011). Each item consists of two distractor answer choices and one open answer choice. The distractor answer choices are designed by referring to the common preconceptions by the students (see Table 2) as logical choices to distract the students from the correct one. The distractors function to emphasize the item diagnostic strength (Sadler, 1998). Some of the items are adopted from previous studies Osborne & Cosgrove (1983), Renström et al., (1990); Devetak et al., (2004); Tóth & Kiss (2006); Davidowitz et al., (2010); Devetak & Glažar (2010); Slapničar et al., (2017) and (Yildirir & Demirkol, 2018).

Figure 1
Sample of item Q1/PS/SL design



Glass (a) contains ice chunks, glass (b) contains melting ice chunks. How is the size of water particle in solid form (ice) compared to that in liquid form?

a. Size of a water particle in solid form $>$ a water particle in liquid form.
b. Size of a water particle in solid form $<$ a water particle in liquid form.
c. Size of a water particle in solid form $=$ a water particle in liquid form.
d. Other answers

Figure 1 displays a sample of item Q1/PS/SL design, in which Q1 is the number of item 1, PS is particle size, and SL is solid-liquid. The item measures student's capability in determining particle size in form change from solid to liquid. The choice A and B are distractors, the correct choice is C, and choice D is for other answers students may fill if the existing answer choices are not in accordance with their initial knowledge. Every correct answer was given mark 1, and wrong answers got 0 mark. Each student only has a slight probability of 0.25 in choosing the right answer. The students will pick what they think the right answer based on their understanding. If the distractor item choice functions well, the students will not be able to predict the correct answer.

Phase 3: design of result blank, i.e., the correlation between construct map and items (Wilson, 2005). This phase aimed to identify whether the answer the students pick correlates with their conceptual understanding; in simpler terms, it was intended to elaborate the conformity between the variable contents being measured. In order to elaborate on the previous aspect, the TPKP instrument was validated by three independent experts and tested to the students to acquire their feedback. The process acquired 25 items of TPKP. Prior to the data collection process, it was ensured that all students had received formal education on the characteristics of a particle of matter and its changes. The students' response towards the instrument was inputted manually by the written answer sheet. The test was supervised by the teachers in school by referring to the agreed permission and duration. Each student was required to finish all test items within the allocated duration of 45 minutes. The instrument sheets were further collected, and checking process was conducted to ensure that the amount of instrument sheet wass the same with participating students.

Phase 4: Rasch model analysis approach. The analysis integrates algorithm as a result of probabilistic expectation of item 'i' and student 'n', as: The statement is the probability of student n in item i to result in the correct answer (x = 1); with student ability, ßn, and item difficulty level (Bond & Fox, 2015). The above equation was simplified by inserting logarithm function, into , so that the probability of picking the right answer equals to student's ability subtracted by item difficulty level. The student (person) and item units were considered on the same interval scale and were independent of each other. The students' ability level and item difficulty level were measured in the logarithm unit, namely odds or log that variates from -00 to +00 (Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015). The instrument efficiency, when compared to the item distribution towards item difficult level with distribution of student's ability level, was quantifiable in order to measure the students' conceptual understanding. In addition, the student's understanding level was differentiated based on the item size. The previous steps highlighted the main difference of Rasch model analysis when compared to the raw score-based conventional one; the latter lacks accuracy in evaluating students' ability observed from different item difficulty level (Herrmann-Abell & DeBoer, 2011; Lu & Bi, 2016; Sumintono & Widhiarso, 2015).

Data Analysis

The research employed WINSTEPS version 3.75 software to convert raw data into interval data (Bond & Fox, 2015; Linacre, 2012). The conversion result acted as the calibration of data on the student's ability level and item difficulty level within the same interval measurement. Moreover, the analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding by Differential Item Functioning (DIF) item test; and 3) diagnosis of students' preconception by estimation of item response pattern through option probability curve test.

Research Results

Effectiveness of Measuring Instruments

Person and Item Reliability. The first step to elaborate on the effectiveness of measuring instruments was by measuring the person and item reliability. This was conducted to gather information to what extent the measurement produces consistent information in displaying latent trait or the unidimensionality of the measured variable (Sumintono & Widhiarso, 2015). The analysis result is presented in the form of a statistical summary (Table 3).

Table 3
Summary of fit statistics

Parameter (N)	Measure	INFIT		OUTFIT		Separation	Reliability	*SD*	KR-20
		MNSQ	ZSTD	MNSQ	ZSTD				
Person (987)	-.34	1.00	-.11	1.02	-.1	1.55	.71	.88	.72
Items (25)	.00	1.00	-.75	1.02	-.1	8.18	.99	.60	

The above table indicates that the person reliability value of 0.71 is equivalent to the person separation index value of 1.55. This is to say that the consistency of students' response towards the test is deemed good. In addition to that, it is generated that the *Cronbach Alpha Coefficient* (KR-20) value is 0.72, signifying good interaction between students and the test. This further indicates strong correlation between the students' response towards the item, in the context that the students' knowledge tends to be non-fragmented, enabling it to be measured (Adams & Wieman, 2011). To the researchers and educational practitioners, such information is essential to prepare for follow-up plans and development of students' ability (Wei et al., 2012). Moreover, the result generated a relatively high value of item separation index of 8.18 that was equivalent to the item reliability value of 0.99. This indicated very good item consistency or the item was deemed capable of meeting the unidimensionality criteria. In other words, the item performed very good in defining the measured variable. This was confirmed by the infit and outfit value result, in which most of the items were in the acceptable range for the multiple-choice test (Herrmann-Abell & DeBoer, 2011; Bond & Fox, 2015).

Figure 2
Function of Measurement Information

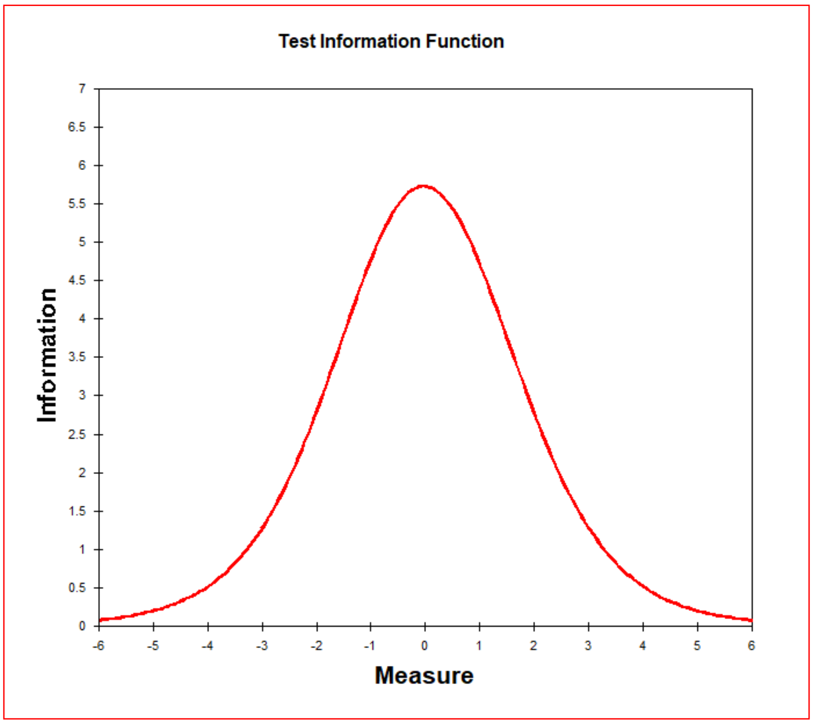


Figure 2 displays the graph of measurement information in order to show the measurement reliability. The higher the tip of information function graph, the measurement reliability value is likely to increase. In the intermediate level of students' ability (-3.0 logit up to +3.0 logit), the measurement information is in very high spot. This indicates that the TPKP instrument is capable of producing optimal information to students with an intermediate level of ability. Such a result means that the instrument possesses high measurement reliability (Bond & Fox, 2015; Kim & Wilson, 2019).

Validity. The next step was to measure the item validity by Fit item test to ensure that all items fit with the Rasch model. The process was aimed to identify whether or not the test item could measure the aspects that intended to be measured, or test validity (Linacre, 2012; Sumintono & Widhiarso, 2014). The criteria used comprise outfit means-square (MNSQ): $0.5 < y < 1.5$; outfit z-standard: $-2.0 < Z < + 2.0$, as well as point measure correlation (PTMEA Corr). The PTMEA Corr is the correlation between the score of item and person measure that is required to be a positive value and not approaching zero (Bond & Fox, 2015). The PTMEA Corr criteria: $0.4 < x < 0.8$. If all three criteria are not met, the item is not good enough and needs further elaboration (Boone et al., 2014). Both Outfiit MNSQ and Infit MNSQ were sensitive chi-squares in detecting outlier response pattern. There were two outlier responses: the right response, guessed by the students with low ability in item with high difficulty level, or the wrong response due to the high-ability students' carelessness in items with a low difficulty level. The expected ideal MNSQ value is 1.0. The analysis result on item appropriateness is displayed in Table 4 as follows:

Table 4

Item Statistics: Misfit Order

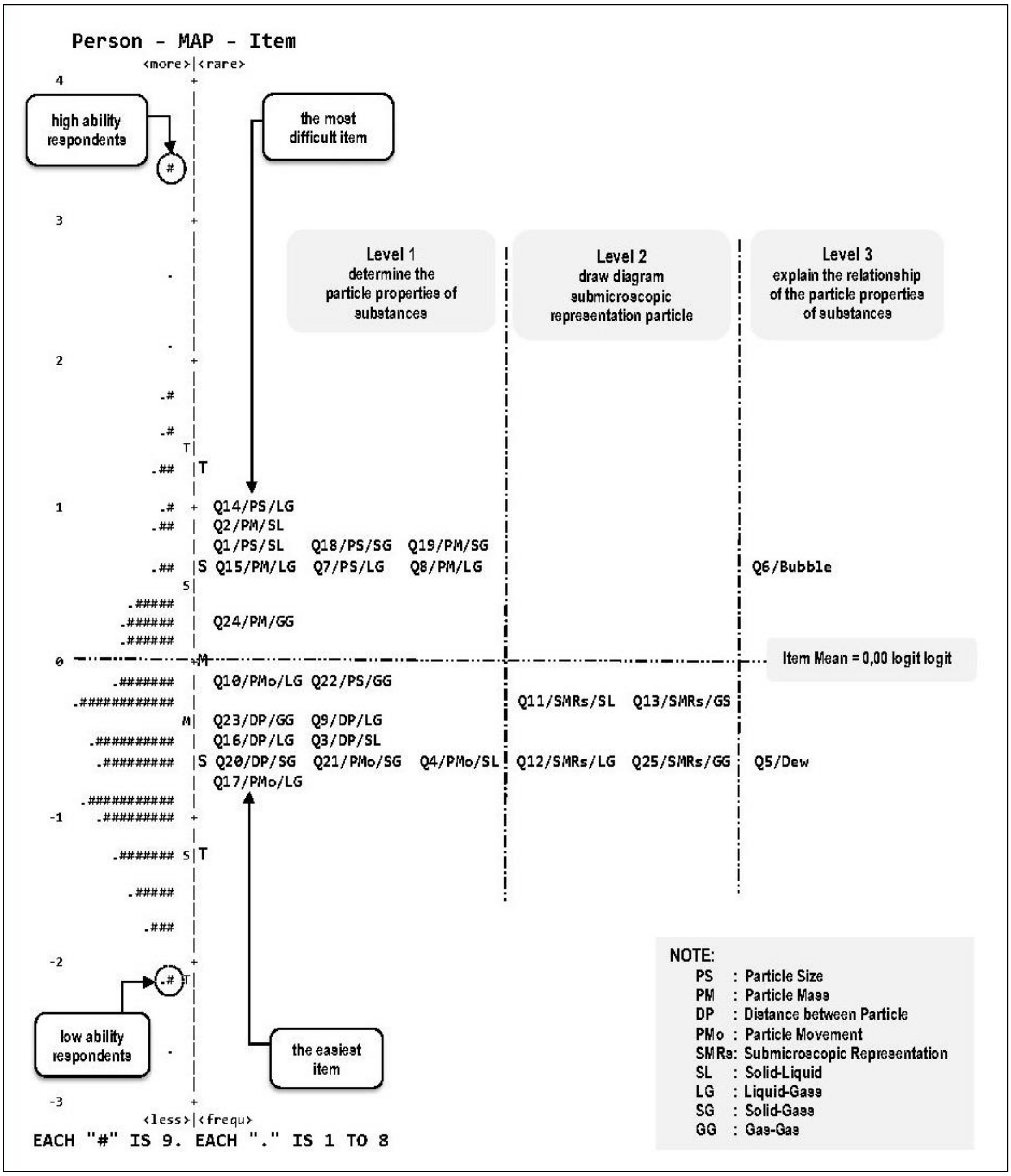
Item	Measure	INFIT		OUTFIT		PTMEA Corr
		MNSQ	ZSTD	MNSQ	ZSTD	
Q6/Bubble	.60	1.26	7.0	1.40	7.5	.07
Q2/PM/SL	.88	1.16	3.7	1.27	4.4	.18
Q15/PM/LG	.66	1.12	3.3	1.20	3.8	.22
Q14/PS/LG	.97	1.03	.8	1.18	2.8	.20
Q18/PS/SG	.71	1.07	1.8	1.15	2.9	.28
Q5/Dew	-.63	1.06	2.6	1.15	3.7	.27
Q7/PS/LG	.66	1.06	1.6	1.14	2.8	.29
Q8/PM/LG	.65	1.07	2.0	1.11	2.1	.28
Q1/PS/SL	.79	1.00	-.1	1.06	1.1	.35
Q24/PM/GG	.25	1.04	1.4	1.04	1.1	.33
Q19/PM/SG	.77	-.3	1.03	.6	.36	.36
Q3/DP/SL	-.44	1.01	.5	1.00	-.1	.34
Q10/PMo/LG	-.07	.98	-.8	.98	-.5	.38
Q13/SMRs/GS	-.24	.98	-1.1	.98	-.6	.38
Q9/DP/LG	-.32	.97	-1.6	.95	-1.6	.39
Q4/PMo/SL	-.66	.96	-2.0	.93	-1.8	.39
Q25/SMRs/GG	-.68	.94	-2.9	.91	-2.4	.41
Q16/DP/LG	-.47	.94	-3.1	.91	-2.8	.42
Q23/DP/GG	-.44	.92	-3.7	.93	-2.1	.43
Q12/SMRs/LG	-.63	.92	-3.8	.87	-3.5	.44
Q21/PMo/SG	-.66	.92	-4.0	.89	-2.9	.43
Q17/PMo/LG	-.71	.91	-4.4	.87	-3.5	.44
Q22/PS/GG	-.07	.90	-4.4	.87	-3.9	.47
Q11/SMRs/SL	-.27	.90	-4.9	.87	-4.0	.47
Q20/DP/SG	-.65	.86	-6.6	.83	-4.6	.49

From the previous Item Statistics, it is generated that all items meet the Outfit MNSA criteria and no negative PTMEA Corr occurs. This means that all items are not deviant, appropriate, and valid. Despite some items do not meet one of the criteria, this by no means decreases the quality of the items. For instance, item (Q6/Bubble, Q2/PM/SL, and Q15/PM/LG) do not meet the criteria of Outfit Z Standard and PTMEA Corr; item (Q1/PS/SL, Q24/PM/GG and Q19/PM/SG) do not meet the criteria of PTMEA Corr; and item (Q25/SMRs/GG, Q16/DP/LG, and Q23/DP/GG) do not meet the criteria of Outfit ZSTD; this is supposedly caused by large size of sample, or N > 500 (Boone et al., 2014).

Wright Map: Person-Map-Item. The third step was to measure the consistency of item difficulty level and student's ability test constructed in Table 2. The higher the item difficulty level, the higher also the student's ability level will result. Information of Wright Map: Person-Map-Item is displayed in Figure 3. The previous Wright map generates that all instrument items encompass almost all the students' ability. The map generates variance from students with very high ability (> 3.0 logit), to those with very low ability (< -2.0 logit) as well. In addition to that, disparity (in which there is no item that is appropriate with the student's ability) was observed within the interval of -3.0 logit up to -0.5 logit and in the interval of +1.0 logit up to +3.7 logit. This signified that the information generated within the interval range was somewhat limited and required further elaboration. On the other hand, the item difficulty level was mostly located in the interval of -1.0 logit up to +1.0 logit;

moreover, the items tended to occur in the same difficulty level. The item Q14/PS/LG was the most difficult item with a logit of +0.97, while item Q17/Pmo/LG was the easiest item with logit of -0.71.

Figure 3
Wright Map: Person-Map-Item



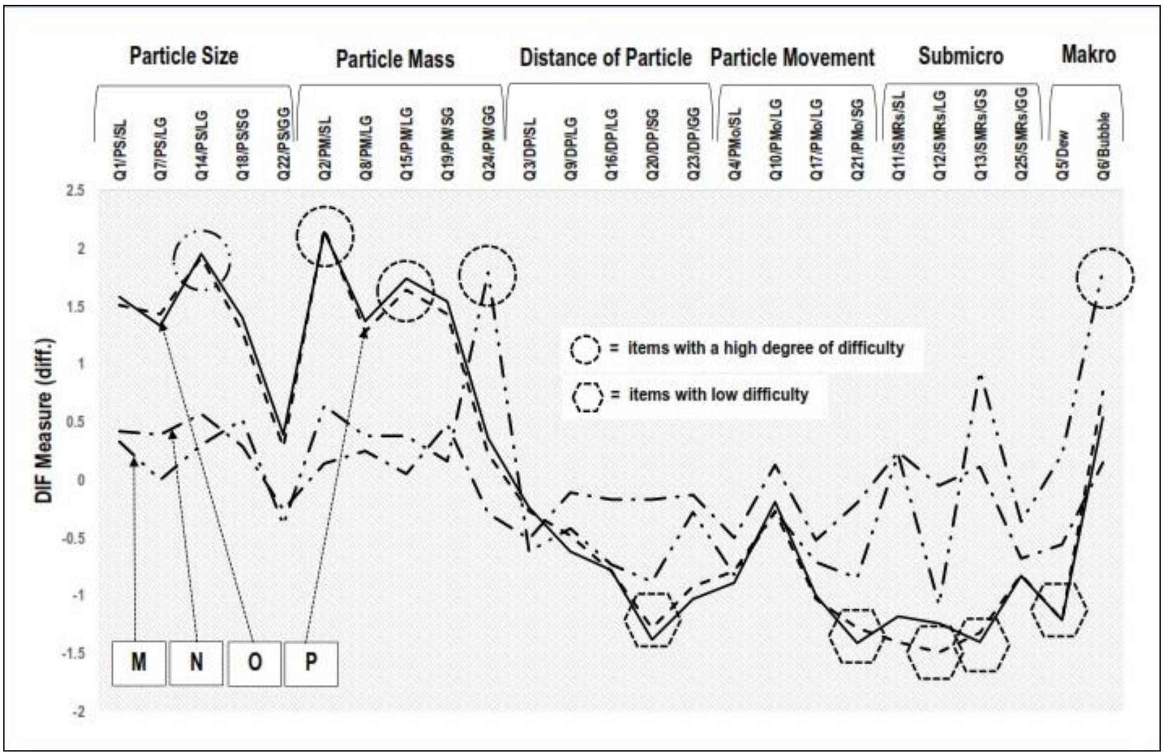
As observed from the differences in item size, some interesting cases were explained as follows: **Firstly**, the items in level 1: Q14/PS/LG (0.97) > Q1/PS/SL (0.79) > Q18/PS/PG (0.71) > Q7/PS/GG (0.66) were instead assumed by the students to possess different difficulty level. The items above, however, were more difficult than item Q6/Bubble in level 3 (0.60). In other words, determining particle size was more difficult than explaining the particle characteristics of matter in the evaporation phenomenon. **Secondly**, the size of item

Q5/Dew (-0.63) < item Q6/Bubble; this indicated that it was harder for the students to elaborate on the particle characteristics of matter in the evaporation phenomenon than in condensation phenomenon, despite that both items were in the same level. **Thirdly**, the size of following items: Q2/PM/SL (0.88) > Q19/PM/SG (0.77) > Q15/PM/LG (0.66) > Q8/PM/LG (0.65) > Q24/PM/GG in level 1 was larger compared to that of items Q13/SMRs/GS (-0.24) > Q11/SMRs/SL (-0.27) > Q12/SMRs/LG (-0.63) > Q25/SMRs/GG (-0.68) in level 2. The finding illustrated that it was harder for the students to determine the particle mass than determining the submicrorepresentation (SMRs) diagram in different form changes of matter. The previous cases identified disparity in students' conceptual understanding, signifying that the level of understanding in particle characteristics of the matter is relatively low. Overall, 80% of test item difficulty level is relatively parallel with the measured constructs. By that, the test possesses good construct validity (Blanc & Rojas, 2018; Lu & Bi, 2016; Neumann et al., 2011).

Disparity in Conceptual Understanding Level

The next step was the measurement of disparity of students' conceptual understanding in the focused topic based on educational level by Differential Item Functioning (DIF).

Figure 4
Person DIF plot based on educational level



Description: M = X Class students,
N = XI Class students,
O = XII Class students, and
P = University students from chemistry department

Figure 4 of DIF plot based on students' educational level depicts that ten items are identified to possess significant disparity. **Firstly**, five curves approaching the upper limit are items with a high difficulty level (Q14/PS/LG, Q2/PM/SL, Q15/PM/SG, Q24/PM/GG and Q6/Bubble); while five curves approaching the lower limit are items with a low difficulty level (Q20/DP/SG, Q21/PMo/SG, Q12/SMRs/LG, Q13/SMRs/GS, and Q5/Dew). **Secondly**, the item Q14/PS/LG (particle size in form change of liquid-gas), Q2/PM/SL (particle mass in form change of solid-liquid), and Q15/PM/SG (particle mass in change form of solid-gas) were deemed very hard by the students of XII class and the university students compared to students in X and XI class. **Thirdly**, the research discovered different results for item Q24/PM/GG and Q6/Bubble. The item Q24/PM/GG (particle mass of O2 in larger volume) and Q6/Bubble (constructing elements of air bubbles during boiling process of water) were deemed very hard for X class students compared to students in XI and XII classes, as well as university students. **Fourthly**, the items Q20/DP/SG (distance between particles in form change of solid-gas), Q21/PMo/SG (motion between particles in form change of solid-gas), Q12/SMRs/LG (SMRs diagram of particle in form change of liquid-gas), Q13/SMRs/GS (SMRs diagram of particle in change form of gas-liquid), and Q5/Dew (condensation) were deemed too easy for students in XII class and university students compared to the students in X and XI classes.

Pattern of Conceptual Understanding and Preconception

The analysis of the pattern of conceptual understanding and preconception employed an option probability curve test (Boone et al., 2014; Linacre, 2012). The option probability curve aims to display the probability of picking every answer choice to elaborate on the performance level of all students in the measured items (Herrmann-Abell & DeBoer, 2011). The test relied on the principle that the curve of the correct answer will rise along with the decrease of the curve of distractor choices (Boone et al., 2014; Haladyna, 2004). For items that are influenced by distractor options, the curve produced tends to be non-parallel with the traditional monotonous item behavior (Sadler, 1998), for this reason, each answer choice was analyzed separately.

The instrument provides four answer choices, thus resulting in four curves. Each curve displays the students' comprehension. Students with low ability tended to pick distractor choice, while students whose high ability were more likely to prefer other preconceptions (Herrmann-Abell & DeBoer, 2011; Perera et al., 2018). Below is the elaboration of the pattern of students' conceptual understanding and preconception based on four option probability curves.

Figure 5

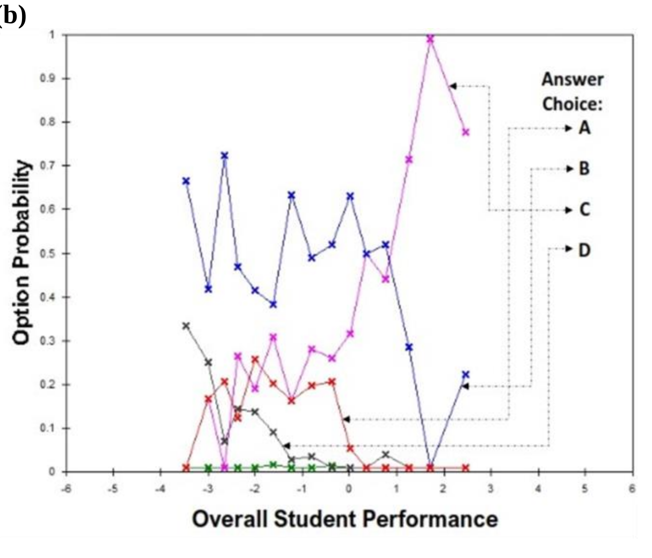
(a) sample of item Q2/PM/SL, (b) option probability curve

(a)

When some ice cubes in the glass melts, some other ice cubes are seen floating on the water surface. How is the comparison between mass of one particle of ice and one particle of water?

A. Mass of one particle of ice is bigger than mass of one particle of water.
B. mass of one particle of ice is smaller than mass of one particle of water.
C. mass of one particle of ice is similar with mass of one particle of water.
D Other answer...

(b)



The first example, i.e., the item Q2/PM/SL (0.88), is shown in Figure 5(a). The item measures students' capability in determining particle size in form change from solid to liquid. The option probability curve is displayed in Figure 5(b). Students with the low ability (< 0.5 logit) tended to pick distractor choice B (mass of one particle of ice is smaller than the mass of one particle of water) or A (mass of one particle of ice is bigger than the mass of one particle of water). In addition, students with very low ability (< -1.0 logit) tended to pick D (other answers). Some students with relatively low ability (> -2.5 logit), however, picked the right answer C (mass of one particle of ice is similar to the mass of one particle of water). One can predict the response pattern of students with low ability, as the distractors A, B, and D contain third preconceptions in level 1 (see Table 2). The students possess the knowledge that mass of particle of matter can change into larger or smaller size by observing the matter's change of form. It is interesting to note that there are students with the high ability (>2.0) who picked B; this indicates the presence of resistant preconception.

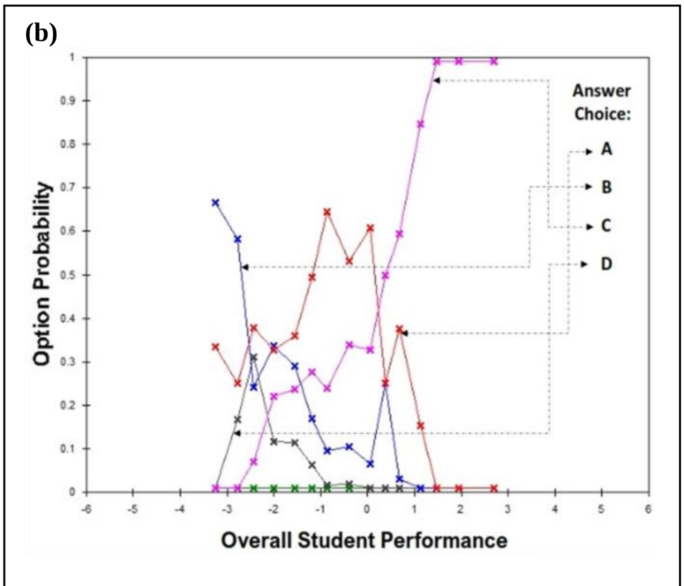
Figure 6

(a) item Q8/PM/LG; (b) option probability curve

(a)

During boiling process of water, a change of form of water from liquid to gas occurs. How is the comparison between mass of one particle of water and one particle of water vapor?

A. Mass of one water particle is bigger than mass of one vapor particle
B. mass of one water particle is smaller than mass of one vapor particle.
C. mass of one water particle is similar with mass of one vapor particle.
D Other answer...



