

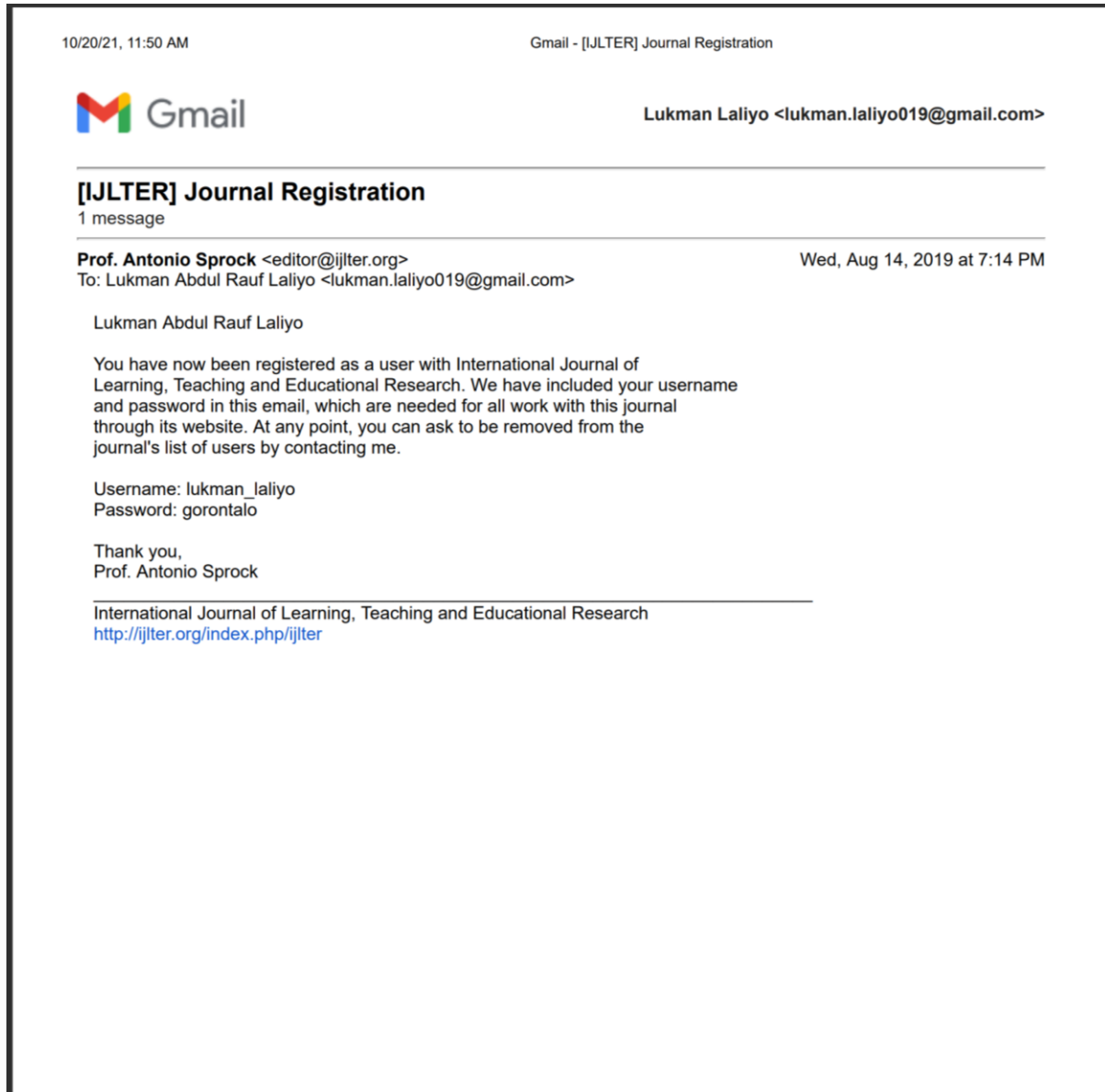
BUKTI KORESPONDENSI

Judul : The Development of Two-Tier Instrument Based On Distractor to Assess Conceptual Understanding Level and Student Misconceptions in Explaining Redox Reactions