The second sample or item Q8/PM/LG (0.65) is shown in figure 6(a) as the item to measure students' ability to determine the mass of the particle in form change of liquid-gas. The option probability curve is displayed in Figure 6(b). The curve of distractor B (mass of one water particle is smaller than the mass of one vapor particle) is chosen by students with low ability (< -2.0 logit), while the curve of choice A (mass of one water particle is bigger than mass of one vapor particle) was chosen by students with ability in a range of -3.5 to 1.5 logit. The correct answer, option C (mass of one water particle is similar to the mass of one vapor particle), was chosen by students with ability in > -2.5 logit. As highlighted in the table, the decline of the curve of distractor A is followed by the increase of curve of right answer C; both curves intersect in the level of 1.0 logit. The shape of curve A indicates the presence of resistant preconception type-three in level 1.

It depicts that the particular item response pattern that signifies students' conceptual understanding patterns in the given level. Moreover, the curve shape of distractors A and B in the items Q2/PM/SL and Q8/PM/SL tend to have an identical pattern. The finding indicated that students with either low or high ability had consistent preconceptions that the mass of the particle can change into larger or smaller in size along with the change in matter form.

Third sample, i.e., item Q5/Dew (-0.63), as shown in Figure 7(a), measures the students' ability in elaborating characteristics of a particle in condensation phenomenon. The option probability curve is displayed in Figure 7(b). Students with low ability (< 1.0 logit) tended to pick distractor A (water drops come from liquid of melting ice that breaks through the glass wall) and option D (other answers). Some students with high ability (> 1.0 logit) also picked distractor B (water drops are the result of the reaction between ice and air nearby the glass). The shape of curve B is wavy and non-linear, even in the interval of 2.0 to 4.0 logit, it can reach option probability value up to 1.0 logit. This is regarded as a deviation from the right answer C (water drops come from condensing water vapor nearby the glass). A worth note, however, is to consider in the unstable, wavy shape of curve C. This indicated the students' inconsistency (particularly those with high ability) in comprehending the concept of condensation. This confirmed that students had their own preconception regarding concept of condensation.

Figure 7

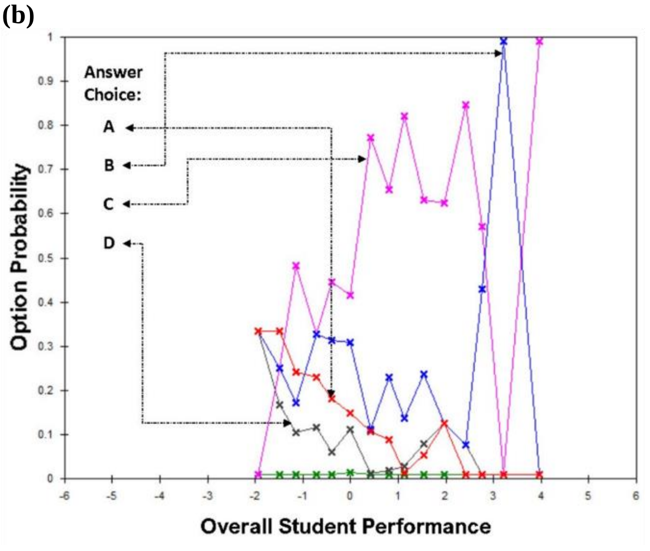
(a) item Q5/Dew; (b) option probability curve

(a)

In a glass filled with ice, you can see water drops at the glass. According to you, where do the water drops come from?

A. Water drops come from melting ice that penetrates the glass wall
B. water drops are the result of reaction between ice and air nearby the glass
C. Water drops come from condensing water vapor nearby the glass
D. Other answer...

(b)



The fourth sample or item Q6/Bubble, as shown in Figure 8(a), measures the students' ability in elaborating characteristics of a particle in the evaporation phenomenon. The option probability curve is displayed in Figure 8(b). The distractor A (air bubbles are Hydrogen and Oxygen particles) was dominantly chosen by students whose ability in a range between -3.0 to 2.0 logit. Moreover, the distractor B (air bubbles are Hydrogen and Oxygen particles) was mostly selected by students whose ability in a range between -3.0 to 0.5 logit. The form of curve A and B were picked by students with low ability was predictable. The curve of right answer C (air bubbles are water molecules), however, shows interesting hint; in the interval range of -2.5 to 3.0 logit, the tip of the curve shows an up-and-down pattern. Moreover, in the level of 1.5 logit, the curve shape of distractors A and B shows a decline pattern, while that of curve C tends to increase. Another finding worth noting was that the curve D (other answers) was picked by some students with high ability (> 2.0 logit). This indicated that particular students had their own preconceptions regarding the evaporation concept.

Figure 8

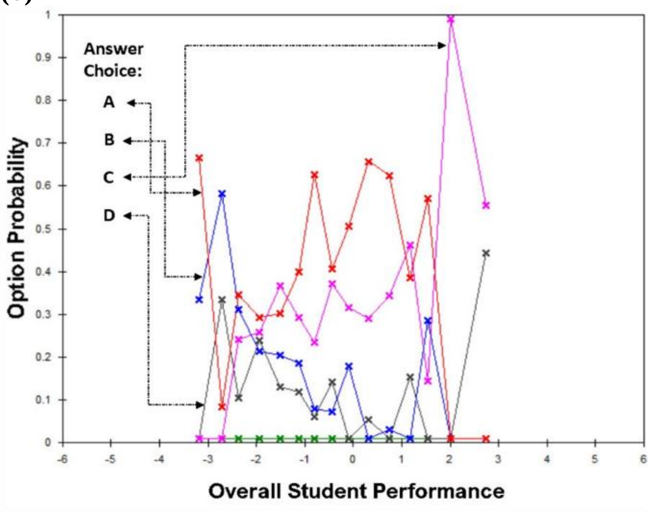
(a) item Q6/Bubble; (b) option probability curve

(a)

In a container filled with boiling water, you can see air bubbles on the top of it. According to you, what are the composing elements of the air bubbles?

A. Hydrogen and Oxygen particles
B. Air that is dissolved in water
C. Water molecules
D. Other answer...

(b)



Discussion and Implications

The research findings indicated that the instruments had good effectiveness, met the requisites of person and item reliability, and showed good construct validity. When applied in evaluating students' conceptual understanding, it was found that: Firstly, almost all students with high ability faced difficulty in understanding the concept of particle size and mass in level 1. The same students found it relatively easy in determining SMRs diagram of particle structure in level 2 or determining the concept of particle regarding evaporation and condensation phenomena in level 3. Secondly, the information of the response pattern of students with high ability was quite consistent, repetitive, and systematic in particular items. This indicates the presence of permanent and latent preconceptions. The analysis of the option probability curve of item Q2/PM/SL (0.88), Q8/PM/LG (0.65), Q5/Dew (-0.63) and Q6/Bubble (0.60) indicates that the approach of item response pattern is able to explore in detail and comprehensively regarding students' conceptual understanding and preconception.

Sequences of verification conducted that involves Rasch model approach shows detailed, accurate, and quantifiable results since the approach integrates development procedure of diagnostic and summative instruments. Several samples of preconception, e.g., item Q2/PM/SL (0.88) and Q8/PM/LG (0.65) indicate that distractor options are potential to be elaborated further in order to investigate tendency of preconception by the students. In addition, it also provides information regarding main idea unknown to the students and their degree of misunderstanding.

The approach employed in this research is an effective illustration to help teacher in evaluating the learning process as well as the students' learning progress. This is due to the integration of qualitative item development procedure and quantitative data analysis, allowing the teachers to explore in-depth on the students' understanding, concepts the students understand and/or do not **understand, and misconception.** Such findings echo Herrmann-Abell & Deboer (2016) that the integration of Rasch model analysis and probability curve is applicable to diagnose how the students' misconception turns into their overall conceptual understanding. Such an attempt is quite hard to conduct by implementing a conventional approach due to the interdependence of person and item. Rasch model, on the other hand, is able to tackle such interdependence, in which the item and the test difficulty remain invariant and not dependent on which sample that is involved in the initial validation. This signifies that the instrument's items have met the unidimensionality and local independence requirements (Jin et al., 2019; Testa et al., 2019; Wei et al., 2012).

Overall, the research indicated empirical evidence that supported findings by Hoe & Subramaniam (2016); Lu & Bi (2016); and Rogat et al., (2011), that students had distinctive preconception as a result of a learning process they experienced. Such preconception was regarded as the inhibitor to the development process of students' conceptual understanding (Soeharto et al., 2019). In this research, students' preconception was found to be repetitive and systematic in each education level. It signifies that the intervention to change students' preconceptions was difficult to conduct by the conventional learning method. A strategic and meaningful learning method is therefore essential to remove students' incorrect preconceptions and develop scientifically correct conceptual understanding. That being said, teachers are demanded to acquire detailed information on the forms and characteristics of students' preconceptions. In conclusion, the item response pattern analysis was an efficient and effective means to acquire such information. The information on students' preconception is important as the basis to develop appropriate and measurable instructional design in solving the students' misconception. This is in line with the previous researches, arguing that the quality of learning progress is highly dependent on the students' learning process and learning experience (Duschl et al., 2011; Park et al, 2017; Wilson, 2009).

Conclusions

The measuring instrument developed performed well in its validity and reliability, thus, it is deemed applicable in measuring students' conceptual understanding and preconception in elaborating particle characteristics of matter. During the implementation of the instruments, the research finds out that: 1) almost all students with high ability face difficulty in understanding the concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, as well as determining the concept of particle regarding evaporation and condensation phenomena in level 3.

2) There is a significant disparity between students' conceptual understanding based on their educational level. 3) In certain cases, it is found that the distractor item response pattern by high-ability students tends to be consistent, indicating a certain tendency of resistant preconception pattern.

The development of diagnostic instruments with Rasch model approach is deemed as the literacy process for practitioners and researchers in Indonesia. The result indicates that there is no single item that is parallel with both the highest ability and lowest ability students. This calls for further elaboration in order to improve the instrument items' quality. Moreover, an anomaly is found that students with high ability (> 1.0 logit) tend to pick distractor choices. This urges further studies to investigate structured comprehension problems. The research regards that further analysis that integrates conceptual understanding level and items designed in a gradual manner is required to define the characteristics of the students' alternative conception and to measure their learning progress. Echoing this notion, one must integrate the item design and basic principles of chemistry as a reference for further researchers and educational practitioners to implement the same approach conducted in the present research. On top of that, despite not focused on discussing matters regarding students' learning progress individually, the instrument is expected to be beneficial for the teachers to diagnose students' conception in developing an effective and meaningful learning experience.

Acknowledgments

This research is funded by the Directorate of Research and Community Outreach/ (Direktorat Riset dan Pengabdian Masyarakat or DRPM), Directorate General of Research and Development (Dirjen Riset dan Pengembangan) of Ministry of Research and Technology of the Republic of Indonesia through Penugasan Penelitian Dasar Unggulan Perguruan Tinggi Grant Program, Research and Community Outreach Institute (Lembaga Penelitian dan Pengabdian Masyarakat) of Universitas Negeri Gorontalo in 2020.

References

Adadan, E., & Savasci, F. (2012). An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, *34*(4), 513–544. https://doi.org/10.1080/09500693.2011.636084

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

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

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

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

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

Blanc, A., & Rojas, A. J. (2018). Use of rasch person-item maps to validate a theoretical model for measuring attitudes toward sexual behaviors. *PLoS ONE*, *13*(8), 1–17. https://doi.org/10.1371/journal.pone.0202551

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

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

Boz, Y. (2006). Turkish pupils' conceptions of the particulate nature of matter. *Journal of Science*

Education and Technology, *15*(2), 203–213. https://doi.org/10.1007/s10956-006-9003-9

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The Development of Two Tier Multiple-Choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Levels of Representation. *Chemistry Education Research and Practice*, *8*(3), 293–307.

Cheng, M. M. W. (2018). Students' visualisation of chemical reactions-insights into the particle model and the atomic model. *Chemistry Education Research and Practice*, *19*(1), 227–239. https://doi.org/10.1039/c6rp00235h

Davidowitz, B., Chittleborough, G., & Murray, E. (2010). Student-generated submicro diagrams: A useful tool for teaching and learning chemical equations and stoichiometry. *Chemistry Education Research and Practice*, *11*(3), 154–164. https://doi.org/10.1039/c005464j

Devetak, I., & Glažar, S. A. (2010). The influence of 16-year-old students' gender, mental abilities, and motivation on their reading and drawing submicrorepresentations achievements. *International Journal of Science Education*, *32*(12), 1561–1593. https://doi.org/10.1080/09500690903150609

Devetak, I., Urbancic, M., Grm, K. S. W., Krnel, D., & Glazar, S. A. (2004). Submicroscopic representations as a tool for evaluating students' chemical misconception. *Acta Chimica Slovenica*, *51*(December 2004), 799–814. https://www.researchgate.net/publication/286849527_Submicroscopic_representations_as_a_tool_for_evaluating_students'_chemical_conceptions

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, *47*(2), 123–182. https://doi.org/10.1080/03057267.2011.604476

Femintasari, V., Effendy, & Munzil. (2015). The effectiveness of two-tier multiple choice test and multiple choice test followed with interview in identifying misconception of students with different scientific reasoning skills. *Jurnal Ilmu Pendidikan*, *21*(2), 192–197. http://journal.um.ac.id/index.php/jip/article/view/8376/4020

Gabel, D. L. (1993). Use of the particle nature of matter in developing conceptual understanding. *Journal of Chemical Education*, *70*(3), 193. https://doi.org/10.1021/ed070p193

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

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

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

Herrmann-Abell, C. F., & Deboer, G. E. (2016). Using Rasch modeling and option probability curves to diagnose students' misconceptions. *American Eduacational Research Assossiation*, 1–12.

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

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

Jin, H., Mikeska, J. N., Hokayem, H., & Mavronikolas, E. (2019). Toward coherence in curriculum, instruction, and assessment: A review of learning progression literature. *Science Education*, *103*(5), 1206–1234. https://doi.org/10.1002/sce.21525

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

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

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

Kapici, H. Ö., & Akcay, H. (2016). Particulate nature of matter misconceptions held by middle and high school students in Turkey. *European Journal of Education Studies*, *2*(8), 43–58. https://doi.org/10.5281/zenodo.163547

Kim, J., & Wilson, M. (2019). Polytomous item explanatory item response theory models. *Educational and Psychological Measurement*. https://doi.org/10.1177/0013164419892667

Kind, V. (2004). *Beyond appearances: Students' misconceptions about basic chemical ideas* (2nd Edition). Royal Society of Chemistry. https://www.researchgate.net/publication/228799159_Beyond_Appearances_Students'_Misconceptions_About_Basic_Chemical_Ideas

Laliyo, L. A. R., Botutihe, D. N., & Panigoro, C. (2019). The development of two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12

Linacre, J. M. (2012). *A user's guide to W I N S T E P S ® M I N I S T E P Rasch-model computer program: Program manual 3.75.0*. https://doi.org/ISBN 0-941938-03-4

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

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

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, *34*(15), 2317–2341. https://doi.org/10.1080/09500693.2012.684433

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

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*(9), 825–838. https://doi.org/10.1002/tea.3660200905

Ozalp, D., & Kahvecib, A. (2015). Diagnostic assessment of student misconceptions about the particulate nature of matter from ontological perspective. *Chemical Education Research and Practice*, *16*, 619–639. https://doi.org/10.1039/C5RP00096C

Ozmen, H. (2011). Turkish primary students´ conceptions about the particulate nature of matter. *International Journal of Environment & Science Education*, *6*(1), 99–121. https://www.researchgate.net/publication/265438996_Turkish_primary_students'_conceptions_about_the_particulate_nature_of_matter

Park, M., Liu, X., & Waight, N. (2017). Development of the connected chemistry as formative assessment pedagogy for high school chemistry teaching. *Journal of Chemical Education*, *94*(3), 273–281. https://doi.org/10.1021/acs.jchemed.6b00299

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

Renström, L., Andersson, B., & Marton, F. (1990). Students' conceptions of matter. *Journal of Educational Psychology*, *82*(3), 555–569. https://doi.org/10.1037/0022-0663.82.3.555

Rogat, A., Anderson, C., Foster, J., Goldberg, F., Hicks, J., Kanter, D., Krajcik, J., Lehrer, R., Reiser, B., & Wiser, M. (2011). *Developing learning progression in support of the new science standards: A RAPID workshop series, New York, NY.* (4), 163. https://doi.org/10.12698/cpre.2011.lprapid

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

Sadler, P. M. (1999). The relevance of multiple-choice testing in assessing science understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 251–274). https://zodml.org/sites/default/files/%5BJoel_J._Mintzes%2C_James_H._Wandersee%2C_Joseph_D._No_0.pdf

Slapničar, M., Devetak, I., Glažar, A. S., & Pavlin, J. (2017). Identification of the understanding of the states of matter of water and air among Slovenian students aged 12, 14 and 16 years through solving authentic tasks. *Journal of Baltic Science Education*, *16*(3), 308–323.

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

Stojanovska, M. I., Soptrajanov, B. T., & Petrusevski, V. M. (2012). Addressing misconceptions about the particulate nature of matter among secondary-school and high-school students in the Republic of Macedonia. *Creative Education*, *03*(05), 619–631. https://doi.org/10.4236/ce.2012.35091

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

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*

[Application of Rasch Model in Social Sciences Research]. Tri Komunikata. https://www.researchgate.net/publication/268688933%0AAplikasi

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan [Application of Rasch Model in Education Assessment]*. Tim Komunikata. https://www.researchgate.net/publication/282673464%0AAplikasi

Taber, K. S. (2015). Alternative Conceptions/Frameworks/Misconceptions. In G. Richard (Ed.), *Encyclopedia of Science Education* (pp. 37–41). https://doi.org/10.1007/978-94-007-6165-0

Testa, I., Capasso, G., Colantonio, A., Galano, S., Marzoli, I., Scotti di Uccio, U., Trani., F & Zappia, A. (2019). Development and validation of a university students' progression in learning quantum mechanics through exploratory factor analysis and Rasch analysis. *International Journal of Science Education*, *41*(3), 388–417. https://doi.org/10.1080/09500693.2018.1556414

Tóth, Z., & Kiss, E. (2006). Using particulate drawings to study 13-17 year olds' understanding of physical and chemical composition of matter as well as the state of matter. *Practice and Theory in Systems of Education*, *1*(1), 109–125. http://eduscience.fw.hu/

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169. https://doi.org/10.1080/0950069880100204

Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, *4*(6), 626–631. http://www.academicjournals.org/SRE

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335–345. https://doi.org/10.1021/ed100852t

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

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

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

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

Yaşar, I. Z., İnce, E., & Kırbaşlar, F. G. (2014). 7. Class Science and Technology Course "Structure of Atom" Subject Readiness Improvement Test. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2014.09.259

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

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

200419.jbse_JL-suo_review
Re-evaluation

Abstract. *This research aimed to evaluate the students' conceptual understanding and to diagnose the students' preconceptions in elaborating the particle characteristics of matter by development of diagnostic instrument as well as Rasch model response pattern analysis approach. Data were acquired by 25 multiple-choice written test items distributed to 987 students in North Sulawesi, Indonesia. Analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding; and 3) diagnosis of students' preconception by estimation of item response pattern. The result generated information on the diagnostic and summative measurement on students' conceptual understanding in elaborating the topic; information also acts as empirical evidence on the measurement's reliability and validity. Moreover, the result discovered a significant disparity between students' conceptual understanding based on their educational level. It was found that the distractor item response pattern tended to be consistent, indicating a certain tendency of resistant preconception pattern. The findings are expected to be a recommendation for future researchers and educational practitioners that integrates diagnostic and summative measurement with Rasch model in evaluating conceptual understanding and diagnosing misconception.*
Keywords: *conceptual understanding, item response, particle, Rasch model.*

Commented [Reviewer1]: Note that a keyword does not have to be made of only one word!

Introduction

Central to the notion of learning about characteristics of a particle of matter is the process of developing an understanding on abstract concepts (Johnstone, 1991) without directly interacting with the object/fact (Stojanovska et al., 2012); therefore it is considered a difficult subject for the students to learn. Echoing this, the disparity in understanding is almost inevitable (Kapici & Akcay, 2016) since different students may develop ones' own distinctive way of understanding a concept (Yildirir & Demirkol, 2018). The idea is also coined by experts as misconception (Johnstone, 2006, 2010; Taber, 2015), or alternative framework and preconception (Lu & Bi, 2016). The experts have discovered that students always have their own preconception that is not in line with scientific concepts (Alamina & Etokeren, 2018; Yaşar et al., 2014); therefore, one needs to conduct identification and improvement on the conceptual learning (Allen, 2014; Soeharto et al., 2019).

In diagnosing preconceptions, several researchers have developed diagnostic instruments in different mechanisms (McClary & Bretz, 2012), i.e., conceptual map, essay test, interview, essay test with interview, or multiple-choice test (Femintasari et al., 2015). Two-step multiple choice diagnostic test (Adadan & Savasci, 2012; Chandrasegaran et al., 2007; Treagust, 1988; Tüysüz, 2009) is preferred due to its ability to diagnose preconception and describe the underlying reasons. The instrument is indeed considered qualitatively effective in elaborating differences in students' thought processes; however, it does not provide summative measurement features due to lack of internal consistency and the instrument's unidimensionality (Lu & Bi, 2016). In addition to that, the measurement conclusion generated is considered weak due to extracted from analysis on the raw score (Sumintono, 2018)

Studies on preconception have found that the concept is somewhat resistant. In the early 2000s, it is discovered that students' preconceptions persisted even when they already undergo formal education experience (Hoe & Subramaniam, 2016). Preconception can also change along with the development of students' conceptual understanding; it also varies in different levels of understanding (Aktan, 2013). If one conducts a two-step test and raw score analysis approach to diagnose resistant preconception, the result generated will only provide limited feedback information (Sumintono, 2018) due to the instrument's limitation in measuring students' conceptual understanding. Instead of supporting, the information will only make it harder for teachers to implement proper instructional decisions (Wilson, 2008).

During the middle of the 2000s, the Rasch model analysis was commonly used in studies of chemistry education (Herrmann-Abell & DeBoer, 2011; Liu, 2012; Wein et al., 2012). The approach provides a testing apparatus that integrates diagnostic and summative measurement. Recently, this approach is used to develop formative assessment with the intention to conduct learning construction mapping, e.g., measuring the students' way of constructing their understanding process (Hadenfeldt et al., 2013). It is worth to note, however, that there are studies that integrate diagnostic and summative measurement with a different approach (Hoe & Subramaniam, 2016); despite that, trends in chemistry education studies highlight that diagnostic-summative measurement by Rasch model analysis is more common to be carried out (Laliyo et al., 2019; Lu & Bi, 2016).

Research Problem

The characteristics of a particle of matter is a fundamental concept in chemistry, usually taught in middle education level. Adequate comprehension regarding the particle characteristics of matter both in macroscopic and microscopic level is essential as the knowledge basis in understanding more advanced topics such as the concept of atoms and molecules as the submicroscopic component that is invisible to plain eyesight but exists in all real-world phenomena (Cheng, 2018; Ozmen, 2011; Yildirir & Demirkol, 2018). The fact signifies the relevance and reasoning of complexity in chemistry learning that is considered difficult for both students and teachers to conduct (Alamina & Etokeren, 2018). In simpler terms, to ensure that the chemistry learning is conducted effectively, one requires to nurture students' comprehensive understanding regarding particle characteristics of matter and its change of state.

To evaluate the students' conceptual understanding on the aforementioned topic, one also needs to measure the students' capability in interpreting particle state during change process of a matter's form (Alamina & Etokeren, 2018; Barbera, 2013; Boz, 2006; Cheng, 2018; Gabel, 1993; Hadenfeldt et al., 2013; Kapici & Akcay, 2016; Kind, 2004; Naah & Sanger, 2012; Ozalp & Kahvecib, 2015; Ozmen, 2011; Renström et al., 1990; Slapničar et al., 2017; Stojanovska et al., 2012; Yildirir & Demirkol, 2018). Researches on particle characteristics and changes of matter generally employ diagnostic instruments in the form of essay tests and/or essays followed by interview; the instruments are further analyzed based on raw score results. The approach is considered inefficient and somewhat lacked accuracy in measuring students' conceptual understanding and misconception pattern. Despite its ineffectiveness, the conventional method is used by most teachers in Indonesia to measure and determine students' learning progress. The teachers argue that measuring the students' raw score is effective in determining how far the students have progressed in the learning process. The students' raw score is regarded by many as an early premature indication regarding the measured variable and is not eligible to be the final measurement indicator due to its temporary nature. In addition to that, regarding the decision-making process, the raw score contains only limited information for it to be treated as reference (He et al., 2016; Sumintono & Widhiarso, 2015)

Research Focus

The research focuses on developing a diagnostic instrument that integrates measurement of conceptual understanding and diagnosis of students' preconceptions regarding the aforementioned topic by the approach of Rasch model item response pattern analysis. The analysis employs different test apparatuses to provide extensive information for practitioners and researchers in science education in evaluating students' learning progress in different topics.

Research Questions

This research aimed to figure out the following questions: 1) How is the effectiveness of measurement instrument to evaluate the students' conceptual understanding and diagnose their preconceptions on the characteristics of a particle of matter? 2) Is there any significant difference between students in elaborating on the aforementioned topic based on their educational level? 3) How is the pattern of students' conceptual understanding and preconception regarding the topic?

Research Methodology

General Background

The descriptive-quantitative research employed a non-experimental approach, in which the students' conceptual understanding in explaining the characteristics of a particle of matter was treated as the measurable variable. Prior to conducting the research, it was ensured that the students already experience formal learning of the aforementioned topic. The researchers did not conduct any intervention on the learning process or the learning material. In other words, no treatment was implemented to the students for them to be able to answer all test items in the measurement instrument.

The data collection step was implemented for four months in the even semester of the 2019-2020 academic year; the process was conducted after obtaining approval from the Government of Province of Gorontalo and heads of universities in the Northern part of Sulawesi, Indonesia. Moreover, the schools' and parents' approval was obtained in cooperation with the school committee. The school administrators were willing to facilitate the data collection process that adjusted with the schedule.

Respondent

The respondents were 987 people consisting of students of eleventh grade from eight junior high schools as well as university students of the chemistry department in Northern Sulawesi, Indonesia. The distribution of respondents is displayed in Table 1 below.

Commented [Reviewer2]: Lower-secondary school

Table 1
Demographic profile of respondents (N=947)

Demography	Code	Respondents	Percentage (%)
Gender			
Male	M	320	67.68
Female	F	667	32.42
Education level			
X Class students	M	168	17.02
XI Class students	N	473	47.92
XII Class students	O	186	18.84
University students from the chemistry department	P	160	16.21

The respondents were chosen randomly and have voluntarily agreed to participate in the research. In addition, they received no learning treatment and other special treatments that allow them to complete the measurement instrument. Students were asked to write down their responses in the answer sheet; the process was supervised by teachers in the respective schools and lecturers in the respective university. All students were instructed to answer all questions in the instruments within 45 minutes. All instrument sheets and answer sheets were collected by the researchers shortly after the session ended; it was ensured that the numbers of instruments matched the numbers of participants.

Commented [Reviewer3]: Ethical aspects?

Instrument and Procedures Development

The design process refers to a recommendation by Wilson (2005), which consists of four key steps: definition of construct map, item design, result blank, and measurement model.

Phase 1: Definition of construct map. The map offers a substantive definition of measured constructs; the more constructs measured, the constructs' level will vary qualitatively (Wilson, 2009). In simpler words, it aims to develop the students' understanding map to measure the students' progress (Wilson, 2012). The instrument involved variables, i.e., the students conceptual understanding and preconception in elaborating the characteristics of a particle of matter; it was conducted in accordance with the Curriculum Standard of Chemistry Subject in Tenth Grade in Indonesia, as presented in Table 2.

Table 2

Conceptual Understanding Level

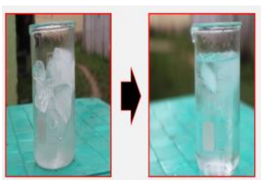
Level 3 The students are able to connect between characteristics of a particle of matter in macroscopic and submicroscopic level	
Phenomenon	Evaporation: item Q6/Bubble
10. Preconception	Air bubble consists of Hydrogen and Oxygen particles
9. Preconception	Air bubble is water-soluble
Phenomenon	Condensation: item Q5/Dew
8. Preconception	Water drops come from melting ice that penetrates the glass wall
7. Preconception	Water drops are the result of the reaction between ice and air nearby the glass
Level 2 The students are able to determine SMRs diagram of particle structure during a change of form: item Q11/SMRs/SL; Q12/SMRs/LG; Q13/SMRs/GS; Q25/SMRs/GG	
6. Preconception	The SMRs diagram of particle structure follows the physical form of matter
5. Preconception	The SMRs diagram of O_2 molecule shape undergoes change as a result of an increase in the volume of the container.
Level 1 The students are able to determine the characteristics of a particle of matter during the change process of matter's form.	
4. Preconception	The particle size of matter changes into (large/small) as a result of change in matter form: item Q1/PS/SL; Q7/PS/LG; Q14/PS/LG; Q18/PS/SG; Q22/PS/GG
3. Preconception	The particle mass of matter changes into (large/small) due to change in matter form: item Q2/PM/SL; Q8/PM/LG; Q15/PM/LG; Q19/PM/SG; Q24/PM/GG
2. Preconception	Distance between matter particles changes into (faster/slower) due to change in matter form: item Q3/DP/SL; Q9/DP/SL; Q16/DP/LG; Q20/DP/SG; Q23/PM/GG
1. Preconception	Motion between matter particles changes into (dense/loose) due to change in matter form: item item Q4/PMo/SL; Q10/PMo/LG; Q17/PMo/LG; Q21/PMo/SG

Variation in conceptual understanding level illustrates the development process of the students' conceptual understanding. In the first level, the students were asked to determine particle characteristics (size, mass, motion, and distance) in the change process of matter form. In the second level, the students were asked to determine the submicroscopic representation diagram of particle structure. Further, in the third level, the students were asked to connect between characteristics of a particle of matter at the macroscopic and submicroscopic level. In each level, the construct map also features the students' tendency of preconception.

Phase 2: item design and evaluation The phase involved the determination process of items to be used in acquiring evidence of students' construct understanding regarding the construct map (Wilson, 2005). Certain items may have a different extent of effectiveness to measure students' conceptual understanding (Sadler, 1999); however, multiple choices item is considered more practical and effective (Wilson, 2008). The instrument of concept understanding test of the particle (or TPKP) is adapted from multiple-choice instruments by (Herrmann-Abell & DeBoer, 2011). Each item consists of two distractor answer choices and one open answer choice. The distractor answer choices are designed by referring to the common preconceptions by the students (see Table 2) as logical choices to distract the students from the correct one. The distractors function to emphasize the item diagnostic strength (Sadler, 1998). Some of the items are adopted from previous studies Osborne & Cosgrove (1983), Renström et al., (1990); Devetak et al., (2004); Tóth & Kiss (2006); Davidowitz et al., (2010); Devetak & Glažar (2010); Slapničar et al., (2017) and (Yildirir & Demirkol, 2018).

Figure 1
Sample of item Q1/PS/SL design

(a) (b)



Glass (a) contains ice chunks, glass (b) contains melting ice chunks. How is the size of water particle in solid form (ice) compared to that in liquid form?

a. Size of a water particle in solid form > a water particle in liquid form.
b. Size of a water particle in solid form < a water particle in liquid form.
c. Size of a water particle in solid form = a water particle in liquid form.
d. Other answers

Figure 1 displays a sample of item Q1/PS/SL design, in which Q1 is the number of item 1, PS is particle size, and SL is solid-liquid. The item measures student's capability in determining particle size in form change from solid to liquid. The choice A and B are distractors, the correct choice is C, and choice D is for other answers students may fill if the existing answer choices are not in accordance with their initial knowledge. Every correct answer was given mark 1, and wrong answers got 0 mark. Each student only has a slight probability of 0.25 in choosing the right answer. The students will pick what they think the right answer based on their understanding. If the distractor item choice functions well, the students will not be able to predict the correct answer.

Phase 3: design of result blank, i.e., the correlation between construct map and items (Wilson, 2005). This phase aimed to identify whether the answer the students pick correlates with their conceptual understanding; in simpler terms, it was intended to elaborate the conformity between the variable contents being measured. In order to elaborate on the previous aspect, the TPKP instrument was validated by three independent experts and tested to the students to acquire their feedback. The process acquired 25 items of TPKP. Prior to the data collection process, it was ensured that all students had received formal education on the characteristics of a particle of matter and its changes. The students' response towards the instrument was inputted manually by the written answer sheet. The test was supervised by the teachers in school by referring to the agreed permission and duration. Each student was required to finish all test items within the allocated duration of 45 minutes. The instrument sheets were further collected, and checking process was conducted to ensure that the amount of instrument sheet wass the same with participating students.

Phase 4: Rasch model analysis approach. The analysis integrates algorithm as a result of probabilistic expectation of item 'i' and student 'n', as: The statement is the probability of student n in item i to result in the correct answer (x = 1); with student ability, ßn, and item difficulty level (Bond & Fox, 2015). The above equation was simplified by inserting logarithm function, into , so that the probability of picking the right answer equals to student's ability subtracted by item difficulty level. The student (person) and item units were considered on the same interval scale and were independent of each other. The students' ability level and item difficulty level were measured in the logarithm unit, namely odds or log that variates from -00 to +00 (Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015). The instrument efficiency, when compared to the item distribution towards item difficult level with distribution of student's ability level, was quantifiable in order to measure the students' conceptual understanding. In addition, the student's understanding level was differentiated based on the item size. The previous steps highlighted the main difference of Rasch model analysis when compared to the raw score-based conventional one; the latter lacks accuracy in evaluating students' ability observed from different item difficulty level (Herrmann-Abell & DeBoer, 2011; Lu & Bi, 2016; Sumintono & Widhiarso, 2015).

Data Analysis

The research employed WINSTEPS version 3.75 software to convert raw data into interval data (Bond & Fox, 2015; Linacre, 2012). The conversion result acted as the calibration of data on the student's ability level and item difficulty level within the same interval measurement. Moreover, the analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding by Differential Item Functioning (DIF) item test; and 3) diagnosis of students' preconception by estimation of item response pattern through option probability curve test.

Research Results

Effectiveness of Measuring Instruments

Person and Item Reliability. The first step to elaborate on the effectiveness of measuring instruments was by measuring the person and item reliability. This was conducted to gather information to what extent the measurement produces consistent information in displaying latent trait or the unidimensionality of the measured variable (Sumintono & Widhiarso, 2015). The analysis result is presented in the form of a statistical summary (Table 3).

Table 3
Summary of fit statistics

Parameter (N)	**Measure**	**INFIT**		**OUTFIT**		**Separation**	**Reliability**	***SD***	**KR-20**
		MNSQ	ZSTD	MNSQ	ZSTD				
?									
Person (987)	-.34	1.00	-.11	1.02	-.1	1.55	.71	.88	.72
Items (25)	.00	1.00	-.75	1.02	-.1	8.18	.99	.60	

Commented [Reviewer4]: Remove an empty row

The above table indicates that the person reliability value of 0.71 is equivalent to the person separation index value of 1.55. This is to say that the consistency of students' response towards the test is deemed good. In addition to that, it is generated that the *Cronbach Alpha Coefficient* (KR-20) value is 0.72, signifying good interaction between students and the test. This further indicates strong correlation between the students' response towards the item, in the context that the students' knowledge tends to be non-fragmented, enabling it to be measured (Adams & Wieman, 2011). To the researchers and educational practitioners, such information is essential to prepare for follow-up plans and development of students' ability (Wei et al., 2012). Moreover, the result generated a relatively high value of item separation index of 8.18 that was equivalent to the item reliability value of 0.99. This indicated very good item consistency or the item was deemed capable of meeting the unidimensionality criteria. In other words, the item performed very good in defining the measured variable. This was confirmed by the infit and outfit value result, in which most of the items were in the acceptable range for the multiple-choice test (Herrmann-Abell & DeBoer, 2011; Bond & Fox, 2015).

Commented [Reviewer5]: The APA Manual (6/7th ed.) says: "Order the citations of two or more works within the same parentheses alphabetically" (6.16 on page 177).

Figure 2
Function of Measurement Information

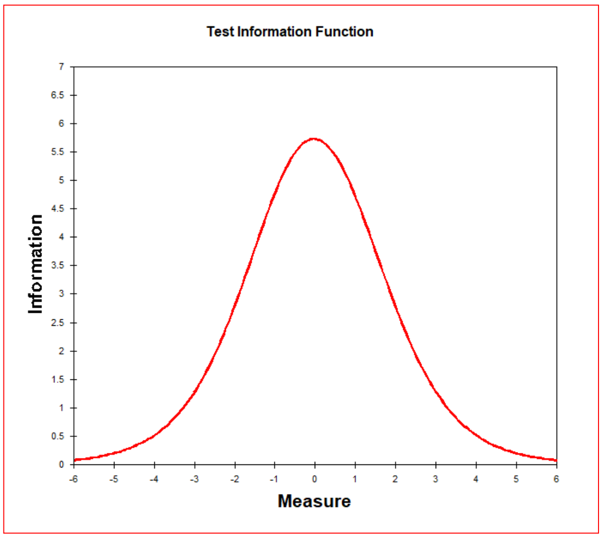


Figure 2 displays the graph of measurement information in order to show the measurement reliability. The higher the tip of information function graph, the measurement reliability value is likely to increase. In the intermediate level of students' ability (-3.0 logit up to +3.0 logit), the measurement information is in very high spot. This indicates that the TPKP instrument is capable of producing optimal information to students with an intermediate level of ability. Such a result means that the instrument possesses high measurement reliability (Bond & Fox, 2015; Kim & Wilson, 2019).

Validity. The next step was to measure the item validity by Fit item test to ensure that all items fit with the Rasch model. The process was aimed to identify whether or not the test item could measure the aspects that intended to be measured, or test validity (Linacre, 2012; Sumintono & Widhiarso, 2014). The criteria used comprise outfit means-square (MNSQ): 0.5 <

y < 1.5; outfit z-standard: -2.0 < Z < + 2.0, as well as point measure correlation (PTMEA Corr). The PTMEA Corr is the correlation between the score of item and person measure that is required to be a positive value and not approaching zero (Bond & Fox, 2015). The PTMEA Corr criteria: 0.4 < x < 0.8. If all three criteria are not met, the item is not good enough and needs further elaboration (Boone et al., 2014). Both Outfiit MNSQ and Infit MNSQ were sensitive chi-squares in detecting outlier response pattern. There were two outlier responses: the right response, guessed by the students with low ability in item with high difficulty level, or the wrong response due to the high-ability students' carelessness in items with a low difficulty level. The expected ideal MNSQ value is 1.0. The analysis result on item appropriateness is displayed in Table 4 as follows:

Table 4

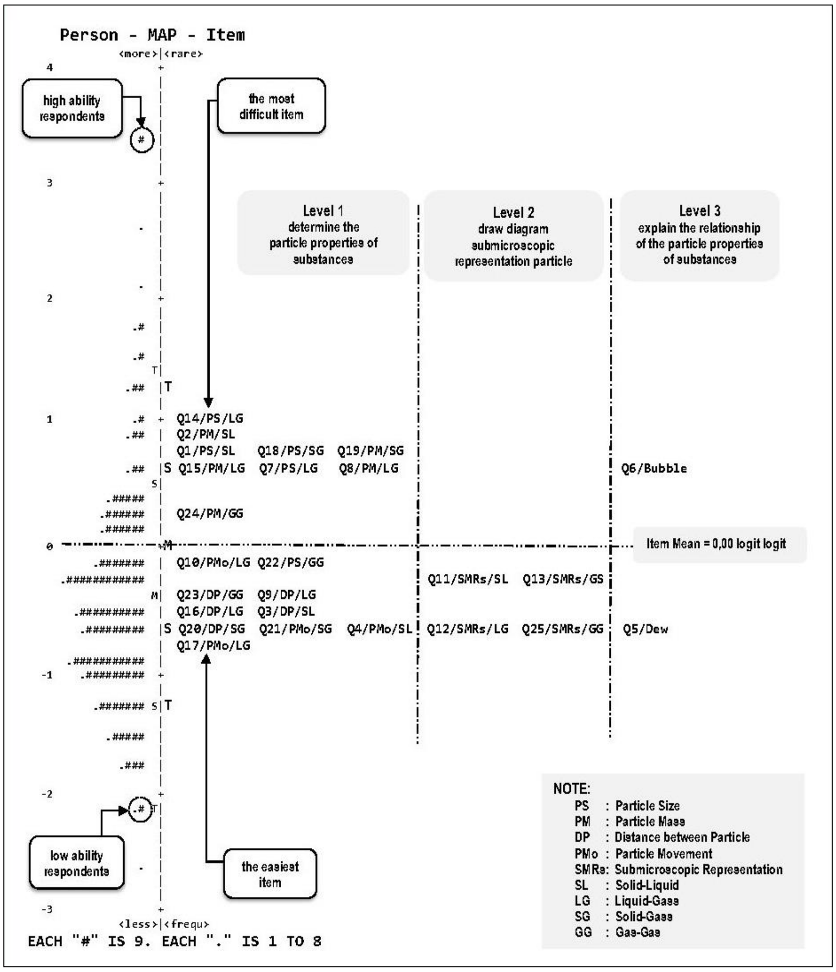
Item Statistics: Misfit Order

Item	Measure	INFIT		OUTFIT		PTMEA Corr
		MNSQ	ZSTD	MNSQ	ZSTD	
Q6/Bubble	.60	1.26	7.0	1.40	7.5	.07
Q2/PM/SL	.88	1.16	3.7	1.27	4.4	.18
Q15/PM/LG	.66	1.12	3.3	1.20	3.8	.22
Q14/PS/LG	.97	1.03	.8	1.18	2.8	.20
Q18/PS/SG	.71	1.07	1.8	1.15	2.9	.28
Q5/Dew	-.63	1.06	2.6	1.15	3.7	.27
Q7/PS/LG	.66	1.06	1.6	1.14	2.8	.29
Q8/PM/LG	.65	1.07	2.0	1.11	2.1	.28
Q1/PS/SL	.79	1.00	-.1	1.06	1.1	.35
Q24/PM/GG	.25	1.04	1.4	1.04	1.1	.33
Q19/PM/SG	.77	-.3	1.03	.6	.36	.36
Q3/DP/SL	-.44	1.01	.5	1.00	-.1	.34
Q10/PMo/LG	-.07	.98	-.8	.98	-.5	.38
Q13/SMRs/GS	-.24	.98	-1.1	.98	-.6	.38
Q9/DP/LG	-.32	.97	-1.6	.95	-1.6	.39
Q4/PMo/SL	-.66	.96	-2.0	.93	-1.8	.39
Q25/SMRs/GG	-.68	.94	-2.9	.91	-2.4	.41
Q16/DP/LG	-.47	.94	-3.1	.91	-2.8	.42
Q23/DP/GG	-.44	.92	-3.7	.93	-2.1	.43
Q12/SMRs/LG	-.63	.92	-3.8	.87	-3.5	.44
Q21/PMo/SG	-.66	.92	-4.0	.89	-2.9	.43
Q17/PMo/LG	-.71	.91	-4.4	.87	-3.5	.44
Q22/PS/GG	-.07	.90	-4.4	.87	-3.9	.47
Q11/SMRs/SL	-.27	.90	-4.9	.87	-4.0	.47
Q20/DP/SG	-.65	.86	-6.6	.83	-4.6	.49

From the previous Item Statistics, it is generated that all items meet the Outfit MNSA criteria and no negative PTMEA Corr occurs. This means that all items are not deviant, appropriate, and valid. Despite some items do not meet one of the criteria, this by no means decreases the quality of the items. For instance, item (Q6/Bubble, Q2/PM/SL, and Q15/PM/LG) do not meet the criteria of Outfit Z Standard and PTMEA Corr; item (Q1/PS/SL, Q24/PM/GG and Q19/PM/SG) do not meet the criteria of PTMEA Corr; and item (Q25/SMRs/GG, Q16/DP/LG, and Q23/DP/GG) do not meet the criteria of Outfit ZSTD; this is supposedly caused by large size of sample, or N > 500 (Boone et al., 2014).

Wright Map: Person-Map-Item. The third step was to measure the consistency of item difficulty level and student's ability test constructed in Table 2. The higher the item difficulty level, the higher also the student's ability level will result. Information of Wright Map: Person-Map-Item is displayed in Figure 3. The previous Wright map generates that all instrument items encompass almost all the students' ability. The map generates variance from students with very high ability (> 3.0 logit), to those with very low ability (< -2.0 logit) as well. In addition to that, disparity (in which there is no item that is appropriate with the student's ability) was observed within the interval of -3.0 logit up to -0.5 logit and in the interval of +1.0 logit up to +3.7 logit. This signified that the information generated within the interval range was somewhat limited and required further elaboration. On the other hand, the item difficulty level was mostly located in the interval of -1.0 logit up to +1.0 logit; moreover, the items tended to occur in the same difficulty level. The item Q14/PS/LG was the most difficult item with a logit of +0.97, while item Q17/Pmo/LG was the easiest item with logit of -0.71.

Figure 3
Wright Map: Person-Map-Item

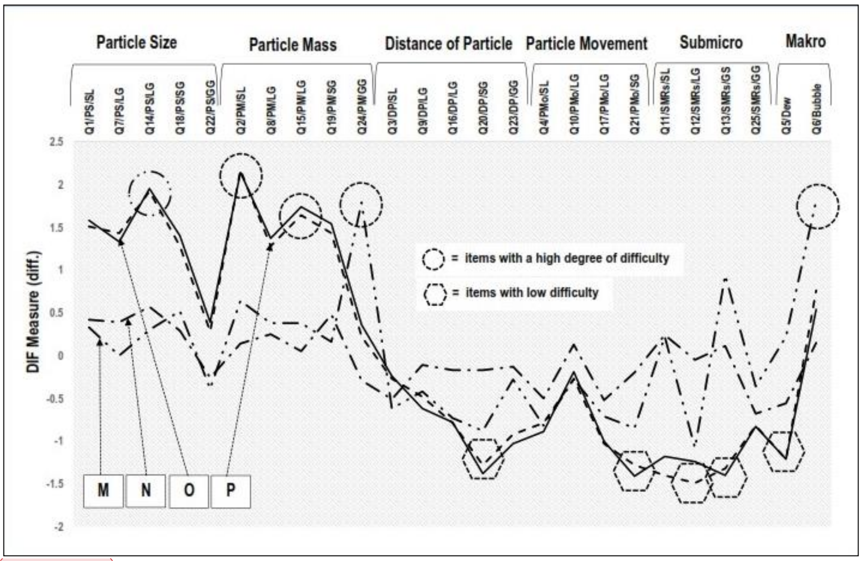


As observed from the differences in item size, some interesting cases were explained as follows: **Firstly**, the items in level 1: Q14/PS/LG (0.97) > Q1/PS/SL (0.79) > Q18/PS/PG (0.71) > Q7/PS/GG (0.66) were instead assumed by the students to possess different difficulty level. The items above, however, were more difficult than item Q6/Bubble in level 3 (0.60). In other words, determining particle size was more difficult than explaining the particle characteristics of matter in the evaporation phenomenon. **Secondly**, the size of item Q5/Dew (-0.63) < item Q6/Bubble; this indicated that it was harder for the students to elaborate on the particle characteristics of matter in the evaporation phenomenon than in condensation phenomenon, despite that both items were in the same level. **Thirdly**, the size of following items: Q2/PM/SL (0.88) > Q19/PM/SG (0.77) > Q15/PM/LG (0.66) > Q8/PM/LG (0.65) > Q24/PM/GG in level 1 was larger compared to that of items Q13/SMRs/GS (-0.24) > Q11/SMRs/SL (-0.27) > Q12/SMRs/LG (-0.63) > Q25/SMRs/GG (-0.68) in level 2. The finding illustrated that it was harder for the students to determine the particle mass than determining the submicrorepresentation (SMRs) diagram in different form changes of matter. The previous cases identified disparity in students' conceptual understanding, signifying that the level of understanding in particle characteristics of the matter is relatively low. Overall, 80% of test item difficulty level is relatively parallel with the measured constructs. By that, the test possesses good construct validity (Blanc & Rojas, 2018; Lu & Bi, 2016; Neumann et al., 2011).

Disparity in Conceptual Understanding Level

The next step was the measurement of disparity of students' conceptual understanding in the focused topic based on educational level by Differential Item Functioning (DIF).

Figure 4
Person DIF plot based on educational level



Description: M = X Class students,
N = XI Class students,
O = XII Class students, and
P = University students from chemistry department

Commented [Reviewer6]: Rewrk as a note in one line

Figure 4 of DIF plot based on students' educational level depicts that ten items are identified to possess significant disparity. **Firstly**, five curves approaching the upper limit are items with a high difficulty level (Q14/PS/LG, Q2/PM/SL, Q15/PM/SG, Q24/PM/GG and Q6/Bubble); while five curves approaching the lower limit are items with a low difficulty level (Q20/DP/SG, Q21/PMo/SG, Q12/SMRs/LG, Q13/SMRs/GS, and Q5/Dew). **Secondly**, the item Q14/PS/LG (particle size in form change of liquid-gas), Q2/PM/SL (particle mass in form change of solid-liquid), and Q15/PM/SG (particle mass in change form of solid-gas) were deemed very hard by the students of XII class and the university students compared to students in X and XI class. **Thirdly**, the research discovered different results for item Q24/PM/GG and Q6/Bubble. The item Q24/PM/GG (particle mass of O2 in larger volume) and Q6/Bubble (constructing elements of air bubbles during boiling process of water) were deemed very hard for

X class students compared to students in XI and XII classes, as well as university students. **Fourthly**, the items Q20/DP/SG (distance between particles in form change of solid-gas), Q21/PMo/SG (motion between particles in form change of solid-gas), Q12/SMRs/LG (SMRs diagram of particle in form change of liquid-gas), Q13/SMRs/GS (SMRs diagram of particle in change form of gas-liquid), and Q5/Dew (condensation) were deemed too easy for students in XII class and university students compared to the students in X and XI classes.

Pattern of Conceptual Understanding and Preconception

The analysis of the pattern of conceptual understanding and preconception employed an option probability curve test (Boone et al., 2014; Linacre, 2012). The option probability curve aims to display the probability of picking every answer choice to elaborate on the performance level of all students in the measured items (Herrmann-Abell & DeBoer, 2011). The test relied on the principle that the curve of the correct answer will rise along with the decrease of the curve of distractor choices (Boone et al., 2014; Haladyna, 2004). For items that are influenced by distractor options, the curve produced tends to be non-parallel with the traditional monotonous item behavior (Sadler, 1998), for this reason, each answer choice was analyzed separately.

The instrument provides four answer choices, thus resulting in four curves. Each curve displays the students' comprehension. Students with low ability tended to pick distractor choice, while students whose high ability were more likely to prefer other preconceptions (Herrmann-Abell & DeBoer, 2011; Perera et al., 2018). Below is the elaboration of the pattern of students' conceptual understanding and preconception based on four option probability curves.

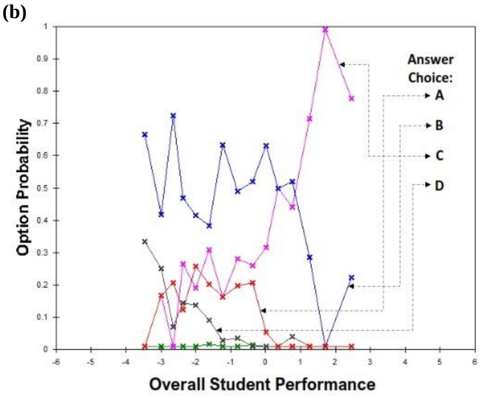
Figure 5

(a) sample of item Q2/PM/SL, (b) option probability curve

(a)
When some ice cubes in the glass melts, some other ice cubes are seen floating on the water surface. How is the comparison between mass of one particle of ice and one particle of water?

A. Mass of one particle of ice is bigger than mass of one particle of water.
B. mass of one particle of ice is smaller than mass of one particle of water.
C. mass of one particle of ice is similar with mass of one particle of water.
D Other answer...

(b)



The first example, i.e., the item Q2/PM/SL (0.88), is shown in Figure 5(a). The item measures students' capability in determining particle size in form change from solid to liquid. The option probability curve is displayed in Figure 5(b). Students with the low ability (< 0.5 logit) tended to pick distractor choice B (mass of one particle of ice is smaller than the mass of one particle of water) or A (mass of one particle of ice is bigger than the mass of one particle of water). In addition, students with very low ability (< -1.0 logit) tended to pick D (other answers). Some students with relatively low ability (> -2.5 logit), however, picked the right answer C (mass of one particle of ice is similar to the mass of one particle of water). One can predict the response pattern of students with low ability, as the distractors A, B, and D contain third preconceptions in level 1 (see Table 2). The students possess the knowledge that mass of particle of matter can change into larger or smaller size by observing the matter's change of form. It is interesting to note that there are students with the high ability (>2.0) who picked B; this indicates the presence of resistant preconception.

Figure 6

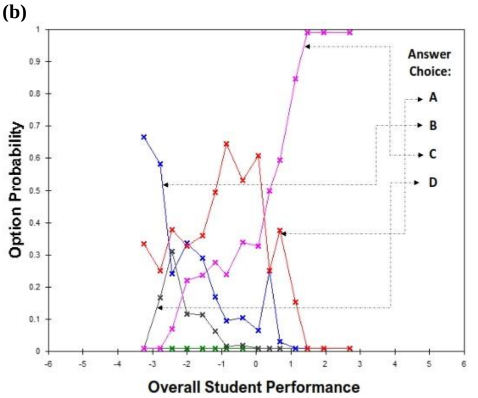
(a) item Q8/PM/LG; (b) option probability curve

(a)

During boiling process of water, a change of form of water from liquid to gas occurs. How is the comparison between mass of one particle of water and one particle of water vapor?

A. Mass of one water particle is bigger than mass of one vapor particle
B. mass of one water particle is smaller than mass of one vapor particle.
C. mass of one water particle is similar with mass of one vapor particle.
D Other answer...

(b)



The second sample or item Q8/PM/LG (0.65) is shown in figure 6(a) as the item to measure students' ability to determine the mass of the particle in form change of liquid-gas. The option probability curve is displayed in Figure 6(b). The curve of distractor B (mass of one water particle is smaller than the mass of one vapor particle) is chosen by students with low ability (< -2.0 logit), while the curve of choice A (mass of one water particle is bigger than mass of one vapor particle) was chosen by students with ability in a range of -3.5 to 1.5 logit. The correct answer, option C (mass of one water particle is similar to the mass of one vapor particle), was chosen by students with ability in > -2.5 logit. As highlighted in the table, the decline of the curve of distractor A is followed by the increase of curve of right answer C; both curves intersect in the level of 1.0 logit. The shape of curve A indicates the presence of resistant preconception type-three in level 1.

It depicts that the particular item response pattern that signifies students' conceptual understanding patterns in the given level. Moreover, the curve shape of distractors A and B in the items Q2/PM/SL and Q8/PM/SL tend to have an identical pattern. The finding indicated that students with either low or high ability had consistent preconceptions that the mass of the particle can change into larger or smaller in size along with the change in matter form.