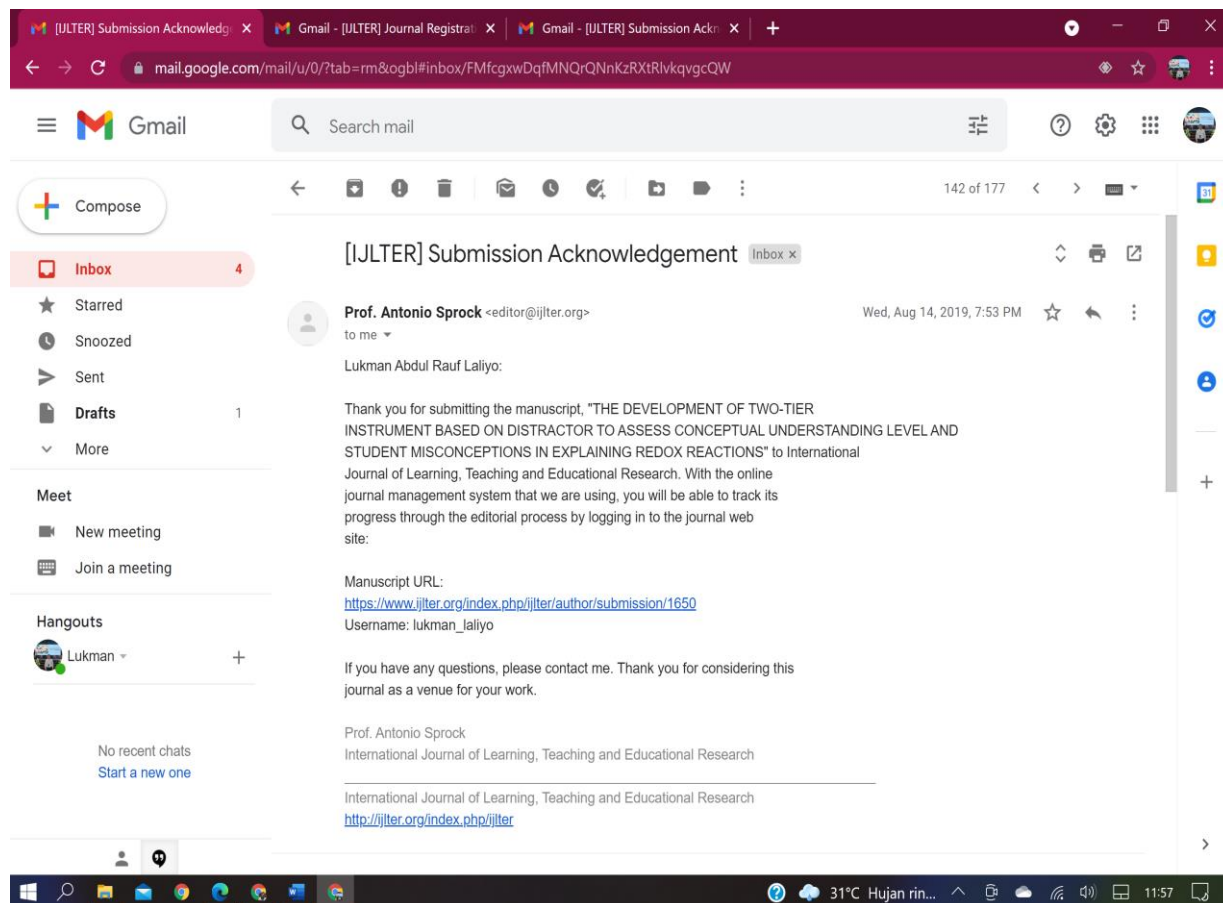
Jurnal : International Journal of Learning, Teaching and Educational Research Vol. 18, No. 9, pp. 216-237, September 2019
<https://doi.org/10.26803/ijlter.18.9.12>

Proses:

1. Registrasi (14 Agustus 2019)



2. Bukti Proses Submitted (14 Agustus 2019)



[IJLTER] Submission Acknowledg... x Gmail - [IJLTER] Journal Registr... x Gmail - [IJLTER] Submission Ackr... x +

mail.google.com/mail/u/0/?tab=rm&ogbl#inbox/FMfcgxdqfMNQrQnKzRXtRlvkqvgcQW

Gmail Search mail

Compose

Inbox 4

Starred

Snoozed

Sent

Drafts 1

More

Meet

New meeting

Join a meeting

Hangouts

Lukman +

No recent chats
Start a new one

142 of 177

IJLTER ORG <ijlter.org@gmail.com>
to me
Aug 16, 2019, 10:52 AM

Dear Dr Laliyo,

Thank you for your submission.

We would like to inform you that our publication fees are \$350.
Are you agreeable to pay the fees if the paper is accepted for publication?

--

Editorial Office
International Journal of Learning, Teaching and Educational Research
ISSN: 1694-2116 (Online)
ISSN: 1694-2493 (Print)
Website: <http://ijlter.org/>
Email: ijlter.org@gmail.com

Lukman Laliyo <lukman.laliyo019@gmail.com>
to IJLTER
Aug 16, 2019, 9:43 PM

Dear Editorial Team of International Journal of Learning, Teaching and Educational Research

We would like to confirm that we agree to pay the publication fee as much you mention above.

Thank you!

[IJLTER] Submission Acknowledg... x Gmail - [IJLTER] Journal Registr... x Gmail - [IJLTER] Submission Ackr... x +

mail.google.com/mail/u/0/?tab=rm&ogbl#inbox/FMfcgxdqfMNQrQnKzRXtRlvkqvgcQW

Gmail Search mail

Compose

Inbox 4

Starred

Snoozed

Sent

Drafts 1

More

Meet

New meeting

Join a meeting

Hangouts

Lukman +

No recent chats
Start a new one

142 of 177

Lukman Laliyo <lukman.laliyo019@gmail.com>
to IJLTER
Aug 16, 2019, 9:43 PM

Dear Editorial Team of International Journal of Learning, Teaching and Educational Research

We would like to confirm that we agree to pay the publication fee as much you mention above.

Thank you!

IJLTER ORG <ijlter.org@gmail.com>
to me
Aug 19, 2019, 12:06 AM

Thank you Dr Laliyo. Your paper is now under review.

Thank you for your response. Thank you for the update. Thanks a lot.

Reply Forward

THE DEVELOPMENT OF TWO-TIER INSTRUMENT BASED ON DISTRACTOR TO ASSESS CONCEPTUAL UNDERSTANDING LEVEL AND STUDENT MISCONCEPTIONS IN EXPLAINING REDOX REACTIONS

ABSTRACT

Fifteen distractor two-level multiple choice items were developed as diagnostic instruments to evaluate the level of conceptual understanding and structure of students' misconceptions in explaining redox reactions. Questions at the first tier (Q1) assess the level of knowledge, and questions at the second tier (Q2) assess the level of reasoning of students. This instrument was given to 1150 participants. The participants were 11th grade students, from eight senior high schools, in the Eastern part of Indonesia. The collected data was analyzed using the Rasch model approach. The results of this study provide diagnostic and summative information on the progressiveness of student learning outcomes, as well as evidence of empirical validity and reliability of measurement. In addition, by comparing the size of items Q1 with Q2, it was found that the level of student knowledge is not always proportional to the level of reasoning, even in some cases, the level of knowledge is lower than the level of reasoning, and vice versa. The results of the investigation using the option probability curve; it was revealed that there **was** students misconceptions and inconsistencies about the concepts of reduction, oxidation and oxidation numbers. This result confirms why students have difficulty interpreting and converting redox reaction equations.

Keywords: two-tier instrument, distractor, understanding, reasoning, redox reaction.

INTRODUCTION

The framework of this research is based on constructivism approaches, which assume that students actively build their own conceptual understanding. In this context, teachers play an important role in facilitating students to develop their learning creativity, to construct their knowledge and conceptual understanding as optimally as possible. Early concepts that have been understood by students -will determine the learning outcomes (Rahayu, 2017). However, this effort is not easy for students. Many studies have showed that students must study hard to understand chemical concepts deeply and meaningfully (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013).

Students **are** difficult to understand chemical concepts due to their complex and abstract characteristics (Johnstone, 1991; Taber, 2013; Gabel, 1999). Therefore, students develop their own conceptions, which tend to be different from the scientific community **presumption**. Conception that has been firmly embedded in the minds of students as a coherent but wrong conceptual structure, especially when it comes to students' daily life experiences (Chandrasegaran, Treagust, & Mocerino, 2007). This student conception is commonly known as **misconception** (Johnstone, 2006; Taber, 2009; Taber, 2013).