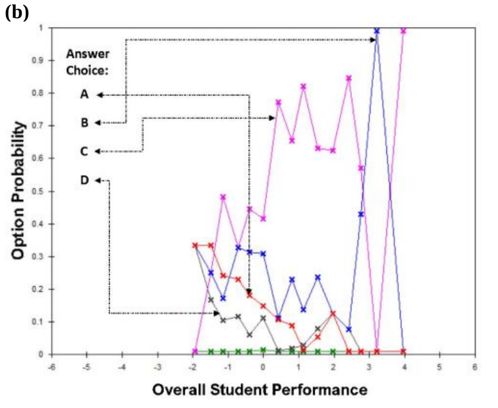
Third sample, i.e., item Q5/Dew (-0.63), as shown in Figure 7(a), measures the students' ability in elaborating characteristics of a particle in condensation phenomenon. The option probability curve is displayed in Figure 7(b). Students with low ability (< 1.0 logit) tended to pick distractor A (water drops come from liquid of melting ice that breaks through the glass wall) and option D (other answers). Some students with high ability (> 1.0 logit) also picked distractor B (water drops are the result of the reaction between ice and air nearby the glass). The shape of curve B is wavy and non-linear, even in the interval of 2.0 to 4.0 logit, it can reach option probability value up to 1.0 logit. This is regarded as a deviation from the right answer C (water drops come from condensing water vapor nearby the glass). A worth note, however, is to consider in the unstable, wavy shape of curve C. This indicated the students' inconsistency (particularly those with high ability) in comprehending the concept of condensation. This confirmed that students had their own preconception regarding concept of condensation.

Figure 7
(a) item Q5/Dew; (b) option probability curve

(a)
In a glass filled with ice, you can see water drops at the glass. According to you, where do the water drops come from?

A. Water drops come from melting ice that penetrates the glass wall
B. water drops are the result of reaction between ice and air nearby the glass
C. Water drops come from condensing water vapor nearby the glass
D. Other answer...

(b)



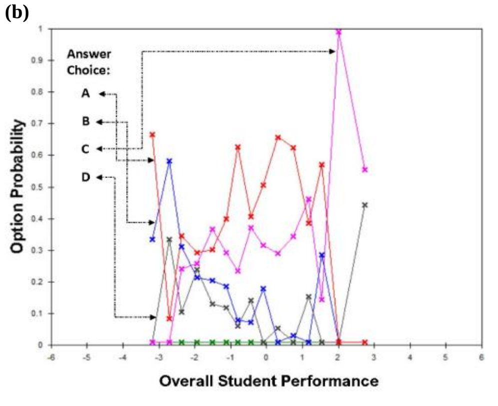
The fourth sample or item Q6/Bubble, as shown in Figure 8(a), measures the students' ability in elaborating characteristics of a particle in the evaporation phenomenon. The option probability curve is displayed in Figure 8(b). The distractor A (air bubbles are Hydrogen and Oxygen particles) was dominantly chosen by students whose ability in a range between -3.0 to 2.0 logit. Moreover, the distractor B (air bubbles are Hydrogen and Oxygen particles) was mostly selected by students whose ability in a range between -3.0 to 0.5 logit. The form of curve A and B were picked by students with low ability was predictable. The curve of right answer C (air bubbles are water molecules), however, shows interesting hint; in the interval range of -2.5 to 3.0 logit, the tip of the curve shows an up-and-down pattern. Moreover, in the level of 1.5 logit, the curve shape of distractors A and B shows a decline pattern, while that of curve C tends to increase. Another finding worth noting was that the curve D (other answers) was picked by some students with high ability (> 2.0 logit). This indicated that particular students had their own preconceptions regarding the evaporation concept.

Figure 8
(a) item Q6/Bubble; (b) option probability curve

(a)
In a container filled with boiling water, you can see air bubbles on the top of it. According to you, what are the composing elements of the air bubbles?

A. Hydrogen and Oxygen particles
B. Air that is dissolved in water
C. Water molecules
D. Other answer...

(b)



Discussion ~~and Implications~~

The research findings indicated that the instruments had good effectiveness, met the requisites of person and item reliability, and showed good construct validity. When applied in evaluating students' conceptual understanding, it was found that: Firstly, almost all students with high ability faced difficulty in understanding the concept of particle size and mass in level 1. The same students found it relatively easy in determining SMRs diagram of particle structure in level 2 or determining the concept of particle regarding evaporation and condensation phenomena in level 3. Secondly, the information of the response pattern of students with high ability was quite consistent, repetitive, and systematic in particular items. This indicates the presence of permanent and latent preconceptions. The analysis of the option probability curve of item Q2/PM/SL (0.88), Q8/PM/LG (0.65), Q5/Dew (-0.63) and Q6/Bubble (0.60) indicates that the approach of item response pattern is able to explore in detail and comprehensively regarding students' conceptual understanding and preconception.

Sequences of verification conducted that involves Rasch model approach shows detailed, accurate, and quantifiable results since the approach integrates development procedure of diagnostic and summative instruments. Several samples of preconception, e.g., item Q2/PM/SL (0.88) and Q8/PM/LG (0.65) indicate that distractor options are potential to be elaborated further in order to investigate tendency of preconception by the students. In addition, it also provides information regarding main idea unknown to the students and their degree of misunderstanding.

The approach employed in this research is an effective illustration to help teacher in evaluating the learning process as well as the students' learning progress. This is due to the integration of qualitative item development procedure and quantitative data analysis, allowing the teachers to explore in-depth on the students' understanding, concepts the students understand and/or do not **understand, and misconception.** Such findings echo Herrmann-Abell & Deboer (2016) that the integration of Rasch model analysis and probability curve is applicable to diagnose how the students' misconception turns into their overall conceptual understanding. Such an attempt is quite hard to conduct by implementing a conventional approach due to the interdependence of person and item. Rasch model, on the other hand, is able to tackle such interdependence, in which the item and the test difficulty remain invariant and not dependent on which sample that is involved in the initial validation. This signifies that the instrument's items have met the unidimensionality and local independence requirements (Jin et al., 2019; Testa et al., 2019; Wei et al., 2012).

Overall, the research indicated empirical evidence that supported findings by Hoe & Subramaniam (2016); Lu & Bi (2016); and Rogat et al., (2011), that students had distinctive preconception as a result of a learning process they experienced. Such preconception was regarded as the inhibitor to the development process of students' conceptual understanding (Soeharto et al., 2019). In this research, students' preconception was found to be repetitive and systematic in each education level. It signifies that the intervention to change students' preconceptions was difficult to conduct by the conventional learning method. A strategic and meaningful learning method is therefore essential to remove students' incorrect preconceptions and develop scientifically correct conceptual understanding. That being said, teachers are demanded to acquire detailed information on the forms and characteristics of students' preconceptions. In conclusion, the item response pattern analysis was an efficient and effective

Commented [Reviewer7]: results

Commented [Reviewer8]: and

means to acquire such information. The information on students' preconception is important as the basis to develop appropriate and measurable instructional design in solving the students' misconception. This is in line with the previous researches, arguing that the quality of learning progress is highly dependent on the students' learning process and learning experience (Duschl et al., 2011; Park et al, 2017; Wilson, 2009).

Conclusions

The measuring instrument developed performed well in its validity and reliability, thus, it is deemed applicable in measuring students' conceptual understanding and preconception in elaborating particle characteristics of matter. During the implementation of the instruments, the research finds out that: 1) almost all students with high ability face difficulty in understanding the concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, as well as determining the concept of particle regarding evaporation and condensation phenomena in level 3.

2) There is a significant disparity between students' conceptual understanding based on their educational level. 3) In certain cases, it is found that the distractor item response pattern by high-ability students tends to be consistent, indicating a certain tendency of resistant preconception pattern.

The development of diagnostic instruments with Rasch model approach is deemed as the literacy process for practitioners and researchers in Indonesia. The result indicates that there is no single item that is parallel with both the highest ability and lowest ability students. This calls for further elaboration in order to improve the instrument items' quality. Moreover, an anomaly is found that students with high ability (> 1.0 logit) tend to pick distractor choices. This urges further studies to investigate structured comprehension problems. The research regards that further analysis that integrates conceptual understanding level and items designed in a gradual manner is required to define the characteristics of the students' alternative conception and to measure their learning progress. Echoing this notion, one must integrate the item design and basic principles of chemistry as a reference for further researchers and educational practitioners to implement the same approach conducted in the present research. On top of that, despite not focused on discussing matters regarding students' learning progress individually, the instrument is expected to be beneficial for the teachers to diagnose students' conception in developing an effective and meaningful learning experience.

Acknowledgments

This research is funded by the Directorate of Research and Community Outreach/ (Direktorat Riset dan Pengabdian Masyarakat or DRPM), Directorate General of Research and Development (Dirjen Riset dan Pengembangan) of Ministry of Research and Technology of the Republic of Indonesia through Penugasan Penelitian Dasar Unggulan Perguruan Tinggi Grant Program, Research and Community Outreach Institute (Lembaga Penelitian dan Pengabdian Masyarakat) of Universitas Negeri Gorontalo in 2020.

References

Adadan, E., & Savasci, F. (2012). An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, *34*(4), 513–544. https://doi.org/10.1080/09500693.2011.636084

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

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

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

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

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

Blanc, A., & Rojas, A. J. (2018). Use of rasch person-item maps to validate a theoretical model for measuring attitudes toward sexual behaviors. *PLoS ONE*, *13*(8), 1–17. https://doi.org/10.1371/journal.pone.0202551

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

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

Boz, Y. (2006). Turkish pupils' conceptions of the particulate nature of matter. *Journal of Science Education and Technology*, *15*(2), 203–213. https://doi.org/10.1007/s10956-006-9003-9

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of two tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, *8*(3), 293–307.

Cheng, M. M. W. (2018). Students' visualisation of chemical reactions-insights into the particle model and the atomic model. *Chemistry Education Research and Practice*, *19*(1), 227–239. https://doi.org/10.1039/c6rp00235h

Davidowitz, B., Chittleborough, G., & Murray, E. (2010). Student-generated submicro diagrams: A useful tool for teaching and learning chemical equations and stoichiometry. *Chemistry Education Research and Practice*, *11*(3), 154–164. https://doi.org/10.1039/c005464j

Devetak, I., & Glažar, S. A. (2010). The influence of 16-year-old students' gender, mental abilities, and motivation on their reading and drawing submicrorepresentations achievements. *International Journal of Science Education*, *32*(12), 1561–1593. https://doi.org/10.1080/09500690903150609

Devetak, I., Urbancic, M., Grm, K. S. W., Krnel, D., & Glazar, S. A. (2004). Submicroscopic representations as a tool for evaluating students' chemical misconception. *Acta Chimica Slovenica*, *51*(December 2004), 799–814. https://www.researchgate.net/publication/286849527_Submicroscopic_representations_as_a_tool_for_evaluating_students'_chemical_conceptions

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, *47*(2), 123–182. https://doi.org/10.1080/03057267.2011.604476

Femintasari, V., Effendy, & Munzil. (2015). The effectiveness of two-tier multiple choice test and multiple choice test followed with interview in identifying misconception of students with different scientific reasoning skills. *Jurnal Ilmu Pendidikan*, *21*(2), 192–197. http://journal.um.ac.id/index.php/jip/article/view/8376/4020

Gabel, D. L. (1993). Use of the particle nature of matter in developing conceptual understanding. *Journal* *of Chemical Education*, *70*(3), 193. https://doi.org/10.1021/ed070p193

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

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

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

Herrmann-Abell, C. F., & Deboer, G. E. (2016). Using Rasch modeling and option probability curves to diagnose students' misconceptions. *American Eduacational Research Assossiation*, 1–12.

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

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

Jin, H., Mikeska, J. N., Hokayem, H., & Mavronikolas, E. (2019). Toward coherence in curriculum, instruction, and assessment: A review of learning progression literature. *Science Education*, *103*(5), 1206–1234. https://doi.org/10.1002/sce.21525

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

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

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

Kapici, H. Ö., & Akcay, H. (2016). Particulate nature of matter misconceptions held by middle and high school students in Turkey. *European Journal of Education Studies*, *2*(8), 43–58. https://doi.org/10.5281/zenodo.163547

Kim, J., & Wilson, M. (2019). Polytomous item explanatory item response theory models. *Educational and Psychological Measurement*. https://doi.org/10.1177/0013164419892667

Kind, V. (2004). *Beyond appearances: Students' misconceptions about basic chemical ideas* (2nd Edition). Royal Society of Chemistry. https://www.researchgate.net/publication/228799159_Beyond_Appearances_Students'_Misconceptions_About_Basic_Chemical_Ideas

Laliyo, L. A. R., Botutihe, D. N., & Panigoro, C. (2019). The development of two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12

Linacre, J. M. (2012). *A user's guide to W I N S T E P S ® M I N I S T E P Rasch-model computer program: Program manual 3.75.0*. https://doi.org/ISBN 0-941938-03-4

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

665). Springer Netherlands,.

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

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, *34*(15), 2317–2341. https://doi.org/10.1080/09500693.2012.684433

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

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*(9), 825–838. https://doi.org/10.1002/tea.3660200905

Ozalp, D., & Kahvecib, A. (2015). Diagnostic assessment of student misconceptions about the particulate nature of matter from ontological perspective. *Chemical Education Research and Practice*, *16*, 619–639. https://doi.org/10.1039/C5RP00096C

Ozmen, H. (2011). Turkish primary students´ conceptions about the particulate nature of matter. *International Journal of Environment & Science Education, 6*(1), 99–121. https://www.researchgate.net/publication/265438996_Turkish_primary_students'_conceptions_about_the_particulate_nature_of_matter

Park, M., Liu, X., & Waight, N. (2017). Development of the connected chemistry as formative assessment pedagogy for high school chemistry teaching. *Journal of Chemical Education*, *94*(3), 273–281. https://doi.org/10.1021/acs.jchemed.6b00299

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

Renström, L., Andersson, B., & Marton, F. (1990). Students' conceptions of matter. *Journal of Educational Psychology*, *82*(3), 555–569. https://doi.org/10.1037/0022-0663.82.3.555

Rogat, A., Anderson, C., Foster, J., Goldberg, F., Hicks, J., Kanter, D., Krajcik, J., Lehrer, R., Reiser, B., & Wiser, M. (2011). *Developing learning progression in support of the new science standards: A RAPID workshop series, New York, NY.* (4), 163. https://doi.org/10.12698/cpre.2011.lprapid

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

Sadler, P. M. (1999). The relevance of multiple-choice testing in assessing science understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 251–274). https://zodml.org/sites/default/files/%5BJoel_J._Mintzes%2C_James_H._Wandersee%2C_Joseph_D._No_0.pdf

Slapničar, M., Devetak, I., Glažar, A. S., & Pavlin, J. (2017). Identification of the understanding of the states of matter of water and air among Slovenian students aged 12, 14 and 16 years through solving authentic tasks. *Journal of Baltic Science Education*, *16*(3), 308–323.

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

Stojanovska, M. I., Soptrajanov, B. T., & Petrusevski, V. M. (2012). Addressing misconceptions about the particulate nature of matter among secondary-school and high-school students in the Republic of Macedonia. *Creative Education*, *03*(05), 619–631. https://doi.org/10.4236/ce.2012.35091

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

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial [Application of Rasch Model in Social Sciences Research].* Tri Komunikata. https://www.researchgate.net/publication/268688933%0AAplikasi

regular

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan [Application of Rasch Model in Education Assessment].* Tim Komunikata. https://www.researchgate.net/publication/282673464%0AAplikasi

regular

Taber, K. S. (2015). Alternative Conceptions/Frameworks/Misconceptions. In G. Richard (Ed.), *Encyclopedia of Science Education* (pp. 37–41). https://doi.org/10.1007/978-94-007-6165-0

Testa, I., Capasso, G., Colantonio, A., Galano, S., Marzoli, I., Scotti di Uccio, U., Trani., F & Zappia, A. (2019). Development and validation of a university students' progression in learning quantum mechanics through exploratory factor analysis and Rasch analysis. *International Journal of Science Education*, *41*(3), 388–417. https://doi.org/10.1080/09500693.2018.1556414

Tóth, Z., & Kiss, E. (2006). Using particulate drawings to study 13-17 year olds' understanding of physical and chemical composition of matter as well as the state of matter. *Practice and Theory in Systems of Education*, *1*(1), 109–125. http://eduscience.fw.hu/

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169. https://doi.org/10.1080/0950069880100204

Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, *4*(6), 626–631. http://www.academicjournals.org/SRE

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335–345. https://doi.org/10.1021/ed100852t

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

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

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

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

Yaşar, I. Z., İnce, E., & Kırbaşlar, F. G. (2014). 7. Class Science and Technology Course "Structure of Atom" Subject Readiness Improvement Test. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2014.09.259

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

The list of references should be carefully revised. Text should be justified: align text to both the left and right margins, adding extra space between words as necessary

Accept without revision	
Accept with minor revisions	X
Accept: with moderate revisions	
Pre-Accept: major revisions and re-evaluation	
Reject: Rework and re-submit	
Reject: do NOT re-submit	

Explanation / rebuttal letter (10 Juli 2020)

No.	Comments from Reviewers	Section & Page	Explanation/Rebuttal
1	Note that a keyword does not have to be made of only one word!	Abstract: Page 1	The keywords have been changed according to the reviewer's request. (DONE)
2	"employed a non-experimental approach"	General Background: Page 3	The authors have no idea why it has been highlighted.
3	Lower-secondary school	Respondent: Page 3	The ethical aspects have been added according to the reviewer's request. (DONE)
4	Ethical aspects?	Respondent: Page 4	It has been changed according to the reviewer's request. (DONE)
5	Remove an empty row	Table 3: Page 6	It has been removed according to the reviewer's request. (DONE)
6	The APA Manual (6/7th ed.) says: "Order the citations of two or more works within the same parentheses alphabetically" (6.16 on page 177).	Effectiveness of Measuring Instruments: Page 7	It has been changed according to the reviewer's request. (DONE)
7	Rewrk as a note in one line	Figure 4's Note: Page 11	It has been changed according to the reviewer's request. (DONE)
8	"and Implication" was asked to removed	Discussion: Page 14	It has been changed according to the reviewer's request. (DONE)
9	Results	Discussion: Page 14	It has been changed according to the reviewer's request. (DONE)
10	and	Discussion: Page 15	It has been changed according to the reviewer's request. (DONE)
11	The list of references should be carefully revised. Text should be justified: align text to both the left and right margins, adding extra space between words as necessary	References	It has been changed according to the reviewer's request. (DONE)

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

200419.jbse_JL-suo_review
Re-evaluation

Abstract. *This research aimed to evaluate the students' conceptual understanding and to diagnose the students' preconceptions in elaborating the particle characteristics of matter by development of diagnostic instrument as well as Rasch model response pattern analysis approach. Data were acquired by 25 multiple-choice written test items distributed to 987 students in North Sulawesi, Indonesia. Analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding; and 3) diagnosis of students' preconception by estimation of item response pattern. The result generated information on the diagnostic and summative measurement on students' conceptual understanding in elaborating the topic; information also acts as empirical evidence on the measurement's reliability and validity. Moreover, the result discovered a significant disparity between students' conceptual understanding based on their educational level. It was found that the distractor item response pattern tended to be consistent, indicating a certain tendency of resistant preconception pattern. The findings are expected to be a recommendation for future researchers and educational practitioners that integrates diagnostic and summative measurement with Rasch model in evaluating conceptual understanding and diagnosing misconception.*
Keywords: *conceptual understanding, item response, particle of matter, Rasch model*.

Introduction

Central to the notion of learning about characteristics of a particle of matter is the process of developing an understanding on abstract concepts (Johnstone, 1991) without directly interacting with the object/fact (Stojanovska et al., 2012); therefore it is considered a difficult subject for the students to learn. Echoing this, the disparity in understanding is almost inevitable (Kapici & Akcay, 2016) since different students may develop ones' own distinctive way of understanding a concept (Yildirir & Demirkol, 2018). The idea is also coined by experts as misconception (Johnstone, 2006, 2010; Taber, 2015), or alternative framework and preconception (Lu & Bi, 2016). The experts have discovered that students always have their own preconception that is not in line with scientific concepts (Alamina & Etokeren, 2018; Yaşar et al., 2014); therefore, one needs to conduct identification and improvement on the conceptual learning (Allen, 2014; Soeharto et al., 2019).

In diagnosing preconceptions, several researchers have developed diagnostic instruments in different mechanisms (McClary & Bretz, 2012), i.e., conceptual map, essay test, interview, essay test with interview, or multiple-choice test (Femintasari et al., 2015). Two-step multiple choice diagnostic test (Adadan & Savasci, 2012; Chandrasegaran et al., 2007; Treagust, 1988; Tüysüz, 2009) is preferred due to its ability to diagnose preconception and describe the underlying reasons. The instrument is indeed considered qualitatively effective in elaborating differences in students' thought processes; however, it does not provide summative measurement features due to lack of internal consistency and the instrument's unidimensionality (Lu & Bi, 2016). In addition to that, the measurement conclusion generated is considered weak due to extracted from analysis on the raw score (Sumintono, 2018)

Studies on preconception have found that the concept is somewhat resistant. In the early 2000s, it is discovered that students' preconceptions persisted even when they already undergo formal education experience (Hoe & Subramaniam, 2016). Preconception can also change along with the development of students' conceptual understanding; it also varies in different levels of understanding (Aktan, 2013). If one conducts a two-step test and raw score analysis approach to diagnose resistant preconception, the result generated will only provide limited feedback information (Sumintono, 2018) due to the instrument's limitation in measuring students' conceptual understanding. Instead of supporting, the information will only make it harder for teachers to implement proper instructional decisions (Wilson, 2008).

During the middle of the 2000s, the Rasch model analysis was commonly used in studies of chemistry education (Herrmann-Abell & DeBoer, 2011; Liu, 2012; Wein et al., 2012). The approach provides a testing apparatus that integrates diagnostic and summative measurement. Recently, this approach is used to develop formative assessment with the intention to conduct learning construction mapping, e.g., measuring the students' way of constructing their understanding process (Hadenfeldt et al., 2013). It is worth to note, however, that there are studies that integrate diagnostic and summative measurement with a different approach (Hoe & Subramaniam, 2016); despite that, trends in chemistry education studies highlight that diagnostic-summative measurement by Rasch model analysis is more common to be carried out (Laliyo et al., 2019; Lu & Bi, 2016).

Research Problem

The characteristics of a particle of matter is a fundamental concept in chemistry, usually taught in middle education level. Adequate comprehension regarding the particle characteristics of matter both in macroscopic and microscopic level is essential as the knowledge basis in understanding more advanced topics such as the concept of atoms and molecules as the submicroscopic component that is invisible to plain eyesight but exists in all real-world phenomena (Cheng, 2018; Ozmen, 2011; Yildirir & Demirkol, 2018). The fact signifies the relevance and reasoning of complexity in chemistry learning that is considered difficult for both students and teachers to conduct (Alamina & Etokeren, 2018). In simpler terms, to ensure that the chemistry learning is conducted effectively, one requires to nurture students' comprehensive understanding regarding particle characteristics of matter and its change of state.

To evaluate the students' conceptual understanding on the aforementioned topic, one also needs to measure the students' capability in interpreting particle state during change process of a matter's form (Alamina & Etokeren, 2018; Barbera, 2013; Boz, 2006; Cheng, 2018; Gabel, 1993; Hadenfeldt et al., 2013; Kapici & Akcay, 2016; Kind, 2004; Naah & Sanger, 2012; Ozalp & Kahvecib, 2015; Ozmen, 2011; Renström et al., 1990; Slapničar et al., 2017; Stojanovska et al., 2012; Yildirir & Demirkol, 2018). Researches on particle characteristics and changes of matter generally employ diagnostic instruments in the form of essay tests and/or essays followed by interview; the instruments are further analyzed based on raw score results. The approach is considered inefficient and somewhat lacked accuracy in measuring students' conceptual understanding and misconception pattern. Despite its ineffectiveness, the conventional method is used by most teachers in Indonesia to measure and determine students' learning progress. The teachers argue that measuring the students' raw score is effective in determining how far the students have progressed in the learning process. The students' raw score is regarded by many as an early premature indication regarding the measured variable and is not eligible to be the final measurement indicator due to its temporary nature. In addition to that, regarding the decision-making process, the raw score contains only limited information for it to be treated as reference (He et al., 2016; Sumintono & Widhiarso, 2015)

Research Focus

The research focuses on developing a diagnostic instrument that integrates measurement of conceptual understanding and diagnosis of students' preconceptions regarding the aforementioned topic by the approach of Rasch model item response pattern analysis. The analysis employs different test apparatuses to provide extensive information for practitioners and researchers in science education in evaluating students' learning progress in different topics.

Research Questions

This research aimed to figure out the following questions: 1) How is the effectiveness of measurement instrument to evaluate the students' conceptual understanding and diagnose their preconceptions on the characteristics of a particle of matter? 2) Is there any significant difference between students in elaborating on the aforementioned topic based on their educational level? 3) How is the pattern of students' conceptual understanding and preconception regarding the topic?

Research Methodology

General Background

The descriptive-quantitative research employed a non-experimental approach, in which the students' conceptual understanding in explaining the characteristics of a particle of matter was treated as the measurable variable. Prior to conducting the research, it was ensured that the students already experience formal learning of the aforementioned topic. The researchers did not conduct any intervention on the learning process or the learning material. In other words, no treatment was implemented to the students for them to be able to answer all test items in the measurement instrument.

The data collection step was implemented for four months in the even semester of the 2019-2020 academic year; the process was conducted after obtaining approval from the Government of Province of Gorontalo and heads of universities in the Northern part of Sulawesi, Indonesia. Moreover, the schools' and parents' approval was obtained in cooperation with the school committee. The school administrators were willing to facilitate the data collection process that adjusted with the schedule.

Respondent

The respondents were 987 people consisting of students of eleventh grade from eight lower-secondary schools well as university students of the chemistry department in Northern Sulawesi, Indonesia. The distribution of respondents is displayed in Table 1 below.

Table 1
Demographic profile of respondents (N=947)

Demography	Code	Respondents	Percentage (%)
Gender			
Male	M	320	67.68
Female	F	667	32.42
Education level			
X Class students	M	168	17.02
XI Class students	N	473	47.92
XII Class students	O	186	18.84
University students from the chemistry department	P	160	16.21

The respondents were chosen randomly and have voluntarily agreed to participate in the research. In addition, they received no learning treatment and other special treatments that allow them to complete the measurement instrument. Students were asked to write down their responses in the answer sheet; the process was supervised by teachers in the respective schools and lecturers in the respective university. All students were instructed to answer all questions in the instruments within 45 minutes. All instrument sheets and answer sheets were collected by the researchers shortly after the session ended; it was ensured that the numbers of instruments matched the numbers of participants. For the certainty in ethical consideration, permission was obtained from the school administration after coordinating with students' parents through the school committee. This process was conducted before the students were invited to participate in research. Permission for the students was obtained from the department leaders of the university, and student written statements. All students were told that the confidentiality of their identity was fully guaranteed, and the results of the study would only be used for research purposes.

Instrument and Procedures Development

The design process refers to a recommendation by Wilson (2005), which consists of four key steps: definition of construct map, item design, result blank, and measurement model.

Phase 1: Definition of construct map. The map offers a substantive definition of measured constructs; the more constructs measured, the constructs' level will vary qualitatively (Wilson, 2009). In simpler words, it aims to develop the students' understanding map to measure the students' progress (Wilson, 2012). The instrument involved variables, i.e., the students conceptual understanding and preconception in elaborating the characteristics of a particle of matter; it was conducted in accordance with the Curriculum Standard of Chemistry Subject in Tenth Grade in Indonesia, as presented in Table 2.

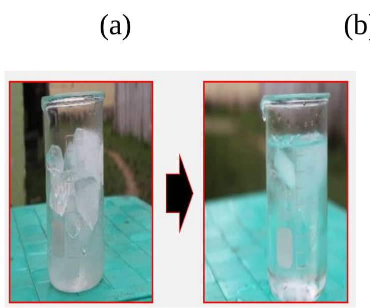
Table 2
Conceptual Understanding Level

Level 3 The students are able to connect between characteristics of a particle of matter in macroscopic and submicroscopic level	
Phenomenon	Evaporation: item Q6/Bubble
10. Preconception	Air bubble consists of Hydrogen and Oxygen particles
9. Preconception	Air bubble is water-soluble
Phenomenon	Condensation: item Q5/Dew
8. Preconception	Water drops come from melting ice that penetrates the glass wall
7. Preconception	Water drops are the result of the reaction between ice and air nearby the glass
Level 2 The students are able to determine SMRs diagram of particle structure during a change of form: item Q11/SMRs/SL; Q12/SMRs/LG; Q13/SMRs/GS; Q25/SMRs/GG	
6. Preconception	The SMRs diagram of particle structure follows the physical form of matter
5. Preconception	The SMRs diagram of O_2 molecule shape undergoes change as a result of an increase in the volume of the container.
Level 1 The students are able to determine the characteristics of a particle of matter during the change process of matter's form.	
4. Preconception	The particle size of matter changes into (large/small) as a result of change in matter form: item Q1/PS/SL; Q7/PS/LG; Q14/PS/LG; Q18/PS/SG; Q22/PS/GG
3. Preconception	The particle mass of matter changes into (large/small) due to change in matter form: item Q2/PM/SL; Q8/PM/LG; Q15/PM/LG; Q19/PM/SG; Q24/PM/GG
2. Preconception	Distance between matter particles changes into (faster/slower) due to change in matter form: item Q3/DP/SL; Q9/DP/SL; Q16/DP/LG; Q20/DP/SG; Q23/PM/GG
1. Preconception	Motion between matter particles changes into (dense/loose) due to change in matter form: item item Q4/PMo/SL; Q10/PMo/LG; Q17/PMo/LG; Q21/PMo/SG

Variation in conceptual understanding level illustrates the development process of the students' conceptual understanding. In the first level, the students were asked to determine particle characteristics (size, mass, motion, and distance) in the change process of matter form. In the second level, the students were asked to determine the submicroscopic representation diagram of particle structure. Further, in the third level, the students were asked to connect between characteristics of a particle of matter at the macroscopic and submicroscopic level. In each level, the construct map also features the students' tendency of preconception.

Phase 2: item design and evaluation The phase involved the determination process of items to be used in acquiring evidence of students' construct understanding regarding the construct map (Wilson, 2005). Certain items may have a different extent of effectiveness to measure students' conceptual understanding (Sadler, 1999); however, multiple choices item is considered more practical and effective (Wilson, 2008). The instrument of concept understanding test of the particle (or TPKP) is adapted from multiple-choice instruments by (Herrmann-Abell & DeBoer, 2011). Each item consists of two distractor answer choices and one open answer choice. The distractor answer choices are designed by referring to the common preconceptions by the students (see Table 2) as logical choices to distract the students from the correct one. The distractors function to emphasize the item diagnostic strength (Sadler, 1998). Some of the items are adopted from previous studies Osborne & Cosgrove (1983), Renström et al., (1990); Devetak et al., (2004); Tóth & Kiss (2006); Davidowitz et al., (2010); Devetak & Glažar (2010); Slapničar et al., (2017) and (Yildirir & Demirkol, 2018).

Figure 1
Sample of item Q1/PS/SL design



Glass (a) contains ice chunks; glass (b) contains melting ice chunks. How is the size of water particle in solid form (ice) compared to that in liquid form?

a. Size of a water particle in solid form > a water particle in liquid form.
b. Size of a water particle in solid form < a water particle in liquid form.
c. Size of a water particle in solid form = a water particle in liquid form.
d. Other answers

Figure 1 displays a sample of item Q1/PS/SL design, in which Q1 is the number of item 1, PS is particle size, and SL is solid-liquid. The item measures student's capability in determining particle size in form change from solid to liquid. The choice A and B are distractors, the correct choice is C, and choice D is for other answers students may fill if the existing answer choices are not in accordance with their initial knowledge. Every correct answer was given mark 1, and wrong answers got 0 mark. Each student only has a slight probability of 0.25 in choosing the right answer. The students will pick what they think the right answer based on their understanding. If the distractor item choice functions well, the students will not be able to predict the correct answer.

Phase 3: design of result blank, i.e., the correlation between construct map and items (Wilson, 2005). This phase aimed to identify whether the answer the students pick correlates with their conceptual understanding; in simpler terms, it was intended to elaborate the conformity between the variable contents being measured. In order to elaborate on the previous aspect, the TPKP instrument was validated by three independent experts and tested to the students to acquire their feedback. The process acquired 25 items of TPKP. Prior to the data collection process, it was ensured that all students had received formal education on the characteristics of a particle of matter and its changes. The students' response towards the instrument was inputted manually by the written answer sheet. The test was supervised by the teachers in school by referring to the agreed permission and duration. Each student was required to finish all test items within the allocated duration of 45 minutes. The instrument sheets were further collected, and checking process was conducted to ensure that the amount of instrument sheet wass the same with participating students.

Phase 4: Rasch model analysis approach. The analysis integrates algorithm as a result of probabilistic expectation of item 'i' and student 'n', as: The statement is the probability of student n in item i to result in the correct answer (x = 1); with student ability, ßn, and item difficulty level (Bond & Fox, 2015). The above equation was simplified by inserting logarithm function, into , so that the probability of picking the right answer equals to student's ability subtracted by item difficulty level. The student (person) and item units were considered on the same interval scale and were independent of each other. The students' ability level and item difficulty level were measured in the logarithm unit, namely odds or log that variates from -00 to +00 (Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015). The instrument efficiency, when compared to the item distribution towards item difficult level with distribution of student's ability level, was quantifiable in order to measure the students' conceptual understanding. In addition, the student's understanding level was differentiated based on the item size. The previous steps highlighted the main difference of Rasch model analysis when compared to the raw score-based conventional one; the latter lacks accuracy in evaluating students' ability observed from different item difficulty level (Herrmann-Abell & DeBoer, 2011; Lu & Bi, 2016; Sumintono & Widhiarso, 2015).

Data Analysis

The research employed WINSTEPS version 3.75 software to convert raw data into interval data (Bond & Fox, 2015; Linacre, 2012). The conversion result acted as the calibration of data on the student's ability level and item difficulty level within the same interval measurement. Moreover, the analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding by Differential Item Functioning (DIF) item test; and 3) diagnosis of students' preconception by estimation of item response pattern through option probability curve test.

Research Results

Effectiveness of Measuring Instruments

Person and Item Reliability. The first step to elaborate on the effectiveness of measuring instruments was by measuring the person and item reliability. This was conducted to gather information to what extent the measurement produces consistent information in displaying latent trait or the unidimensionality of the measured variable (Sumintono & Widhiarso, 2015). The analysis result is presented in the form of a statistical summary (Table 3).

Table 3
Summary of fit statistics

Parameter (N)	Measure	INFIT		OUTFIT		Separation	Reliability	*SD*	KR-20
		MNSQ	ZSTD	MNSQ	ZSTD				
Person (987)	-.34	1.00	-.11	1.02	-.1	1.55	.71	.88	.72
Items (25)	.00	1.00	-.75	1.02	-.1	8.18	.99	.60	

The above table indicates that the person reliability value of 0.71 is equivalent to the person separation index value of 1.55. This is to say that the consistency of students' response towards the test is deemed good. In addition to that, it is generated that the *Cronbach Alpha Coefficient* (KR-20) value is 0.72, signifying good interaction between students and the test. This further indicates strong correlation between the students' response towards the item, in the context that the students' knowledge tends to be non-fragmented, enabling it to be measured (Adams & Wieman, 2011). To the researchers and educational practitioners, such information is essential to prepare for follow-up plans and development of students' ability (Wei et al., 2012). Moreover, the result generated a relatively high value of item separation index of 8.18 that was equivalent to the item reliability value of 0.99. This indicated very good item consistency or the item was deemed capable of meeting the unidimensionality criteria. In other words, the item performed very good in defining the measured variable. This was confirmed by the infit and outfit value result, in which most of the items were in the acceptable range for the multiple-choice test (Bond & Fox, 2015; Herrmann-Abell & DeBoer, 2011).

Figure 2
Function of Measurement Information

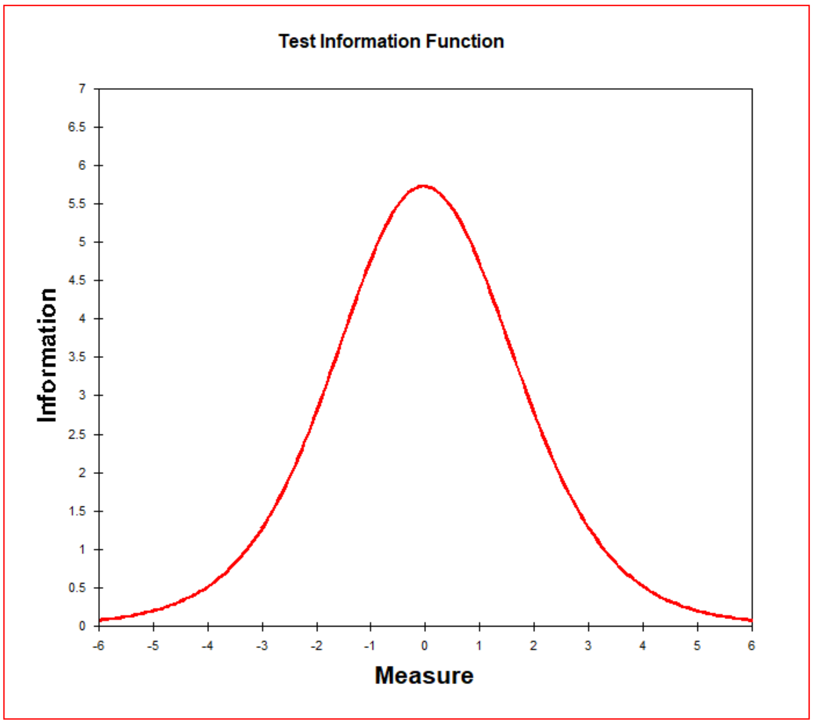


Figure 2 displays the graph of measurement information in order to show the measurement reliability. The higher the tip of information function graph, the measurement reliability value is likely to increase. In the intermediate level of students' ability (-3.0 logit up to +3.0 logit), the measurement information is in very high spot. This indicates that the TPKP instrument is capable of producing optimal information to students with an intermediate level of ability. Such a result means that the instrument possesses high measurement reliability (Bond & Fox, 2015; Kim & Wilson, 2019).

Validity. The next step was to measure the item validity by Fit item test to ensure that all items fit with the Rasch model. The process was aimed to identify whether or not the test item could measure the aspects that intended to be measured, or test validity (Linacre, 2012; Sumintono & Widhiarso, 2014). The criteria used comprise outfit means-square (MNSQ): $0.5 < y < 1.5$; outfit z-standard: $-2.0 < Z < +2.0$, as well as point measure correlation (PTMEA Corr). The PTMEA Corr is the correlation between the score of item and person measure that is required to be a positive value and not approaching zero (Bond & Fox, 2015). The PTMEA Corr criteria: 0.4 < x < 0.8. If all three criteria are not met, the item is not good enough and needs further elaboration (Boone et al., 2014). Both Outfiit MNSQ and Infit MNSQ were sensitive chi-squares in detecting outlier response pattern. There were two outlier responses: the right response, guessed by the students with low ability in item with high difficulty level, or the wrong response due to the high-ability students' carelessness in items with a low difficulty level. The expected ideal MNSQ value is 1.0. The analysis result on item appropriateness is displayed in Table 4 as follows:

Table 4

Item Statistics: Misfit Order

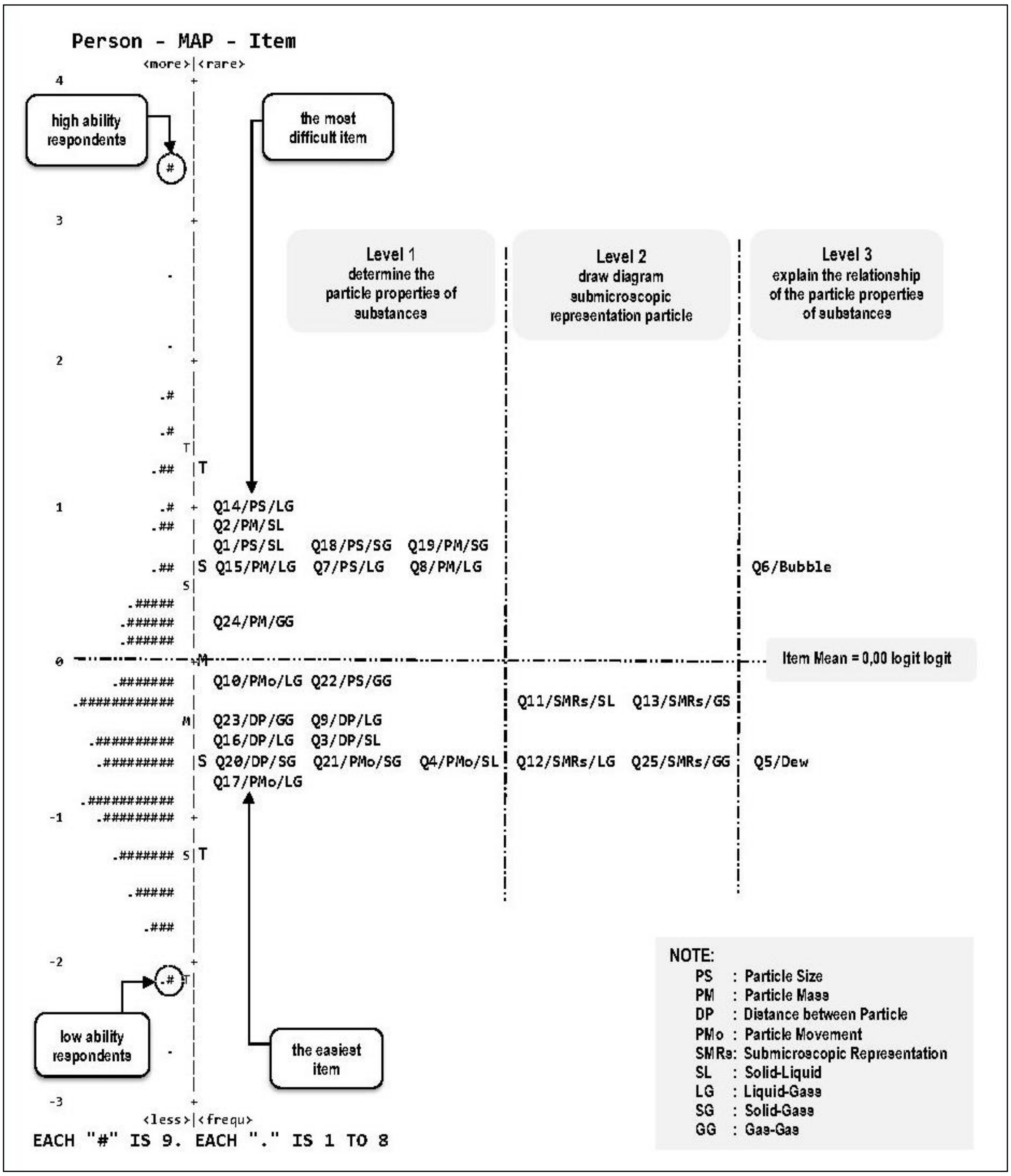
Item	Measure	INFIT		OUTFIT		PTMEA Corr
		MNSQ	ZSTD	MNSQ	ZSTD	
Q6/Bubble	.60	1.26	7.0	1.40	7.5	.07
Q2/PM/SL	.88	1.16	3.7	1.27	4.4	.18
Q15/PM/LG	.66	1.12	3.3	1.20	3.8	.22
Q14/PS/LG	.97	1.03	.8	1.18	2.8	.20
Q18/PS/SG	.71	1.07	1.8	1.15	2.9	.28
Q5/Dew	-.63	1.06	2.6	1.15	3.7	.27
Q7/PS/LG	.66	1.06	1.6	1.14	2.8	.29
Q8/PM/LG	.65	1.07	2.0	1.11	2.1	.28
Q1/PS/SL	.79	1.00	-.1	1.06	1.1	.35
Q24/PM/GG	.25	1.04	1.4	1.04	1.1	.33
Q19/PM/SG	.77	-.3	1.03	.6	.36	.36
Q3/DP/SL	-.44	1.01	.5	1.00	-.1	.34
Q10/PMo/LG	-.07	.98	-.8	.98	-.5	.38
Q13/SMRs/GS	-.24	.98	-1.1	.98	-.6	.38
Q9/DP/LG	-.32	.97	-1.6	.95	-1.6	.39
Q4/PMo/SL	-.66	.96	-2.0	.93	-1.8	.39
Q25/SMRs/GG	-.68	.94	-2.9	.91	-2.4	.41
Q16/DP/LG	-.47	.94	-3.1	.91	-2.8	.42
Q23/DP/GG	-.44	.92	-3.7	.93	-2.1	.43
Q12/SMRs/LG	-.63	.92	-3.8	.87	-3.5	.44
Q21/PMo/SG	-.66	.92	-4.0	.89	-2.9	.43
Q17/PMo/LG	-.71	.91	-4.4	.87	-3.5	.44
Q22/PS/GG	-.07	.90	-4.4	.87	-3.9	.47
Q11/SMRs/SL	-.27	.90	-4.9	.87	-4.0	.47
Q20/DP/SG	-.65	.86	-6.6	.83	-4.6	.49

From the previous Item Statistics, it is generated that all items meet the Outfit MNSA criteria and no negative PTMEA Corr occurs. This means that all items are not deviant, appropriate, and valid. Despite some items do not meet one of the criteria, this by no means decreases the quality of the items. For instance, item (Q6/Bubble, Q2/PM/SL, and Q15/PM/LG) do not meet the criteria of Outfit Z Standard and PTMEA Corr; item (Q1/PS/SL, Q24/PM/GG and Q19/PM/SG) do not meet the criteria of PTMEA Corr; and item (Q25/SMRs/GG, Q16/DP/LG, and Q23/DP/GG) do not meet the criteria of Outfit ZSTD; this is supposedly caused by large size of sample, or N > 500 (Boone et al., 2014).

Wright Map: Person-Map-Item. The third step was to measure the consistency of item difficulty level and student's ability test constructed in Table 2. The higher the item difficulty level, the higher also the student's ability level will result. Information of Wright Map: Person-Map-Item is displayed in Figure 3. The previous Wright map generates that all instrument items encompass almost all the students' ability. The map generates variance from students with very high ability (> 3.0 logit), to those with very low ability (< -2.0 logit) as well. In addition to that, disparity (in which there is no item that is appropriate with the student's ability) was observed within the interval of -3.0 logit up to -0.5 logit and in the interval of +1.0 logit up to +3.7 logit. This signified that the information generated within the interval range was somewhat limited and required further elaboration. On the other hand, the item difficulty level was mostly located in the interval of -1.0 logit up to +1.0 logit; moreover, the items tended to occur in the same difficulty level. The item Q14/PS/LG was the most difficult item with a logit of +0.97, while item Q17/Pmo/LG was the easiest item with logit of -0.71.

Figure 3

Wright Map: Person-Map-Item

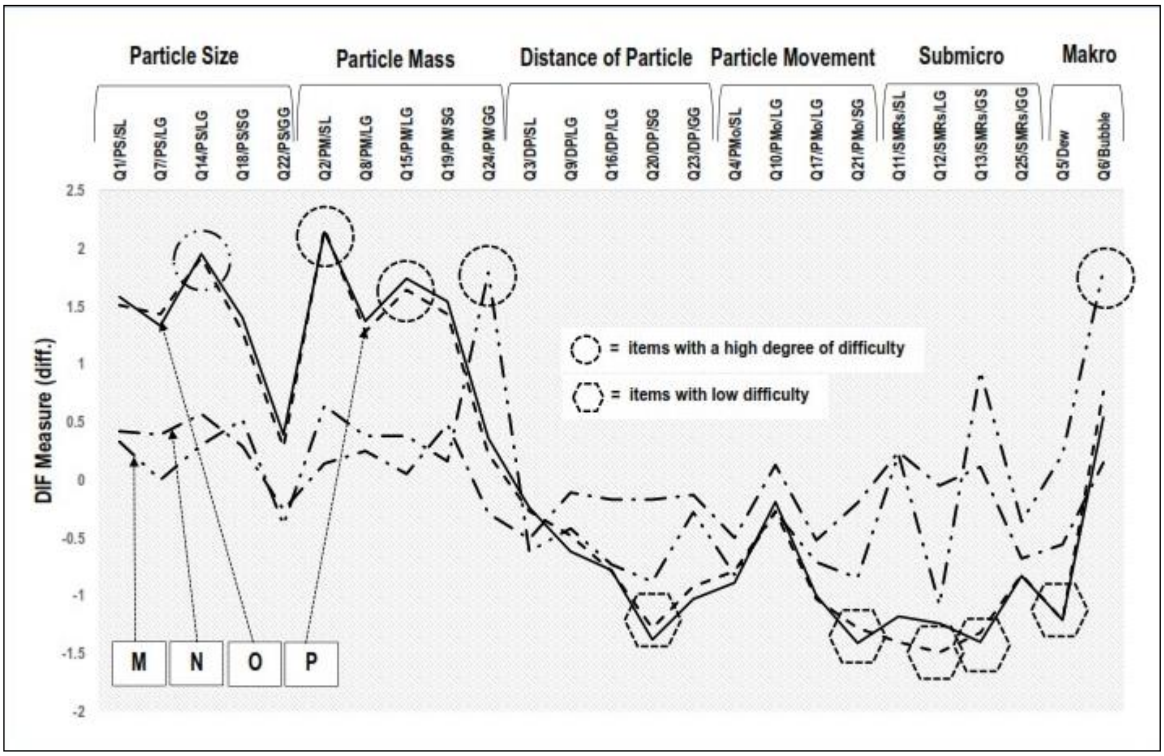


As observed from the differences in item size, some interesting cases were explained as follows: **Firstly**, the items in level 1: Q14/PS/LG (0.97) > Q1/PS/SL (0.79) > Q18/PS/PG (0.71) > Q7/PS/GG (0.66) were instead assumed by the students to possess different difficulty level. The items above, however, were more difficult than item Q6/Bubble in level 3 (0.60). In other words, determining particle size was more difficult than explaining the particle characteristics of matter in the evaporation phenomenon. **Secondly**, the size of item Q5/Dew (-0.63) < item Q6/Bubble; this indicated that it was harder for the students to elaborate on the particle characteristics of matter in the evaporation phenomenon than in condensation phenomenon, despite that both items were in the same level. **Thirdly**, the size of following items: Q2/PM/SL (0.88) > Q19/PM/SG (0.77) > Q15/PM/LG (0.66) > Q8/PM/LG (0.65) > Q24/PM/GG in level 1 was larger compared to that of items Q13/SMRs/GS (-0.24) > Q11/SMRs/SL (-0.27) > Q12/SMRs/LG (-0.63) > Q25/SMRs/GG (-0.68) in level 2. The finding illustrated that it was harder for the students to determine the particle mass than determining the submicrorepresentation (SMRs) diagram in different form changes of matter. The previous cases identified disparity in students' conceptual understanding, signifying that the level of understanding in particle characteristics of the matter is relatively low. Overall, 80% of test item difficulty level is relatively parallel with the measured constructs. By that, the test possesses good construct validity (Blanc & Rojas, 2018; Lu & Bi, 2016; Neumann et al., 2011).

Disparity in Conceptual Understanding Level

The next step was the measurement of disparity of students' conceptual understanding in the focused topic based on educational level by Differential Item Functioning (DIF).

Figure 4
Person DIF plot based on educational level



Note: M = X Class students, N = XI Class students, O = XII Class students, and P = University students from chemistry department

Figure 4 of DIF plot based on students' educational level depicts that ten items are identified to possess significant disparity. **Firstly**, five curves approaching the upper limit are items with a high difficulty level (Q14/PS/LG, Q2/PM/SL, Q15/PM/SG, Q24/PM/GG and Q6/Bubble); while five curves approaching the lower limit are items with a low difficulty level (Q20/DP/SG, Q21/PMo/SG, Q12/SMRs/LG, Q13/SMRs/GS, and Q5/Dew). **Secondly**, the item Q14/PS/LG (particle size in form change of liquid-gas), Q2/PM/SL (particle mass in form change of solid-liquid), and Q15/PM/SG (particle mass in change form of solid-gas) were deemed very hard by the students of XII class and the university students compared to students in X and XI class. **Thirdly**, the research discovered different results for item Q24/PM/GG and Q6/Bubble. The item Q24/PM/GG (particle mass of O2 in larger volume) and Q6/Bubble (constructing elements of air bubbles during boiling process of water) were deemed very hard for X class students compared to students in XI and XII classes, as well as university students. **Fourthly**, the items Q20/DP/SG (distance between particles in form change of solid-gas), Q21/PMo/SG (motion between particles in form change of solid-gas), Q12/SMRs/LG (SMRs diagram of particle in form change of liquid-gas), Q13/SMRs/GS (SMRs diagram of particle in change form of gas-liquid), and Q5/Dew (condensation) were deemed too easy for students in XII class and university students compared to the students in X and XI classes.

Pattern of Conceptual Understanding and Preconception

The analysis of the pattern of conceptual understanding and preconception employed an option probability curve test (Boone et al., 2014; Linacre, 2012). The option probability curve aims to display the probability of picking every answer choice to elaborate on the performance level of all students in the measured items (Herrmann-Abell & DeBoer, 2011). The test relied on the principle that the curve of the correct answer will rise along with the decrease of the curve of distractor choices (Boone et al., 2014; Haladyna, 2004). For items that are influenced by distractor options, the curve produced tends to be non-parallel with the traditional monotonous item behavior (Sadler, 1998), for this reason, each answer choice was analyzed separately.

The instrument provides four answer choices, thus resulting in four curves. Each curve displays the students' comprehension. Students with low ability tended to pick distractor choice, while students whose high ability were more likely to prefer other preconceptions (Herrmann-Abell & DeBoer, 2011; Perera et al., 2018). Below is the elaboration of the pattern of students' conceptual understanding and preconception based on four option probability curves.

Figure 5

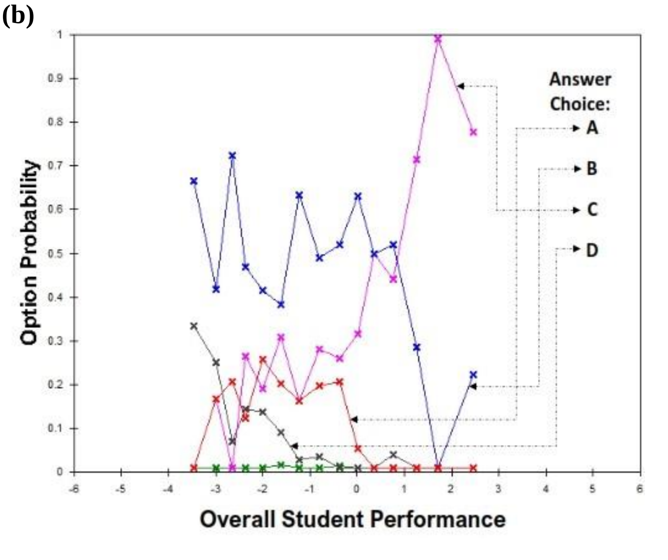
(a) sample of item Q2/PM/SL, (b) option probability curve

(a)

When some ice cubes in the glass melts, some other ice cubes are seen floating on the water surface. How is the comparison between mass of one particle of ice and one particle of water?

A. Mass of one particle of ice is bigger than mass of one particle of water.
B. mass of one particle of ice is smaller than mass of one particle of water.
C. mass of one particle of ice is similar with mass of one particle of water.
D Other answer...