Scientific education researchers have developed many types of instruments to diagnose student misconceptions. One of them that is often used today is a two-tier multiple choice diagnostic instrument (Treagust, 1988; Chandrasegaran et al., 2007; Tüysüz, 2009; Femintasari, 2015). This instrument can not only evaluate conceptual understanding, but also can diagnose student misconceptions. Qualitatively, this diagnostic instrument is relatively effective in providing information regarding the way of students think and who have misconceptions. However this instrument tends to have limitations, such as the weak internal consistency. Because the instrument is not in the same measurement dimension, then it is difficult to use for summative measurement (Lu & Bi, 2016).

Misconceptions studies tend to have interesting progress, when researchers have new concepts of misconceptions. Research findings showed that misconceptions are often resistant and still persist even after formal learning. This new concept of **misconception** developed since the early 2000s, along with the progression of knowledge and reasoning, which varied according to the level of understanding of students (Aktan, 2013). Two-level diagnostic instruments can diagnose misconceptions, but cannot measure the level of conceptual understanding and evaluate the development of students' knowledge and reasoning. As a result the information produced is relatively limited for teachers for making decisions (Wilson, 2008).

Furthermore, Rasch model has been introduced as a measurement method that can integrate the diagnostic evaluation approach of students' misconception with summative evaluation of students (Liu, 2012). Although there are still studies that develop formative evaluations to investigate the progression of student learning outcomes (Claesgens, Scalise, Wilson, & Stacy, 2009; Hadenfeldt et al., 2013), several other studies

Commented [TR1]: Were student

Commented [TR2]: Can find it

Commented [TR3]: reference

Commented [TR4]: Re-phrase these two sentences

Commented [TR5]: You are using this word too often

have developed integrative diagnostic and summative approaches (Hoe & Subramaniam, 2016). In fact, there are studies have developed instruments based on Rasch modeling, which integrate the diagnostic evaluation approach to misconception with summative evaluation in chemistry learning (Chi, Wang, Luo, Yang, & Huang, 2018).

This study was designed using a Rasch-based instrument, in the form of a two-level distractor-based multiple choice item. Weaknesses of multiple choice items, that often criticized such as **only** testing facts and unable to diagnose student understanding (Klassen, 2006), are overcome by adding the power of diagnostic selected items with two stratified questions (Treagust, 1988), and using answer distractor choices on each item (Sadler, 1998; Herrmann-Abell & DeBoer, 2011).

The aim of this research was to develop a distractor-based two-tier multiple-choice diagnostic instrument (TMDDI), to evaluate the level of conceptual understanding and structure of students' misconceptions in explaining the redox reaction.

Redox reaction is one of the main topics in Chemistry curriculum of secondary school in Indonesia. Evaluating the level of students' conceptual understanding in explaining redox reaction, includes two important aspects, namely knowledge and reasoning. Knowledge evaluation is related to measuring students' knowledge about the concepts of oxidation, reduction and changes in oxidation numbers. Reasoning evaluation is a measurement of a student's ability to give reasons/responses to justify what students have understood. The quality of students' conceptual understanding is good, if students are able to know and explain their knowledge well. In this study there are two problems to resolve, namely (1) How is the effectiveness of the instrument when it is used to measure the level of conceptual understanding and diagnose the structure of student misconception? And, (2) What is the description of conceptual understanding level and structure of students' misconceptions in explaining the redox reaction?

METHODS

There are three important components that are interrelated in the progression of measurement instruments: cognition, observations and interpretations. Cognition refers to theories or constructs about how students' understanding develops related to what they learn. Observations refer to student performance based on the type of problem solving or task and situation at the time of measurement. Interpretations refer to the results of data analysis, in the form of a statistical model as a summary of students' understanding patterns (NRC, 2001). These three components become an important part of the development stage of the instrument proposed by Wilson (2008), which includes developmental progress variables, item models, outcome space, and measurement models. The first stage is developmental progress variable, **with** regard to the developmental evaluation variables and progressiveness of student learning outcomes that only focus on one characteristic to be measured. The second stage, the items model relating to the form of items that designed to obtain suitability between the measurement and various understanding diagnoses. Each item is designed to provide a response that is diagnosing student understanding, at least one level for one construct map. The third stage, the outcome space, is student learning outcomes that are categorized for all items related to certain progress variables. The fourth stage, the measurement model, in this study uses the Rasch model.