(b)



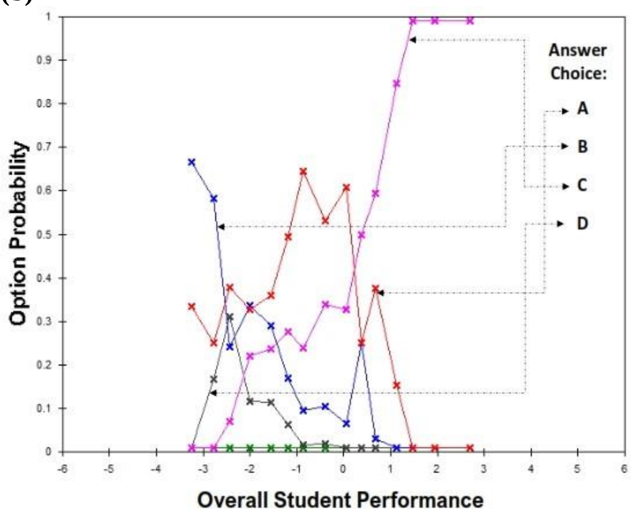
The first example, i.e., the item Q2/PM/SL (0.88), is shown in Figure 5(a). The item measures students' capability in determining particle size in form change from solid to liquid. The option probability curve is displayed in Figure 5(b). Students with the low ability (< 0.5 logit) tended to pick distractor choice B (mass of one particle of ice is smaller than the mass of one particle of water) or A (mass of one particle of ice is bigger than the mass of one particle of water). In addition, students with very low ability (< -1.0 logit) tended to pick D (other answers). Some students with relatively low ability (> -2.5 logit), however, picked the right answer C (mass of one particle of ice is similar to the mass of one particle of water). One can predict the response pattern of students with low ability, as the distractors A, B, and D contain third preconceptions in level 1 (see Table 2). The students possess the knowledge that mass of particle of matter can change into larger or smaller size by observing the matter's change of form. It is interesting to note that there are students with the high ability (>2.0) who picked B; this indicates the presence of resistant preconception.

Figure 6
(a) item Q8/PM/LG; (b) option probability curve

(a)
During boiling process of water, a change of form of water from liquid to gas occurs. How is the comparison between mass of one particle of water and one particle of water vapor?

A. Mass of one water particle is bigger than mass of one vapor particle
B. mass of one water particle is smaller than mass of one vapor particle.
C. mass of one water particle is similar with mass of one vapor particle.
D Other answers...

(b)



The second sample or item Q8/PM/LG (0.65) is shown in figure 6(a) as the item to measure students' ability to determine the mass of the particle in form change of liquid-gas. The option probability curve is displayed in Figure 6(b). The curve of distractor B (mass of one water particle is smaller than the mass of one vapor particle) is chosen by students with low ability (< -2.0 logit), while the curve of choice A (mass of one water particle is bigger than mass of one vapor particle) was chosen by students with ability in a range of -3.5 to 1.5 logit. The correct answer, option C (mass of one water particle is similar to the mass of one vapor particle), was chosen by students with ability in > -2.5 logit. As highlighted in the table, the decline of the curve of distractor A is followed by the increase of curve of right answer C; both curves intersect in the level of 1.0 logit. The shape of curve A indicates the presence of resistant preconception type-three in level 1.

It depicts that the particular item response pattern that signifies students' conceptual understanding patterns in the given level. Moreover, the curve shape of distractors A and B in the items Q2/PM/SL and Q8/PM/SL tend to have an identical pattern. The finding indicated that students with either low or high ability had consistent preconceptions that the mass of the particle can change into larger or smaller in size along with the change in matter form.

Third sample, i.e., item Q5/Dew (-0.63), as shown in Figure 7(a), measures the students' ability in elaborating characteristics of a particle in condensation phenomenon. The option probability curve is displayed in Figure 7(b). Students with low ability (< 1.0 logit) tended to pick distractor A (water drops come from liquid of melting ice that breaks through the glass wall) and option D (other answers). Some students with high ability (> 1.0 logit) also picked distractor B (water drops are the result of the reaction between ice and air nearby the glass). The shape of curve B is wavy and non-linear, even in the interval of 2.0 to 4.0 logit, it can reach option probability value up to 1.0 logit. This is regarded as a deviation from the right answer C (water drops come from condensing water vapor nearby the glass). A worth note, however, is to consider in the unstable, wavy shape of curve C. This indicated the students' inconsistency (particularly those with high ability) in comprehending the concept of condensation. This confirmed that students had their own preconception regarding concept of condensation.

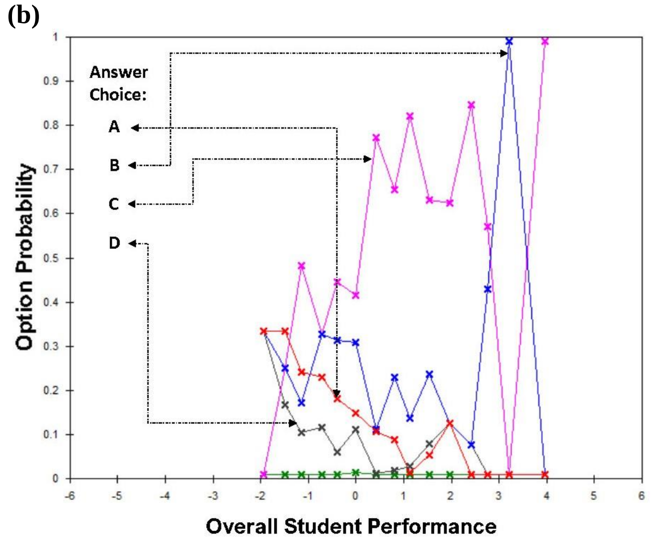
Figure 7

(a) item Q5/Dew; (b) option probability curve

(a)
In a glass filled with ice, you can see water drops at the glass. According to you, where do the water drops come from?

A. Water drops come from melting ice that penetrates the glass wall
B. water drops are the result of reaction between ice and air nearby the glass
C. Water drops come from condensing water vapor nearby the glass
D. Other answer...

(b)

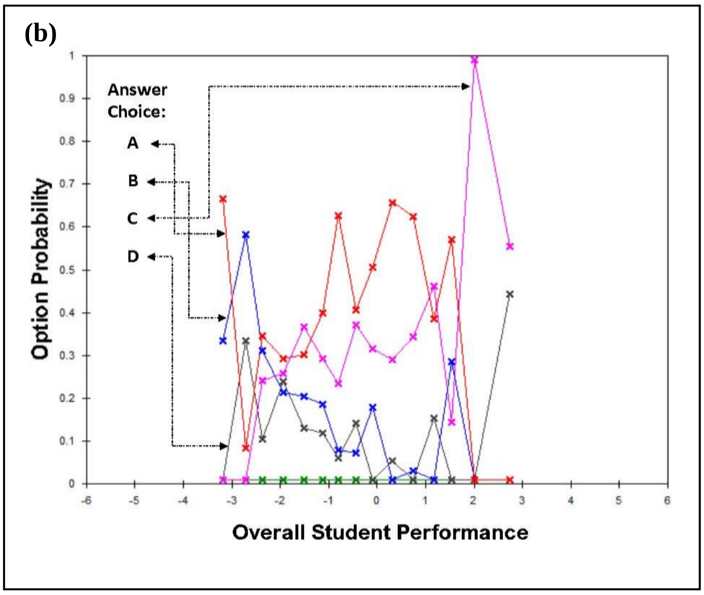


The fourth sample or item Q6/Bubble, as shown in Figure 8(a), measures the students' ability in elaborating characteristics of a particle in the evaporation phenomenon. The option probability curve is displayed in Figure 8(b). The distractor A (air bubbles are Hydrogen and Oxygen particles) was dominantly chosen by students whose ability in a range between -3.0 to 2.0 logit. Moreover, the distractor B (air bubbles are Hydrogen and Oxygen particles) was mostly selected by students whose ability in a range between -3.0 to 0.5 logit. The form of curve A and B were picked by students with low ability was predictable. The curve of right answer C (air bubbles are water molecules), however, shows interesting hint; in the interval range of -2.5 to 3.0 logit, the tip of the curve shows an up-and-down pattern. Moreover, in the level of 1.5 logit, the curve shape of distractors A and B shows a decline pattern, while that of curve C tends to increase. Another finding worth noting was that the curve D (other answers) was picked by some students with high ability (> 2.0 logit). This indicated that particular students had their own preconceptions regarding the evaporation concept.

Figure 8
(a) item Q6/Bubble; (b) option probability curve

(a)
In a container filled with boiling water, you can see air bubbles on the top of it. According to you, what are the composing elements of the air bubbles?

A. Hydrogen and Oxygen particles
B. Air that is dissolved in water
C. Water molecules
D. Other answer...



Discussion

The research results indicated that the instruments had good effectiveness, met the requisites of person and item reliability, and showed good construct validity. When applied in evaluating students' conceptual understanding, it was found that: Firstly, almost all students with high ability faced difficulty in understanding the concept of particle size and mass in level 1. The same students found it relatively easy in determining SMRs diagram of particle structure in level 2 or determining the concept of particle regarding evaporation and condensation phenomena in level 3. Secondly, the information of the response pattern of students with high ability was quite consistent, repetitive, and systematic in particular items. This indicates the presence of permanent and latent preconceptions. The analysis of the option probability curve of item Q2/PM/SL (0.88), Q8/PM/LG (0.65), Q5/Dew (-0.63) and Q6/Bubble (0.60) indicates that the approach of item response pattern is able to explore in detail and comprehensively regarding students' conceptual understanding and preconception.

Sequences of verification conducted that involves Rasch model approach shows detailed, accurate, and quantifiable results since the approach integrates development procedure of diagnostic and summative instruments. Several samples of preconception, e.g., item Q2/PM/SL (0.88) and Q8/PM/LG (0.65) indicate that distractor options are potential to be elaborated further in order to investigate tendency of preconception by the students. In addition, it also provides information regarding main idea unknown to the students and their degree of misunderstanding.

The approach employed in this research is an effective illustration to help teacher in evaluating the learning process as well as the students' learning progress. This is due to the integration of qualitative item development procedure and quantitative data analysis, allowing the teachers to explore in-depth on the students' understanding, concepts the students understand and/or do not **understand, and misconception.** Such findings echo Herrmann-Abell & Deboer (2016) that the integration of Rasch model analysis and probability curve is applicable to diagnose how the students' misconception turns into their overall conceptual understanding. Such an attempt is quite hard to conduct by implementing a conventional approach due to the interdependence of person and item. Rasch model, on the other hand, is able to tackle such interdependence, in which the item and the test difficulty remain invariant and not dependent on which sample that is involved in the initial validation. This signifies that the instrument's items have met the unidimensionality and local independence requirements (Jin et al., 2019; Testa et al., 2019; Wei et al., 2012).

Overall, the research indicated empirical evidence that supported findings by Hoe & Subramaniam (2016); Lu & Bi (2016); Rogat et al., (2011), that students had distinctive preconception as a result of a learning process they experienced. Such preconception was regarded as the inhibitor to the development process of students' conceptual understanding (Soeharto et al., 2019). In this research, students' preconception was found to be repetitive and systematic in each education level. It signifies that the intervention to change students' preconceptions was difficult to conduct by the conventional learning method. A strategic and meaningful learning method is therefore essential to remove students' incorrect preconceptions and develop scientifically correct conceptual understanding. That being said, teachers are demanded to acquire detailed information on the forms and characteristics of students' preconceptions. In conclusion, the item response pattern analysis was an efficient and effective means to acquire such information. The information on students' preconception is important as the basis to develop appropriate and measurable instructional design in solving the students' misconception. This is in line with the previous researches, arguing that the quality of learning progress is highly dependent on the students' learning process and learning experience (Duschl et al., 2011; Park et al, 2017; Wilson, 2009).

Conclusions

The measuring instrument developed performed well in its validity and reliability, thus, it is deemed applicable in measuring students' conceptual understanding and preconception in elaborating particle characteristics of matter. During the implementation of the instruments, the research finds out that: 1) almost all students with high ability face difficulty in understanding the concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, as well as determining the concept of particle regarding evaporation and condensation phenomena in level 3.

2) There is a significant disparity between students' conceptual understanding based on their educational level. 3) In certain cases, it is found that the distractor item response pattern by high-ability students tends to be consistent, indicating a certain tendency of resistant preconception pattern.

The development of diagnostic instruments with Rasch model approach is deemed as the literacy process for practitioners and researchers in Indonesia. The result indicates that there is no single item that is parallel with both the highest ability and lowest ability students. This calls for further elaboration in order to improve the instrument items' quality. Moreover, an anomaly is found that students with high ability (> 1.0 logit) tend to pick distractor choices. This urges further studies to investigate structured comprehension problems. The research regards that further analysis that integrates conceptual understanding level and items designed in a gradual manner is required to define the characteristics of the students' alternative conception and to measure their learning progress. Echoing this notion, one must integrate the item design and basic principles of chemistry as a reference for further researchers and educational practitioners to implement the same approach conducted in the present research. On top of that, despite not focused on discussing matters regarding students' learning progress individually, the instrument is expected to be beneficial for the teachers to diagnose students' conception in developing an effective and meaningful learning experience.

Acknowledgments

This research is funded by the Directorate of Research and Community Outreach/ (Direktorat Riset dan Pengabdian Masyarakat or DRPM), Directorate General of Research and Development (Dirjen Riset dan Pengembangan) of Ministry of Research and Technology of the Republic of Indonesia through Penugasan Penelitian Dasar Unggulan Perguruan Tinggi Grant Program, Research and Community Outreach Institute (Lembaga Penelitian dan Pengabdian Masyarakat) of Universitas Negeri Gorontalo in 2020.

References

Adadan, E., & Savasci, F. (2012). An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, *34*(4), 513–544. https://doi.org/10.1080/09500693.2011.636084

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

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

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

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

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

Blanc, A., & Rojas, A. J. (2018). Use of rasch person-item maps to validate a theoretical model for measuring attitudes toward sexual behaviors. *PLoS ONE*, *13*(8), 1–17. https://doi.org/10.1371/journal.pone.0202551

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

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

Boz, Y. (2006). Turkish pupils' conceptions of the particulate nature of matter. *Journal of Science Education and Technology*, *15*(2), 203–213. https://doi.org/10.1007/s10956-006-9003-9

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, *8*(3), 293–307.

Cheng, M. M. W. (2018). Students' visualisation of chemical reactions-insights into the particle model and the atomic model. *Chemistry Education Research and Practice*, *19*(1), 227–239. https://doi.org/10.1039/c6rp00235h

Davidowitz, B., Chittleborough, G., & Murray, E. (2010). Student-generated submicro diagrams: A useful tool for teaching and learning chemical equations and stoichiometry. *Chemistry Education Research and Practice*, *11*(3), 154–164. https://doi.org/10.1039/c005464j

Devetak, I., & Glažar, S. A. (2010). The influence of 16-year-old students' gender, mental abilities, and motivation on their reading and drawing submicrorepresentations achievements. *International Journal of Science Education*, *32*(12), 1561–1593. https://doi.org/10.1080/09500690903150609

Devetak, I., Urbancic, M., Grm, K. S. W., Krnel, D., & Glazar, S. A. (2004). Submicroscopic representations as a tool for evaluating students' chemical misconception. *Acta Chimica Slovenica*, *51*(December 2004), 799–814. https://www.researchgate.net/publication/286849527_Submicroscopic_representations_as_a_tool_for_evaluating_students'_chemical_conceptions

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, *47*(2), 123–182. https://doi.org/10.1080/03057267.2011.604476

Femintasari, V., Effendy, & Munzil. (2015). The effectiveness of two-tier multiple choice test and multiple choice test followed with interview in identifying misconception of students with different scientific reasoning skills. *Jurnal Ilmu Pendidikan*, *21*(2), 192–197. http://journal.um.ac.id/index.php/jip/article/view/8376/4020

Gabel, D. L. (1993). Use of the particle nature of matter in developing conceptual understanding. *Journal of Chemical Education*, *70*(3), 193. https://doi.org/10.1021/ed070p193

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

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

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

Herrmann-Abell, C. F., & Deboer, G. E. (2016). Using Rasch modeling and option probability curves to diagnose students' misconceptions. *American Eduacational Research Assossiation*, 1–12.

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

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

Jin, H., Mikeska, J. N., Hokayem, H., & Mavronikolas, E. (2019). Toward coherence in curriculum, instruction, and assessment: A review of learning progression literature. *Science Education*, *103*(5), 1206–1234. https://doi.org/10.1002/sce.21525

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

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

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

Kapici, H. Ö., & Akcay, H. (2016). Particulate nature of matter misconceptions held by middle and high school students in Turkey. *European Journal of Education Studies*, *2*(8), 43–58. https://doi.org/10.5281/zenodo.163547

Kim, J., & Wilson, M. (2019). Polytomous item explanatory item response theory models. *Educational and Psychological Measurement*. https://doi.org/10.1177/0013164419892667

Kind, V. (2004). *Beyond appearances: Students' misconceptions about basic chemical ideas* (2nd Edition). Royal Society of Chemistry. https://www.researchgate.net/publication/228799159_Beyond_Appearances_Students'_Misconceptions_About_Basic_Chemical_Ideas

Laliyo, L. A. R., Botutihe, D. N., & Panigoro, C. (2019). The development of a two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12

Linacre, J. M. (2012). *A user's guide to W I N S T E P S ® M I N I S T E P Rasch-model computer program: Program manual 3.75.0*. https://doi.org/ISBN 0-941938-03-4

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

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

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, *34*(15), 2317–2341. https://doi.org/10.1080/09500693.2012.684433

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

https://doi.org/10.1039/C2RP00015F
Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297
Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*(9), 825–838. https://doi.org/10.1002/tea.3660200905
Ozalp, D., & Kahvecib, A. (2015). Diagnostic assessment of student misconceptions about the particulate nature of matter from ontological perspective. *Chemical Education Research and Practice*, *16*, 619–639. https://doi.org/10.1039/C5RP00096C
Ozmen, H. (2011). Turkish primary students´ conceptions about the particulate nature of matter. *International Journal of Environment & Science Education*, *6*(1), 99–121. https://www.researchgate.net/publication/265438996_Turkish_primary_students'_conceptions_about_the_particulate_nature_of_matter
Park, M., Liu, X., & Waight, N. (2017). Development of the connected chemistry as formative assessment pedagogy for high school chemistry teaching. *Journal of Chemical Education*, *94*(3), 273–281. https://doi.org/10.1021/acs.jchemed.6b00299
Perera, C. J., Sumintono, B., & Jiang, N. (2018). The psychometric validation of the principal practices questionnaire based on Item Response Theory. *International Online Journal of Educational Leadership*, *2*(1), 21–38. https://doi.org/10.22452/iojel.vol2no1.3
Renström, L., Andersson, B., & Marton, F. (1990). Students' conceptions of matter. *Journal of Educational Psychology*, *82*(3), 555–569. https://doi.org/10.1037/0022-0663.82.3.555
Rogat, A., Anderson, C., Foster, J., Goldberg, F., Hicks, J., Kanter, D., Krajcik, J., Lehrer, R., Reiser, B., & Wiser, M. (2011). *Developing learning progression in support of the new science standards: A RAPID workshop series*, *New York, NY.* (4), 163. https://doi.org/10.12698/cpre.2011.lprapid
Sadler, P. M. (1998). Psychometric models for student-conceptions in science: Reconciling qualitative studies and distractor-driver assessment instruments. *Journal of Research in Science Teaching*, *35*(3), 265–296.
Sadler, P. M. (1999). The relevance of multiple-choice testing in assessing science understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 251–274). https://zodml.org/sites/default/files/%5BJoel_J._Mintzes%2C_James_H._Wandersee%2C_Joseph_D._No_0.pdf
Slapničar, M., Devetak, I., Glažar, A. S., & Pavlin, J. (2017). Identification of the understanding of the states of matter of water and air among Slovenian students aged 12, 14 and 16 years through solving authentic tasks. *Journal of Baltic Science Education*, *16*(3), 308–323.
Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649
Stojanovska, M. I., Soptrajanov, B. T., & Petrusevski, V. M. (2012). Addressing misconceptions about the particulate nature of matter among secondary-school and high-school students in the Republic of Macedonia. *Creative Education*, *03*(05), 619–631. https://doi.org/10.4236/ce.2012.35091
Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)*, (October 2017). https://doi.org/10.2991/icei-17.2018.11
Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial* [Application of Rasch Model in Social Sciences Research]. Tri Komunikata. https://www.researchgate.net/publication/268688933%0AAplikasi
Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan* [Application of Rasch Model in Education Assessment]. Tim Komunikata. https://www.researchgate.net/publication/282673464%0AAplikasi
Taber, K. S. (2015). Alternative Conceptions/Frameworks/Misconceptions. In G. Richard (Ed.), *Encyclopedia of Science Education* (pp. 37–41). https://doi.org/10.1007/978-94-007-6165-0
Testa, I., Capasso, G., Colantonio, A., Galano, S., Marzoli, I., Scotti di Uccio, U., Trani., F & Zappia, A. (2019). Development and validation of a university students' progression in learning quantum mechanics through exploratory factor analysis and Rasch analysis. *International Journal of Science Education*, *41*(3), 388–417. https://doi.org/10.1080/09500693.2018.1556414
Tóth, Z., & Kiss, E. (2006). Using particulate drawings to study 13-17 year olds' understanding of

physical and chemical composition of matter as well as the state of matter. *Practice and Theory in Systems of Education*, *1*(1), 109–125.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169. https://doi.org/10.1080/0950069880100204

Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, *4*(6), 626–631. http://www.academicjournals.org/SRE

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335–345. https://doi.org/10.1021/ed100852t

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

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

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

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

Yaşar, I. Z., İnce, E., & Kırbaşlar, F. G. (2014). 7. Class Science and Technology Course "Structure of Atom" Subject Readiness Improvement Test. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2014.09.259

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

Accept without revision	
Accept with minor revisions	X
Accept: with moderate revisions	
Pre-Accept: major revisions and re-evaluation	
Reject: Rework and re-submit	
Reject: do NOT re-submit	

JOURNAL OF BALTIC SCIENCE EDUCATION
ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

ANALYTIC APPROACH OF RESPONSE PATTERN OF DIAGNOSTIC TEST ITEMS IN EVALUATING STUDENTS' CONCEPTUAL UNDERSTANDING OF CHARACTERISTICS OF PARTICLE OF MATTER

Lukman Abdul Rauf Laliyo,
Julhim S. Tangio,
Bambang Sumintono,
Mohamad Jahja,
Citra Panigoro

Abstract. *This research aimed to evaluate the students' conceptual understanding and to diagnose the students' preconceptions in elaborating the particle characteristics of matter by development of diagnostic instrument as well as Rasch model response pattern analysis approach. Data were acquired by 25 multiple-choice written test items distributed to 987 students in North Sulawesi, Indonesia. Analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding; and 3) diagnosis of students' preconception by estimation of item response pattern. The result generated information on the diagnostic and summative measurement on students' conceptual understanding in elaborating the topic; information also acts as empirical evidence on the measurement's reliability and validity. Moreover, the result discovered a significant disparity between students' conceptual understanding based on their educational level. It was found that the distractor item response pattern tended to be consistent, indicating a certain tendency of resistant preconception pattern. The findings are expected to be a recommendation for future researchers and educational practitioners that integrate diagnostic and summative measurement with Rasch model in evaluating conceptual understanding and diagnosing misconception.*
Keywords: *conceptual understanding, item response, particle of matter, Rasch model*

Lukman Abdul Rauf Laliyo,
Julhim S. Tangio
State University of Gorontalo, Indonesia
Bambang Sumintono
University of Malaya, Malaysia
Mohamad Jahja, Citra Panigoro
State University of Gorontalo, Indonesia

Introduction

Central to the notion of learning about characteristics of a particle of matter is the process of developing an understanding on abstract concepts (Johnstone, 1991) without directly interacting with the object/fact (Stojanovska et al., 2012); therefore it is considered a difficult subject for the students to learn. Echoing this, the disparity in understanding is almost inevitable (Kapici & Akcay, 2016) since different students may develop their own distinctive way of understanding a concept (Yildirir & Demirkol, 2018). The idea is also coined by experts as misconception (Johnstone, 2006, 2010; Taber, 2015), or alternative framework and preconception (Lu & Bi, 2016). The experts have discovered that students always have their own preconception that is not in line with scientific concepts (Alamina & Etokeren, 2018; Yaşar et al., 2014); therefore, one needs to conduct identification and improvement on the conceptual learning (Allen, 2014; Soeharto et al., 2019).

In diagnosing preconceptions, several researchers have developed diagnostic instruments in different mechanisms (McClary & Bretz, 2012), i.e., conceptual map, essay test, interview, essay test with interview, or multiple-choice test (Femintasari et al., 2015). Two-step multiple choice diagnostic test (Adadan & Savasci, 2012; Chandrasegaran et al., 2007; Treagust, 1988; Tüysüz, 2009) is preferred due to its ability to diagnose preconception and describe the underlying reasons. The instrument is indeed considered qualitatively effective in elaborating differences in students' thought processes; however, it does not provide summative measurement features due to lack of internal consistency and the instrument's unidimensionality (Lu & Bi, 2016). In addition to that, the measurement conclusion generated is considered weak due to extracted from analysis on the raw score (Sumintono, 2018)

https://doi.org/10.33225/jbse/20.19.824

Studies on preconception have found that the concept is somewhat resistant. In the early 2000s, it is discovered that students' preconceptions persisted even when they already undergo formal education experience (Hoe & Subramaniam, 2016). Preconception can also change along with the development of students' conceptual understanding; it also varies in different levels of understanding (Aktan, 2013). If one conducts a two-step test and raw score analysis approach to diagnose resistant preconception, the result generated will only provide limited feedback information (Sumintono, 2018) due to the instrument's limitation in measuring students' conceptual understanding. Instead of supporting, the information will only make it harder for teachers to implement proper instructional decisions (Wilson, 2008).

During the middle of the 2000s, the Rasch model analysis was commonly used in studies of chemistry education (Herrmann-Abell & DeBoer, 2011; Liu, 2012; Wein et al., 2012). The approach provides a testing apparatus that integrates diagnostic and summative measurement. Recently, this approach is used to develop formative assessment with the intention to conduct learning construction mapping, e.g., measuring the students' way of constructing their understanding process (Hadenfeldt et al., 2013). It is worth to note, however, that there are studies that integrate diagnostic and summative measurement with a different approach (Hoe & Subramaniam, 2016); despite that, trends in chemistry education studies highlight that diagnostic-summative measurement by Rasch model analysis is more common to be carried out (Laliyo et al., 2019; Lu & Bi, 2016).

Research Problem

The characteristics of a particle of matter is a fundamental concept in chemistry, usually taught in middle education level. Adequate comprehension regarding the particle characteristics of matter both in macroscopic and microscopic level is essential as the knowledge basis in understanding more advanced topics such as the concept of atoms and molecules as the submicroscopic component that is invisible to plain eyesight but exists in all real-world phenomena (Cheng, 2018; Ozmen, 2011; Yildirir & Demirkol, 2018). The fact signifies the relevance and reasoning of complexity in chemistry learning that is considered difficult for both students and teachers to conduct (Alamina & Etokeren, 2018). In simpler terms, to ensure that the chemistry learning is conducted effectively, one requires to nurture students' comprehensive understanding regarding particle characteristics of matter and its change of state.

To evaluate the students' conceptual understanding on the aforementioned topic, one also needs to measure the students' capability in interpreting particle state during change process of a matter's form (Alamina & Etokeren, 2018; Barbera, 2013; Boz, 2006; Cheng, 2018; Gabel, 1993; Hadenfeldt et al., 2013; Kapici & Akcay, 2016; Kind, 2004; Naah & Sanger, 2012; Ozalp & Kahvecib, 2015; Ozmen, 2011; Renström et al., 1990; Slapničar et al., 2017; Stojanovska et al., 2012; Yildirir & Demirkol, 2018). Research studies on particle characteristics and changes of matter generally employ diagnostic instruments in the form of essay tests and/or essays followed by interview; the instruments are further analyzed based on raw score results. The approach is considered inefficient and somewhat lacked accuracy in measuring students' conceptual understanding and misconception pattern. Despite its ineffectiveness, the conventional method is used by most teachers in Indonesia to measure and determine students' learning progress. The teachers argue that measuring the students' raw score is effective in determining how far the students have progressed in the learning process. The students' raw score is regarded by many as an early premature indication regarding the measured variable and is not eligible to be the final measurement indicator due to its temporary nature. In addition to that, regarding the decision-making process, the raw score contains only limited information for it to be treated as reference (He et al., 2016; Sumintono & Widhiarso, 2015)

Research Focus

The research focuses on developing a diagnostic instrument that integrates measurement of conceptual understanding and diagnosis of students' preconceptions regarding the aforementioned topic by the approach of Rasch model item response pattern analysis. The analysis employs different test apparatuses to provide extensive information for practitioners and researchers in science education in evaluating students' learning progress in different topics.

https://doi.org/10.33225/jbse/20.19.824

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

Research Questions

This research aimed to figure out the following questions: 1) How is the effectiveness of measurement instrument to evaluate the students' conceptual understanding and diagnose their preconceptions on the characteristics of a particle of matter? 2) Is there any significant difference between students in elaborating on the aforementioned topic based on their educational level? 3) How is the pattern of students' conceptual understanding and preconception regarding the topic?

Research Methodology

General Background

The descriptive-quantitative research employed a non-experimental approach, in which the students' conceptual understanding in explaining the characteristics of a particle of matter was treated as the measurable variable. Prior to conducting the research, it was ensured that the students already experience formal learning of the aforementioned topic. The researchers did not conduct any intervention on the learning process or the learning material. In other words, no treatment was implemented to the students for them to be able to answer all test items in the measurement instrument.

The data collection step was implemented for four months in the even semester of the 2019-2020 academic year; the process was conducted after obtaining approval from the Government of Province of Gorontalo and heads of universities in the Northern part of Sulawesi, Indonesia. Moreover, the schools' and parents' approval was obtained in cooperation with the school committee. The school administrators were willing to facilitate the data collection process that adjusted with the schedule.

Respondents

The respondents were 987 people consisting of students of eleventh grade from eight lower-secondary schools well as university students of the chemistry department in Northern Sulawesi, Indonesia. The distribution of respondents is displayed in Table 1 below.

Table 1
Demographic profile of respondents (N=947)

Demography	Code	Respondents	Percentage (%)
Gender			
Male	M	320	67.68
Female	F	667	32.42
Education level			
X Class students	M	168	17.02
XI Class students	N	473	47.92
XII Class students	O	186	18.84
University students from the chemistry department	P	160	16.21

The respondents were chosen randomly and have voluntarily agreed to participate in the research. In addition, they received no learning treatment and other special treatments that allow them to complete the measurement instrument. Students were asked to write down their responses in the answer sheet; the process was supervised by teachers in the respective schools and lecturers in the respective university. All students were instructed to answer all questions in the instruments within 45 minutes. All instrument sheets and answer sheets were collected by the researchers shortly after the session ended; it was ensured that the numbers of instruments matched the numbers of participants. For the certainty in ethical consideration, permission was obtained from the school administration after coordinating with students' parents through the school committee. This process was conducted before the

https://doi.org/10.33225/jbse/20.19.824

students were invited to participate in research. Permission for the students was obtained from the department leaders of the university, and student written statements. All students were told that the confidentiality of their identity was fully guaranteed, and the results of the study would only be used for research purposes.

Instrument and Procedures Development

The design process refers to a recommendation by Wilson (2005), which consists of four key steps: definition of construct map, item design, result blank, and measurement model.

Phase 1: Definition of construct map. The map offers a substantive definition of measured constructs; the more constructs measured, the constructs' level will vary qualitatively (Wilson, 2009). In simpler words, it aims to develop the students' understanding map to measure the students' progress (Wilson, 2012). The instrument involved variables, i.e., the students conceptual understanding and preconception in elaborating the characteristics of a particle of matter; it was conducted in accordance with the Curriculum Standard of Chemistry Subject in Tenth Grade in Indonesia, as presented in Table 2.

Table 2
Conceptual Understanding Level

Level 3	**The students are able to connect between characteristics of a particle of matter in macroscopic and submicroscopic level**
Phenomenon	Evaporation: item Q6/Bubble
10. Preconception	Air bubble consists of Hydrogen and Oxygen particles
9. Preconception	Air bubble is water-soluble
Phenomenon	Condensation: item Q5/Dew
8. Preconception	Water drops come from melting ice that penetrates the glass wall
7. Preconception	Water drops are the result of the reaction between ice and air nearby the glass
Level 2	**The students are able to determine SMRs diagram of particle structure during a change of form: item Q11/SMRs/SL; Q12/SMRs/LG; Q13/SMRs/GS; Q25/SMRs/GG**
6. Preconception	The SMRs diagram of particle structure follows the physical form of matter
5. Preconception	The SMRs diagram of O2 molecule shape undergoes change as a result of an increase in the volume of the container.
Level 1	**The students are able to determine the characteristics of a particle of matter during the change process of matter's form.**
4. Preconception	The particle size of matter changes into (large/small) as a result of change in matter form: item Q1/PS/SL; Q7/PS/LG; Q14/PS/LG; Q18/PS/SG; Q22/PS/GG
3. Preconception	The particle mass of matter changes into (large/small) due to change in matter form: item Q2/PM/SL; Q8/PM/LG; Q15/PM/LG; Q19/PM/SG; Q24/PM/GG
2. Preconception	Distance between matter particles changes into (faster/slower) due to change in matter form: item Q3/DP/SL; Q9/DP/SL; Q16/DP/LG; Q20/DP/SG; Q23/PM/GG
1. Preconception	Motion between matter particles changes into (dense/loose) due to change in matter form: item item Q4/PMo/SL; Q10/PMo/LG; Q17/PMo/LG; Q21/PMo/SG

Variation in conceptual understanding level illustrates the development process of the students' conceptual understanding. In the first level, the students were asked to determine particle characteristics (size, mass, motion, and distance) in the change process of matter form. In the second level, the students were asked to determine the submicroscopic representation diagram of particle structure. Further, in the third level, the students were asked to connect between characteristics of a particle of matter at the macroscopic and submicroscopic level. In each level, the construct map also features the students' tendency of preconception.

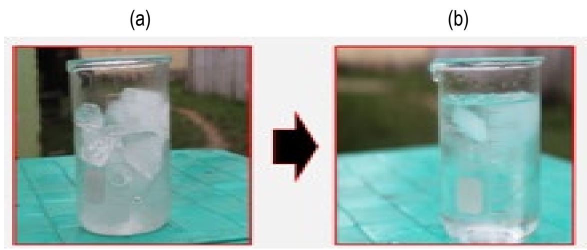
Phase 2: item design and evaluation The phase involved the determination process of items to be used in acquiring evidence of students' construct understanding regarding the construct map (Wilson, 2005). Certain items may have a different extent of effectiveness to measure students' conceptual understanding (Sadler, 1999);

https://doi.org/10.33225/jbse/20.19.824

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

however, multiple choices item is considered more practical and effective (Wilson, 2008). The instrument of concept understanding test of the particle (or TPKP) is adapted from multiple-choice instruments by (Herrmann-Abell & DeBoer, 2011). Each item consists of two distractor answer choices and one open answer choice. The distractor answer choices are designed by referring to the common preconceptions by the students (see Table 2) as logical choices to distract the students from the correct one. The distractors function to emphasize the item diagnostic strength (Sadler, 1998). Some of the items are adopted from previous studies Osborne & Cosgrove (1983), Renström et al., (1990); Devetak et al., (2004); Tóth & Kiss (2006); Davidowitz et al., (2010); Devetak & Glažar (2010); Slapničar et al., (2017) and (Yildirir & Demirkol, 2018).

Figure 1
Sample of item Q1/PS/SL design



Glass (a) contains ice chunks; glass (b) contains melting ice chunks. How is the size of water particle in solid form (ice) compared to that in liquid form?

a. Size of a water particle in solid form > a water particle in liquid form.
b. Size of a water particle in solid form < a water particle in liquid form.
c. Size of a water particle in solid form = a water particle in liquid form.
d. Other answers

Figure 1 displays a sample of item Q1/PS/SL design, in which Q1 is the number of item 1, PS is particle size, and SL is solid-liquid. The item measures student's capability in determining particle size in form change from solid to liquid. The choice A and B are distractors, the correct choice is C, and choice D is for other answers students may fill if the existing answer choices are not in accordance with their initial knowledge. Every correct answer was given mark 1, and wrong answers got 0 mark. Each student only has a slight probability of 0.25 in choosing the right answer. The students will pick what they think the right answer based on their understanding. If the distractor item choice functions well, the students will not be able to predict the correct answer.

Phase 3: design of result blank, i.e., the correlation between construct map and items (Wilson, 2005). This phase aimed to identify whether the answer the students pick correlates with their conceptual understanding; in simpler terms, it was intended to elaborate the conformity between the variable contents being measured. In order to elaborate on the previous aspect, the TPKP instrument was validated by three independent experts and tested to the students to acquire their feedback. The process acquired 25 items of TPKP. Prior to the data collection process, it was ensured that all students had received formal education on the characteristics of a particle of matter and its changes. The students' response towards the instrument was inputted manually by the written answer sheet. The test was supervised by the teachers in school by referring to the agreed permission and duration. Each student was required to finish all test items within the allocated duration of 45 minutes. The instrument sheets were further collected, and checking process was conducted to ensure that the amount of instrument sheet was the same with participating students.

Phase 4: Rasch model analysis approach. The analysis integrates algorithm as a result of probabilistic expectation of item 'i' and student 'n', as: The statement is the probability of student n in item i to result in the correct answer (x = 1); with student ability, ßn, and item difficulty level (Bond & Fox, 2015). The above equation was simplified by inserting logarithm function, into, so that the probability of picking the right answer equals to student's ability subtracted by item difficulty level. The student (person) and item units were considered on the same interval scale and were independent of each other. The students' ability level and item difficulty level were measured in the logarithm unit, namely odds or log that variates from -00 to +00 (Herrmann-Abell & DeBoer, 2011; Sumintono & Widhiarso, 2015). The instrument efficiency, when compared to the item distribution towards item difficult level with distribution of student's ability level, was quantifiable in order to measure the students' conceptual understanding. In addition, the student's understanding level was differentiated based on the item size. The previous steps highlighted the main difference of Rasch model analysis when compared to the raw score-based conventional one; the latter lacks accuracy in evaluating students' ability observed from different item difficulty level (Herrmann-Abell & DeBoer, 2011; Lu & Bi, 2016; Sumintono & Widhiarso, 2015).

https://doi.org/10.33225/jbse/20.19.824

Data Analysis

The research employed WINSTEPS version 3.75 software to convert raw data into interval data (Bond & Fox, 2015; Linacre, 2012). The conversion result acted as the calibration of data on the student's ability level and item difficulty level within the same interval measurement. Moreover, the analysis on diagnostic test items response pattern was conducted in three steps: 1) conversion of raw score to a homogenous interval unit and effectiveness analysis of measurement instruments; 2) measurement of disparity of students' conceptual understanding by Differential Item Functioning (DIF) item test; and 3) diagnosis of students' preconception by estimation of item response pattern through option probability curve test.

Research Results

Effectiveness of Measuring Instruments

Person and Item Reliability. The first step to elaborate on the effectiveness of measuring instruments was by measuring the person and item reliability. This was conducted to gather information to what extent the measurement produces consistent information in displaying latent trait or the unidimensionality of the measured variable (Sumintono & Widhiarso, 2015). The analysis result is presented in the form of a statistical summary (Table 3).

Table 3
Summary of fit statistics

Parameter (*N*)	Measure	INFIT		OUTFIT		Separation	Reliability	*SD*	KR-20
		MNSQ	ZSTD	MNSQ	ZSTD				
Person (987)	-.34	1.00	-.11	1.02	-.1	1.55	.71	.88	.72
Items (25)	.00	1.00	-.75	1.02	-.1	8.18	.99	.60	

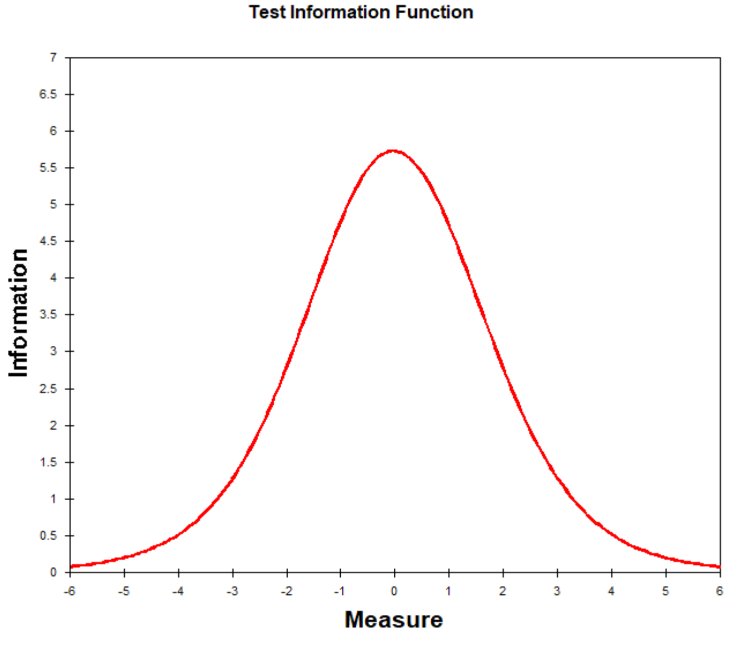
The above table indicates that the person reliability value of 0.71 is equivalent to the person separation index value of 1.55. This is to say that the consistency of students' response towards the test is deemed good. In addition to that, it is generated that the *Cronbach Alpha Coefficient* (KR-20) value is 0.72, signifying good interaction between students and the test. This further indicates strong correlation between the students' response towards the item, in the context that the students' knowledge tends to be non-fragmented, enabling it to be measured (Adams & Wieman, 2011). To the researchers and educational practitioners, such information is essential to prepare for follow-up plans and development of students' ability (Wei et al., 2012). Moreover, the result generated a relatively high value of item separation index of 8.18 that was equivalent to the item reliability value of 0.99. This indicated very good item consistency, or the item was deemed capable of meeting the unidimensionality criteria. In other words, the item performed very good in defining the measured variable. This was confirmed by the infit and outfit value result, in which most of the items were in the acceptable range for the multiple-choice test (Bond & Fox, 2015; Herrmann-Abell & DeBoer, 2011).

Figure 2 displays the graph of measurement information in order to show the measurement reliability. The higher the tip of information function graph, the measurement reliability value is likely to increase. In the intermediate level of students' ability (-3.0 logit up to +3.0 logit), the measurement information is in very high spot. This indicates that the TPKP instrument is capable of producing optimal information to students with an intermediate level of ability. Such a result means that the instrument possesses high measurement reliability (Bond & Fox, 2015; Kim & Wilson, 2019).

https://doi.org/10.33225/jbse/20.19.824

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

Figure 2
Function of Measurement Information



Note: M = X Class students, *N* = XI Class students, O = XII Class students, and P = University students from chemistry department

Validity. The next step was to measure the item validity by Fit item test to ensure that all items fit with the Rasch model. The process was aimed to identify whether or not the test item could measure the aspects that intended to be measured, or test validity (Linacre, 2012; Sumintono & Widhiarso, 2014). The criteria used comprise outfit means-square (MNSQ): 0.5 < y < 1.5; outfit z-standard: -2.0 < Z < + 2.0, as well as point measure correlation (PTMEA Corr). The PTMEA Corr is the correlation between the score of item and person measure that is required to be a positive value and not approaching zero (Bond & Fox, 2015). The PTMEA Corr criteria: 0.4 < x < 0.8. If all three criteria are not met, the item is not good enough and needs further elaboration (Boone et al., 2014). Both Outfit MNSQ and Infit MNSQ were sensitive chi-squares in detecting outlier response pattern. There were two outlier responses: the right response, guessed by the students with low ability in item with high difficulty level, or the wrong response due to the high-ability students' carelessness in items with a low difficulty level. The expected ideal MNSQ value is 1.0. The analysis result on item appropriateness is displayed in Table 4 as follows:

Table 4
Item Statistics: Misfit Order

Item	Measure	INFIT		OUTFIT		PTMEA Corr
		MNSQ	ZSTD	MNSQ	ZSTD	
Q6/Bubble	.60	1.26	7.0	1.40	7.5	.07
Q2/PM/SL	.88	1.16	3.7	1.27	4.4	.18
Q15/PM/LG	.66	1.12	3.3	1.20	3.8	.22
Q14/PS/LG	.97	1.03	.8	1.18	2.8	.20
Q18/PS/SG	.71	1.07	1.8	1.15	2.9	.28
Q5/Dew	-.63	1.06	2.6	1.15	3.7	.27

https://doi.org/10.33225/jbse/20.19.824

Item	Measure	INFIT		OUTFIT		PTMEA Corr
		MNSQ	ZSTD	MNSQ	ZSTD	
Q7/PS/LG	.66	1.06	1.6	1.14	2.8	.29
Q8/PM/LG	.65	1.07	2.0	1.11	2.1	.28
Q1/PS/SL	.79	1.00	-.1	1.06	1.1	.35
Q24/PM/GG	.25	1.04	1.4	1.04	1.1	.33
Q19/PM/SG	.77	-.3	1.03	.6	.36	.36
Q3/DP/SL	-.44	1.01	.5	1.00	-.1	.34
Q10/PMo/LG	-.07	.98	-.8	.98	-.5	.38
Q13/SMRs/GS	-.24	.98	-1.1	.98	-.6	.38
Q9/DP/LG	-.32	.97	-1.6	.95	-1.6	.39
Q4/PMo/SL	-.66	.96	-2.0	.93	-1.8	.39
Q25/SMRs/GG	-.68	.94	-2.9	.91	-2.4	.41
Q16/DP/LG	-.47	.94	-3.1	.91	-2.8	.42
Q23/DP/GG	-.44	.92	-3.7	.93	-2.1	.43
Q12/SMRs/LG	-.63	.92	-3.8	.87	-3.5	.44
Q21/PMo/SG	-.66	.92	-4.0	.89	-2.9	.43
Q17/PMo/LG	-.71	.91	-4.4	.87	-3.5	.44
Q22/PS/GG	-.07	.90	-4.4	.87	-3.9	.47
Q11/SMRs/SL	-.27	.90	-4.9	.87	-4.0	.47
Q20/DP/SG	-.65	.86	-6.6	.83	-4.6	.49

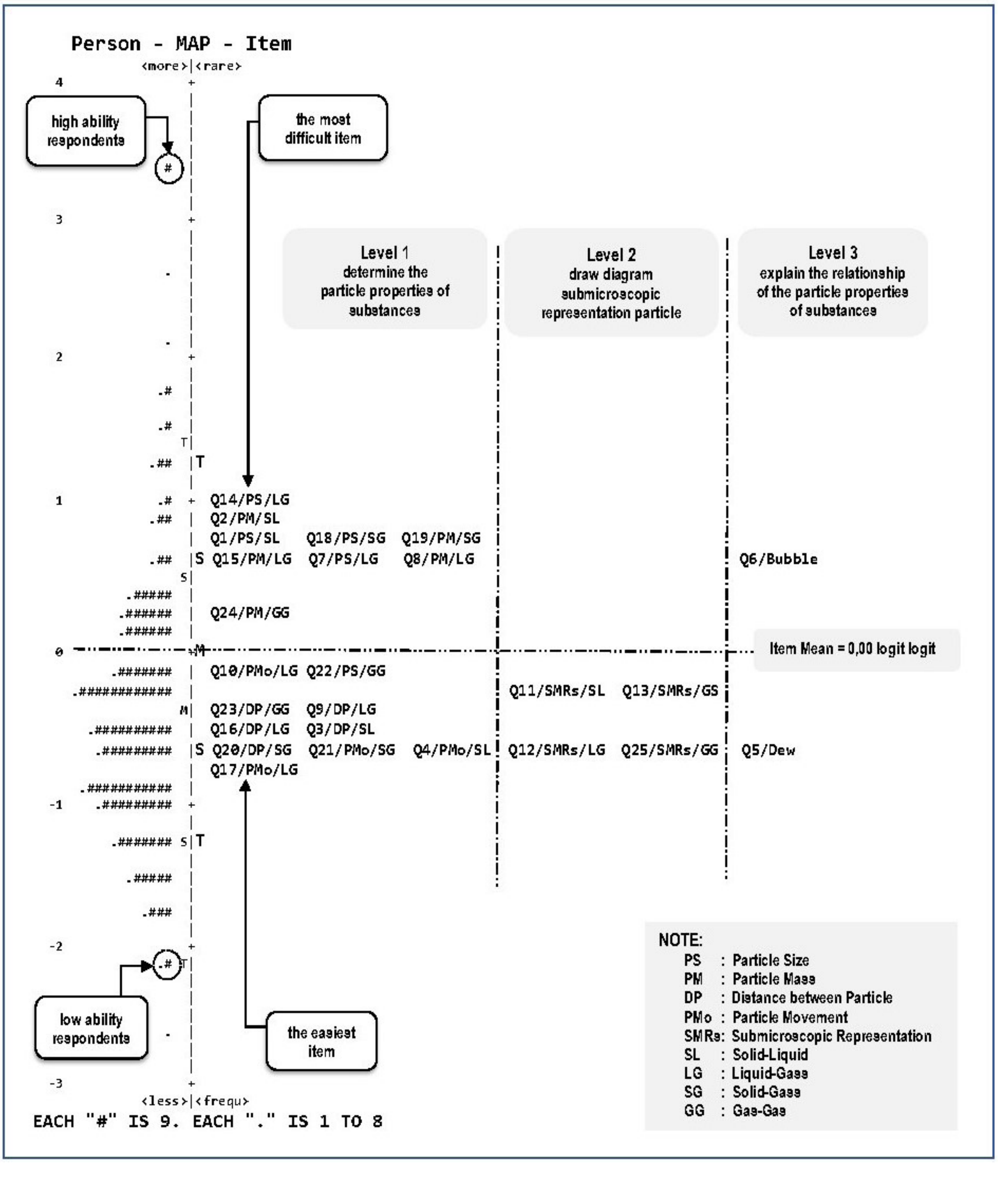
From the previous Item Statistics, it is generated that all items meet the Outfit MNSA criteria and no negative PTMEA Corr occurs. This means that all items are not deviant, appropriate, and valid. Despite some items do not meet one of the criteria, this by no means decreases the quality of the items. For instance, item (Q6/Bubble, Q2/PM/SL, and Q15/PM/LG) do not meet the criteria of Outfit *Z* Standard and PTMEA Corr; item (Q1/PS/SL, Q24/PM/GG and Q19/PM/SG) do not meet the criteria of PTMEA Corr; and item (Q25/SMRs/GG, Q16/DP/LG, and Q23/DP/GG) do not meet the criteria of Outfit ZSTD; this is supposedly caused by large size of sample, or $N > 500$ (Boone et al., 2014).

Wright Map: Person-Map-Item. The third step was to measure the consistency of item difficulty level and student's ability test constructed in Table 2. The higher the item difficulty level, the higher also the student's ability level will result. Information of Wright Map: Person-Map-Item is displayed in Figure 3. The previous Wright map generates that all instrument items encompass almost all the students' ability. The map generates variance from students with very high ability (> 3.0 logit), to those with very low ability (< -2.0 logit) as well. In addition to that, disparity (in which there is no item that is appropriate with the student's ability) was observed within the interval of -3.0 logit up to -0.5 logit and in the interval of +1.0 logit up to +3.7 logit. This signified that the information generated within the interval range was somewhat limited and required further elaboration. On the other hand, the item difficulty level was mostly located in the interval of -1.0 logit up to +1.0 logit; moreover, the items tended to occur in the same difficulty level. The item Q14/PS/LG was the most difficult item with a logit of +0.97, while item Q17/Pmo/LG was the easiest item with logit of -0.71.

https://doi.org/10.33225/jbse/20.19.824

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

Figure 3
Wright Map: Person-Map-Item



As observed from the differences in item size, some interesting cases were explained as follows: Firstly, the items in level 1: Q14/PS/LG (0.97) > Q1/PS/SL (0.79) > Q18/PS/PG (0.71) > Q7/PS/GG (0.66) were instead assumed by the students to possess different difficulty level. The items above, however, were more difficult than item Q6/Bubble in level 3 (0.60). In other words, determining particle size was more difficult than explaining the particle characteristics of matter in the evaporation phenomenon. Secondly, the size of item Q5/Dew (-0.63) < item Q6/Bubble; this indicated that it was harder for the students to elaborate on the particle characteristics of matter in

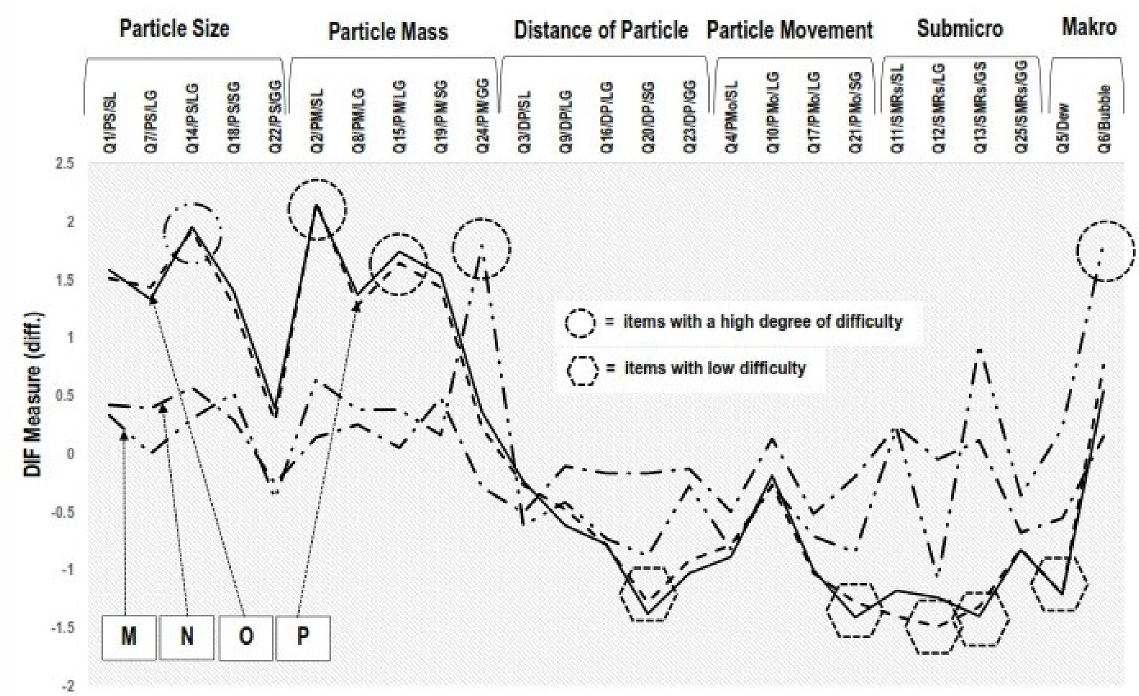
https://doi.org/10.33225/jbse/20.19.824

the evaporation phenomenon than in condensation phenomenon, despite that both items were in the same level. Thirdly, the size of following items: Q2/PM/SL (0.88) > Q19/PM/SG (0.77) > Q15/PM/LG (0.66) > Q8/PM/LG (0.65) > Q24/PM/GG in level 1 was larger compared to that of items Q13/SMRs/GS (-0.24) > Q11/SMRs/SL (-0.27) > Q12/SMRs/LG (-0.63) > Q25/SMRs/GG (-0.68) in level 2. The finding illustrated that it was harder for the students to determine the particle mass than determining the submicrorepresentation (SMRs) diagram in different form changes of matter. The previous cases identified disparity in students' conceptual understanding, signifying that the level of understanding in particle characteristics of the matter is relatively low. Overall, 80% of test item difficulty level is relatively parallel with the measured constructs. By that, the test possesses good construct validity (Blanc & Rojas, 2018; Lu & Bi, 2016; Neumann et al., 2011).

Disparity in Conceptual Understanding Level

The next step was the measurement of disparity of students' conceptual understanding in the focused topic based on educational level by Differential Item Functioning (DIF).

Figure 4
Person DIF plot based on educational level



Note: M = X Class students, N = XI Class students, O = XII Class students, and P = University students from chemistry department

Figure 4 of DIF plot based on students' educational level depicts that ten items are identified to possess significant disparity. Firstly, five curves approaching the upper limit are items with a high difficulty level (Q14/PS/LG, Q2/PM/SL, Q15/PM/SG, Q24/PM/GG and Q6/Bubble); while five curves approaching the lower limit are items with a low difficulty level (Q20/DP/SG, Q21/PMo/SG, Q12/SMRs/LG, Q13/SMRs/GS, and Q5/Dew). Secondly, the item Q14/PS/LG (particle size in form change of liquid-gas), Q2/PM/SL (particle mass in form change of solid-liquid), and Q15/PM/SG (particle mass in change form of solid-gas) were deemed very hard by the students of XII class and the university students compared to students in X and XI class. Thirdly, the research discovered different results for item Q24/PM/GG and Q6/Bubble. The item Q24/PM/GG (particle mass of O2 in larger volume) and Q6/Bubble (constructing elements of air bubbles during boiling process of water) were deemed very hard for X class students compared to students in XI and XII classes, as well as university students. Fourthly, the items Q20/DP/SG

https://doi.org/10.33225/jbse/20.19.824

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

(distance between particles in form change of solid-gas), Q21/PMo/SG (motion between particles in form change of solid-gas), Q12/SMRs/LG (SMRs diagram of particle in form change of liquid-gas), Q13/SMRs/GS (SMRs diagram of particle in change form of gas-liquid), and Q5/Dew (condensation) were deemed too easy for students in XII class and university students compared to the students in X and XI classes.

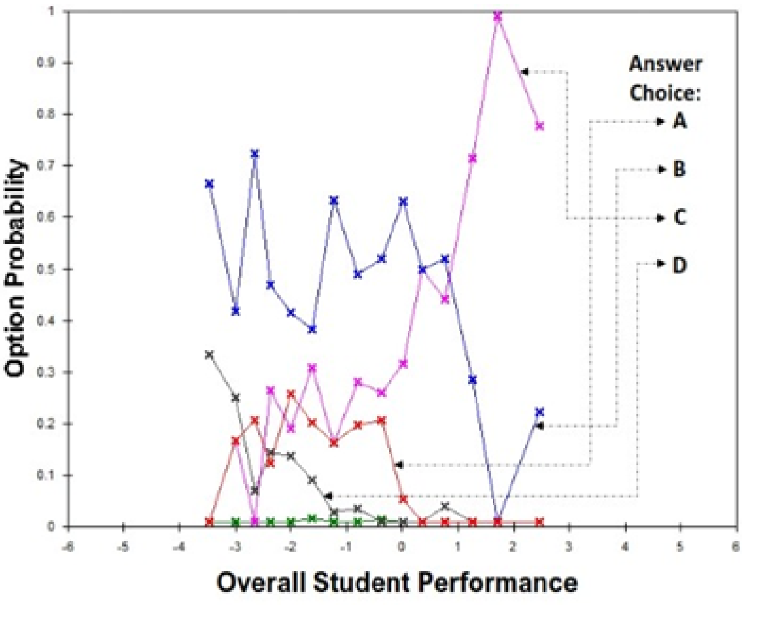
Pattern of Conceptual Understanding and Preconception

The analysis of the pattern of conceptual understanding and preconception employed an option probability curve test (Boone et al., 2014; Linacre, 2012). The option probability curve aims to display the probability of picking every answer choice to elaborate on the performance level of all students in the measured items (Herrmann-Abell & DeBoer, 2011). The test relied on the principle that the curve of the correct answer will rise along with the decrease of the curve of distractor choices (Boone et al., 2014; Haladyna, 2004). For items that are influenced by distractor options, the curve produced tends to be non-parallel with the traditional monotonous item behavior (Sadler, 1998), for this reason, each answer choice was analyzed separately.

The instrument provides four answer choices, thus resulting in four curves. Each curve displays the students' comprehension. Students with low ability tended to pick distractor choice, while students whose high ability were more likely to prefer other preconceptions (Herrmann-Abell & DeBoer, 2011; Perera et al., 2018). Below is the elaboration of the pattern of students' conceptual understanding and preconception based on four option probability curves.

Figure 5
(a) sample of item Q2/PM/SL, (b) option probability curve

(a)
When some ice cubes in the glass melts, some other ice cubes are seen floating on the water surface. How is the comparison between mass of one particle of ice and one particle of water?
A. Mass of one particle of ice is bigger than mass of one particle of water.
B. mass of one particle of ice is smaller than mass of one particle of water.
C. mass of one particle of ice is similar with mass of one particle of water.
D Other answer...

(b)


The first example, i.e., the item Q2/PM/SL (0.88), is shown in Figure 5(a). The item measures students' capability in determining particle size in form change from solid to liquid. The option probability curve is displayed in Figure 5(b). Students with the low ability (< 0.5 logit) tended to pick distractor choice B (mass of one particle of ice is smaller than the mass of one particle of water) or A (mass of one particle of ice is bigger than the mass of one particle of water). In addition, students with very low ability (< -1.0 logit) tended to pick D (other answers). Some students with relatively low ability (> -2.5 logit), however, picked the right answer C (mass of one particle of ice is similar to the mass of one particle of water). One can predict the response pattern of students with low ability, as the distractors A, B, and D contain third preconceptions in level 1 (see Table 2). The students possess the knowledge

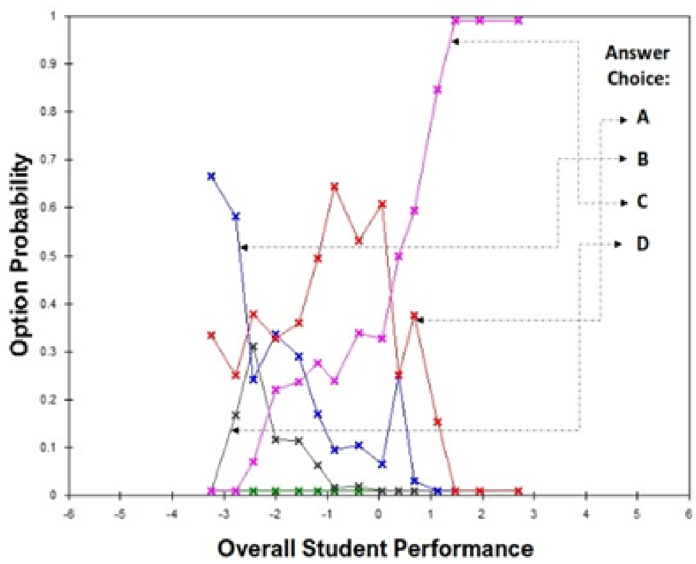
https://doi.org/10.33225/jbse/20.19.824

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

that mass of particle of matter can change into larger or smaller size by observing the matter's change of form. It is interesting to note that there are students with the high ability (>2.0) who picked B; this indicates the presence of resistant preconception.

Figure 6
(a) item Q8/PM/LG; (b) option probability curve

(a)
During boiling process of water, a change of form of water from liquid to gas occurs. How is the comparison between mass of one particle of water and one particle of water vapor?
A. Mass of one water particle is bigger than mass of one vapor particle
B. mass of one water particle is smaller than mass of one vapor particle.
C. mass of one water particle is similar with mass of one vapor particle.
D Other answers...

(b)


The second sample or item Q8/PM/LG (0.65) is shown in figure 6(a) as the item to measure students' ability to determine the mass of the particle in form change of liquid-gas. The option probability curve is displayed in Figure 6(b). The curve of distractor B (mass of one water particle is smaller than the mass of one vapor particle) is chosen by students with low ability (< -2.0 logit), while the curve of choice A (mass of one water particle is bigger than mass of one vapor particle) was chosen by students with ability in a range of -3.5 to 1.5 logit. The correct answer, option C (mass of one water particle is similar to the mass of one vapor particle), was chosen by students with ability in > -2.5 logit. As highlighted in the table, the decline of the curve of distractor A is followed by the increase of curve of right answer C; both curves intersect in the level of 1.0 logit. The shape of curve A indicates the presence of resistant preconception type-three in level 1.

It depicts that the particular item response pattern that signifies students' conceptual understanding patterns in the given level. Moreover, the curve shape of distractors A and B in the items Q2/PM/SL and Q8/PM/SL tend to have an identical pattern. The finding indicated that students with either low or high ability had consistent preconceptions that the mass of the particle can change into larger or smaller in size along with the change in matter form.

Third sample, i.e., item Q5/Dew (-0.63), as shown in Figure 7(a), measures the students' ability in elaborating characteristics of a particle in condensation phenomenon. The option probability curve is displayed in Figure 7(b). Students with low ability (< 1.0 logit) tended to pick distractor A (water drops come from liquid of melting ice that breaks through the glass wall) and option D (other answers). Some students with high ability (> 1.0 logit) also picked distractor B (water drops are the result of the reaction between ice and air nearby the glass). The shape of curve B is wavy and non-linear, even in the interval of 2.0 to 4.0 logit, it can reach option probability value up to 1.0 logit. This is regarded as a deviation from the right answer C (water drops come from condensing water vapor nearby the glass). A worth note, however, is to consider in the unstable, wavy shape of curve C. This indicated the students' inconsistency (particularly those with high ability) in comprehending the concept of condensation. This confirmed that students had their own preconception regarding concept of condensation.

https://doi.org/10.33225/jbse/20.19.824

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

Figure 7

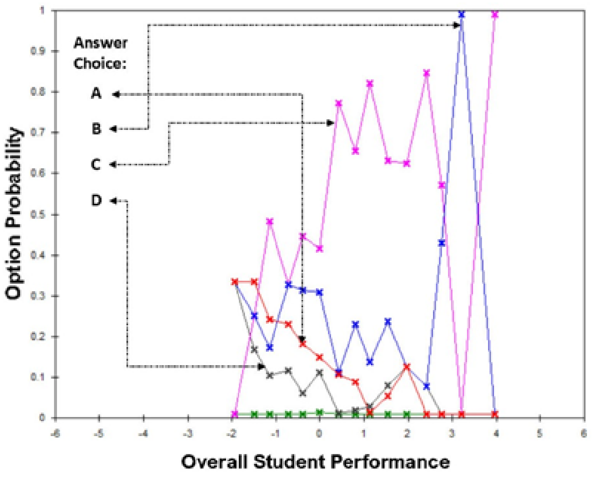
(a) item Q5/Dew; (b) option probability curve

(a)

In a glass filled with ice, you can see water drops at the glass. According to you, where do the water drops come from?

A. Water drops come from melting ice that penetrates the glass wall
B. water drops are the result of reaction between ice and air nearby the glass
C. Water drops come from condensing water vapor nearby the glass
D. Other answer...

(b)



The fourth sample or item Q6/Bubble, as shown in Figure 8(a), measures the students' ability in elaborating characteristics of a particle in the evaporation phenomenon. The option probability curve is displayed in Figure 8(b). The distractor A (air bubbles are Hydrogen and Oxygen particles) was dominantly chosen by students whose ability in a range between -3.0 to 2.0 logit. Moreover, the distractor B (air bubbles are Hydrogen and Oxygen particles) was mostly selected by students whose ability in a range between -3.0 to 0.5 logit. The form of curve A and B were picked by students with low ability was predictable. The curve of right answer C (air bubbles are water molecules), however, shows interesting hint; in the interval range of -2.5 to 3.0 logit, the tip of the curve shows an up-and-down pattern. Moreover, in the level of 1.5 logit, the curve shape of distractors A and B shows a decline pattern, while that of curve C tends to increase. Another finding worth noting was that the curve D (other answers) was picked by some students with high ability (> 2.0 logit). This indicated that particular students had their own preconceptions regarding the evaporation concept.

Figure 8

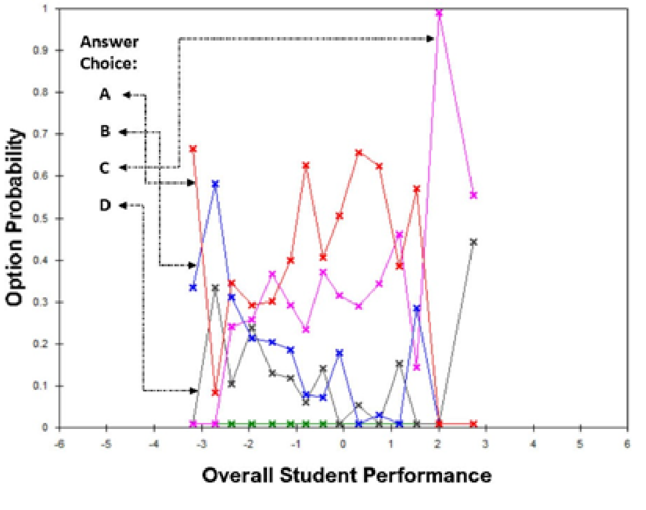
(a) item Q6/Bubble; (b) option probability curve

(a)

In a container filled with boiling water, you can see air bubbles on the top of it. According to you, what are the composing elements of the air bubbles?

A. Hydrogen and Oxygen particles
B. Air that is dissolved in water
C. Water molecules
D. Other answer...

(b)



https://doi.org/10.33225/jbse/20.19.824

Discussion

The research results indicated that the instruments had good effectiveness, met the requisites of person and item reliability, and showed good construct validity. When applied in evaluating students' conceptual understanding, it was found that: Firstly, almost all students with high ability faced difficulty in understanding the concept of particle size and mass in level 1. The same students found it relatively easy in determining SMRs diagram of particle structure in level 2 or determining the concept of particle regarding evaporation and condensation phenomena in level 3. Secondly, the information of the response pattern of students with high ability was quite consistent, repetitive, and systematic in particular items. This indicates the presence of permanent and latent preconceptions. The analysis of the option probability curve of item Q2/PM/SL (0.88), Q8/PM/LG (0.65), Q5/Dew (-0.63) and Q6/Bubble (0.60) indicates that the approach of item response pattern is able to explore in detail and comprehensively regarding students' conceptual understanding and preconception.

Sequences of verification conducted that involves Rasch model approach shows detailed, accurate, and quantifiable results since the approach integrates development procedure of diagnostic and summative instruments. Several samples of preconception, e.g., item Q2/PM/SL (0.88) and Q8/PM/LG (0.65) indicate that distractor options are potential to be elaborated further in order to investigate tendency of preconception by the students. In addition, it also provides information regarding main idea unknown to the students and their degree of misunderstanding.

The approach employed in this research is an effective illustration to help teacher in evaluating the learning process as well as the students' learning progress. This is due to the integration of qualitative item development procedure and quantitative data analysis, allowing the teachers to explore in-depth on the students' understanding, concepts the students understand and/or do not understand, and misconception. Such findings echo Herrmann-Abell and Deboer (2016) that the integration of Rasch model analysis and probability curve is applicable to diagnose how the students' misconception turns into their overall conceptual understanding. Such an attempt is quite hard to conduct by implementing a conventional approach due to the interdependence of person and item. Rasch model, on the other hand, is able to tackle such interdependence, in which the item and the test difficulty remain invariant and not dependent on which sample that is involved in the initial validation. This signifies that the instrument's items have met the unidimensionality and local independence requirements (Jin et al., 2019; Testa et al., 2019; Wei et al., 2012).

Overall, the research indicated empirical evidence that supported findings by Hoe and Subramaniam (2016); Lu and Bi (2016); Rogat et al., (2011), that students had distinctive preconception as a result of a learning process they experienced. Such preconception was regarded as the inhibitor to the development process of students' conceptual understanding (Soeharto et al., 2019). In this research, students' preconception was found to be repetitive and systematic in each education level. It signifies that the intervention to change students' preconceptions was difficult to conduct by the conventional learning method. A strategic and meaningful learning method is therefore essential to remove students' incorrect preconceptions and develop scientifically correct conceptual understanding. That being said, teachers are demanded to acquire detailed information on the forms and characteristics of students' preconceptions. In conclusion, the item response pattern analysis was an efficient and effective means to acquire such information. The information on students' preconception is important as the basis to develop appropriate and measurable instructional design in solving the students' misconception. This is in line with the previous research studies, arguing that the quality of learning progress is highly dependent on the students' learning process and learning experience (Duschl et al., 2011; Park et al, 2017; Wilson, 2009).

Conclusions

The measuring instrument developed performed well in its validity and reliability, thus, it is deemed applicable in measuring students' conceptual understanding and preconception in elaborating particle characteristics of matter. During the implementation of the instruments, the research finds out that:

1) almost all students with high ability face difficulty in understanding the concept of particle size and mass in level 1. The same students find it relatively easy in determining SMRs diagram of particle structure in level 2, as well as determining the concept of particle regarding evaporation and condensation phenomena in level 3.

https://doi.org/10.33225/jbse/20.19.824

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

2) There is a significant disparity between students' conceptual understanding based on their educational level.
3) In certain cases, it is found that the distractor item response pattern by high-ability students tends to be consistent, indicating a certain tendency of resistant preconception pattern.

The development of diagnostic instruments with Rasch model approach is deemed as the literacy process for practitioners and researchers in Indonesia. The result indicates that there is no single item that is parallel with both the highest ability and lowest ability students. This calls for further elaboration in order to improve the instrument items' quality. Moreover, an anomaly is found that students with high ability (> 1.0 logit) tend to pick distractor choices. This urges further studies to investigate structured comprehension problems. The research regards that further analysis that integrates conceptual understanding level and items designed in a gradual manner is required to define the characteristics of the students' alternative conception and to measure their learning progress. Echoing this notion, one must integrate the item design and basic principles of chemistry as a reference for further researchers and educational practitioners to implement the same approach conducted in the present research. On top of that, despite not focused on discussing matters regarding students' learning progress individually, the instrument is expected to be beneficial for the teachers to diagnose students' conception in developing an effective and meaningful learning experience.

Acknowledgments

This research is funded by the Directorate of Research and Community Outreach/ (Direktorat Riset dan Pengabdian Masyarakat or DRPM), Directorate General of Research and Development (Dirjen Riset dan Pengembangan) of Ministry of Research and Technology of the Republic of Indonesia through Penugasan Penelitian Dasar Unggulan Perguruan Tinggi Grant Program, Research and Community Outreach Institute (Lembaga Penelitian dan Pengabdian Masyarakat) of Universitas Negeri Gorontalo in 2020.

References

Adadan, E., & Savasci, F. (2012). An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, *34*(4), 513–544. https://doi.org/10.1080/09500693.2011.636084

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

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

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

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

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

Blanc, A., & Rojas, A. J. (2018). Use of Rasch person-item maps to validate a theoretical model for measuring attitudes toward sexual behaviors. *PLoS ONE*, *13*(8), 1–17. https://doi.org/10.1371/journal.pone.0202551

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

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

Boz, Y. (2006). Turkish pupils' conceptions of the particulate nature of matter. *Journal of Science Education and Technology*, *15*(2), 203–213. https://doi.org/10.1007/s10956-006-9003-9

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, *8*(3), 293–307.

Cheng, M. M. W. (2018). Students' visualization of chemical reactions-insights into the particle model and the atomic model. *Chemistry Education Research and Practice*, *19*(1), 227–239. https://doi.org/10.1039/c6rp00235h

Davidowitz, B., Chittleborough, G., & Murray, E. (2010). Student-generated submicro diagrams: A useful tool for teaching and learning chemical equations and stoichiometry. *Chemistry Education Research and Practice*, *11*(3), 154–164. https://doi.org/10.1039/c005464j

https://doi.org/10.33225/jbse/20.19.824

Devetak, I., & Glažar, S. A. (2010). The influence of 16-year-old students' gender, mental abilities, and motivation on their reading and drawing submicrorepresentations achievements. *International Journal of Science Education*, *32*(12), 1561–1593. https://doi.org/10.1080/09500690903150609

Devetak, I., Urbancic, M., Grm, K. S. W., Krnel, D., & Glazar, S. A. (2004). Submicroscopic representations as a tool for evaluating students' chemical misconception. *Acta Chimica Slovenica*, *51*(December 2004), 799–814. https://www.researchgate.net/publication/286849527_Submicroscopic_representations_as_a_tool_for_evaluating_students'_chemical_conceptions

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, *47*(2), 123–182. https://doi.org/10.1080/03057267.2011.604476

Femintasari, V., Effendy, & Munzil. (2015). The effectiveness of two-tier multiple-choice test and multiple-choice test followed with interview in identifying misconception of students with different scientific reasoning skills. *Jurnal Ilmu Pendidikan*, *21*(2), 192–197. http://journal.um.ac.id/index.php/jip/article/view/8376/4020

Gabel, D. L. (1993). Use of the particle nature of matter in developing conceptual understanding. *Journal of Chemical Education*, *70*(3), 193. https://doi.org/10.1021/ed070p193

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

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

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

Herrmann-Abell, C. F., & Deboer, G. E. (2016). Using Rasch modeling and option probability curves to diagnose students' misconceptions. *American Educational Research Association*, 1–12. https://www.aaas.org/sites/default/files/Project2061_CHA-AERA%20energy%20paper%204-7-16.pdf

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

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

Jin, H., Mikeska, J. N., Hokayem, H., & Mavronikolas, E. (2019). Toward coherence in curriculum, instruction, and assessment: A review of learning progression literature. *Science Education*, *103*(5), 1206–1234. https://doi.org/10.1002/sce.21525

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

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

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

Kapici, H. Ö., & Akcay, H. (2016). Particulate nature of matter misconceptions held by middle and high school students in Turkey. *European Journal of Education Studies*, *2*(8), 43–58. https://doi.org/10.5281/zenodo.163547

Kim, J., & Wilson, M. (2019). Polytomous item explanatory item response theory models. *Educational and Psychological Measurement*. https://doi.org/10.1177/0013164419892667

Kind, V. (2004). *Beyond appearances: Students' misconceptions about basic chemical ideas* (2nd Edition). Royal Society of Chemistry. https://www.researchgate.net/publication/228799159_Beyond_Appearances_Students'_Misconceptions_About_Basic_Chemical_Ideas

Laliyo, L. A. R., Botutihe, D. N., & Panigoro, C. (2019). The development of a two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12

Linacre, J. M. (2012). *A user's guide to W I N S T E P S ® M I N I S T E P Rasch-model computer program: Program manual 3.75.0*. https://doi.org/ISBN 0-941938-03-4

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

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

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, *34*(15), 2317–2341. https://doi.org/10.1080/09500693.2012.684433

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

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

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

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*(9), 825–838. https://doi.org/10.1002/tea.3660200905

Ozalp, D., & Kahvecib, A. (2015). Diagnostic assessment of student misconceptions about the particulate nature of matter from ontological perspective. *Chemical Education Research and Practice*, *16*, 619–639. https://doi.org/10.1039/C5RP00096C

Ozmen, H. (2011). Turkish primary students´ conceptions about the particulate nature of matter. *International Journal of Environment & Science Education*, *6*(1), 99–121. https://www.researchgate.net/publication/265438996_Turkish_primary_students'_conceptions_about_the_particulate_nature_of_matter

Park, M., Liu, X., & Waight, N. (2017). Development of the connected chemistry as formative assessment pedagogy for high school chemistry teaching. *Journal of Chemical Education*, *94*(3), 273–281. https://doi.org/10.1021/acs.jchemed.6b00299

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

Renström, L., Andersson, B., & Marton, F. (1990). Students' conceptions of matter. *Journal of Educational Psychology*, *82*(3), 555–569. https://doi.org/10.1037/0022-0663.82.3.555

Rogat, A., Anderson, C., Foster, J., Goldberg, F., Hicks, J., Kanter, D., Krajcik, J., Lehrer, R., Reiser, B., & Wiser, M. (2011). *Developing learning progression in support of the new science standards: A RAPID workshop series, New York, NY.* (4), 163. https://doi.org/10.12698/cpre.2011.lprapid

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

Sadler, P. M. (1999). The relevance of multiple-choice testing in assessing science understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 251–274). https://zodml.org/sites/default/files/%5BJoel_J._Mintzes%2C_James_H._Wandersee%2C_Joseph_D._No_0.pdf

Slapničar, M., Devetak, I., Glažar, A. S., & Pavlin, J. (2017). Identification of the understanding of the states of matter of water and air among Slovenian students aged 12, 14 and 16 years through solving authentic tasks. *Journal of Baltic Science Education*, *16*(3), 308–323.

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

Stojanovska, M. I., Soptrajanov, B. T., & Petrusevski, V. M. (2012). Addressing misconceptions about the particulate nature of matter among secondary-school and high-school students in the Republic of Macedonia. *Creative Education*, *03*(05), 619–631. https://doi.org/10.4236/ce.2012.35091

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

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial* [Application of Rasch Model in Social Sciences Research]. Tri Komunikata. https://www.researchgate.net/publication/268688933%0AAplikasi

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan* [Application of Rasch Model in Education Assessment]. Tim Komunikata. https://www.researchgate.net/publication/282673464%0AAplikasi

Taber, K. S. (2015). Alternative Conceptions/Frameworks/Misconceptions. In G. Richard (Ed.), *Encyclopedia of Science Education* (pp. 37–41). https://doi.org/10.1007/978-94-007-6165-0

Testa, I., Capasso, G., Colantonio, A., Galano, S., Marzoli, I., Scotti di Uccio, U., Trani., F & Zappia, A. (2019). Development and validation of a university students' progression in learning quantum mechanics through exploratory factor analysis and Rasch analysis. *International Journal of Science Education*, *41*(3), 388–417. https://doi.org/10.1080/09500693.2018.1556414

Tóth, Z., & Kiss, E. (2006). Using particulate drawings to study 13-17year olds' understanding of physical and chemical composition of matter as well as the state of matter. *Practice and Theory in Systems of Education*, *1*(1), 109–125.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169. https://doi.org/10.1080/0950069880100204

Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, *4*(6), 626–631. http://www.academicjournals.org/SRE

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using Rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335–345. https://doi.org/10.1021/ed100852t

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

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

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

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

Yaşar, I. Z., İnce, E., & Kırbaşlar, F. G. (2014). 7. Class Science and Technology Course "Structure of Atom" Subject Readiness

https://doi.org/10.33225/jbse/20.19.824

Improvement Test. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2014.09.259
Yildirir, H. E., & Demirkol, H. (2018). Identifying mental models of students for physical and chemical change. *Journal of Baltic Science Education*, *17*(6), 986–1004. https://doi.org/10.33225/jbse/18.17.986

Received: *April 09, 2020* Accepted: *September 24, 2020*

Cite as: Laliyo, L. A. R., Tangio, J. S., Sumintono, B., Jahja, M., & Panigooro, C. (2020). Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter. *Journal of Baltic Science Education*, *19*(5), 824-841. https://doi.org/10.33225/jbse/20.19.824

Lukman Abdul Rauf Laliyo
(Corresponding author)
Doctor, Associate Professor, Department of Chemistry, Faculty of Mathematics and Natural Sciences, State University of Gorontalo, Indonesia.
E-mail: lukman.laliyo019@gmail.com
ORCID: https://orcid.org/0000-0003-3281-7202

Julhim S. Tangio
Master of Education, Assistant Professor, Department of Chemistry, Faculty of Mathematics and Natural Sciences, State University of Gorontalo, Indonesia.
E-mail: transbahasaid@gmail.com

Bambang Sumintono
PhD, Lecturer in Institute of Educational Leadership, University of Malaya, Kuala Lumpur, Malaysia.
Email: transbahasa.ilmiah@gmail.com

Mohamad Jahja
Doctor, Associate Professor, Department of Chemistry, Faculty of Mathematics and Natural Sciences, State University of Gorontalo, Indonesia.
E-mail: transbahasatb@gmail.com

Citra Panigooro
Master of Science, Assistant Professor, Faculty of Faculty of Fisheries and Marine Science, State University of Gorontalo, Indonesia.
E-mail: citrapanigoro25@gmail.com