The diagnostic instruments developed in this study were adapted from the framework of developing two-tier diagnostic instruments reported by Treagust (1988) and Chandrasegaran, Treagust, & Mocerino (2007). Distractor item design were adapted from Herrmann-Abell & DeBoer (2011) and Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann (2013), so that it can be used to diagnosis what students do, what doesn't they **knew**, the misconceptions and alternative ideas that they have understood. The data obtained were analyzed using the Rasch model approach, through the four steps described below.

1. Description of students' conceptual understanding

The first step is to describe students' conceptual understanding. This requires a construct map that provides the substantive definition and the qualitative level of the construct, according to the complexity of the construct of the variable being measured (Wilson, 2009). Construct definition map is related to what students must understand and how the progress of student learning outcomes in developing their **understanding** (Wilson, 2012). The map definitions of students' conceptual understanding in explaining the redox reaction was elaborated from the Chemistry Curriculum of High Schools in Indonesia. Learning redox reaction concept intended to improve students' ability to identify redox reactions, oxidation numbers and able to analyze the redox reaction based on changes in oxidation numbers, obtained from experimental data or through experiments (Ministry of Education and Culture, 2016). The progressiveness of student learning outcomes is the comprehensive and interrelated students' conceptual understanding in explaining the redox reaction. The word "explain" in the definition, implies that students must be able to interpret and develop their

Commented [TR6]: rephrase

Commented [TR7]: what they don't do

Commented [TR8]: rephrase

reasoning, that form **knowledge** to be used to solve problems related to redox reactions. A map of the definition of level of conceptual understanding is shown in Table 1.

Commented [TR9]: rephrase

Table 1. The level of students' conceptual understanding in explaining the redox reaction

Representation Level	Conceptual Understanding of Redox Reaction	Item
Macro, submicro, symbolic	9. Students can determine non-redox reaction by analyzing submicroscopic diagram.	14E
	8. Students can convert submicroscopic diagrams of iron rusting reaction to symbolic chemical reaction equation	13D
	7. Students can determine autoredox reaction by analyzing submicroscopic diagram that illustrated by reaction between bleaching powder and HCl and dissolve in water.	15F
Symbolic	6. Students can determine elements that undergoing oxidation state change in A and B redox reaction	4skAb, 6skBb
	5. Students can determine substances that act as oxidator and reductor in A and redox reaction.	3skAs, 7skBs
Submicroscopic	4. Students can determine submicroscopic diagram of irreversible reaction between Zn dan CuSO ₄ .	10sbBx
	3. Students can determine net ion equation by submicroscopic diagram of A dan B redox reaction.	2sbAi, 9sbBi
	2. Students can interpret submicroscopic diagram of transfer elctron in B and C redox reaction.	8sbBe, 12sbCe
Macroscopic	1. Students can identify color change of solution due to redox process:	
	A. Iron metal dipped in CuSO ₄ solution	1mA
	B. Zink metal dipped in CuSO ₄ solution	5mB
	C. Iron metal dipped in H ₂ SO ₄ solution	11mC

Table 1 presents nine levels of conceptual understanding of students, in interpreting and converting redox reactions A, B and C, through macro, submicro, and symbolic representations. The purpose using representations are (1) -Macroscopic representation to measure students' abilities in describing real and visible redox reactions, according to everyday student experience, or at least experience when observing change in redox reactions (such as discoloration of solutions, gas formation, deposits) through laboratory experiment. (2) Submicroscopic representation to measure their ability to interpret and explain redox reactions at the particulate level, using images of atoms, molecules and ions. (3) Symbolic representation to measure students' abilities using chemical symbols, formulas and equations, molecular structure, and diagrams. (4) through macro, submicro and symbolic representations, can be measured students' ability to explain redox reactions based on the relationship of the three levels of representation. Therefore the ability of students to interpret and convert macro, submicro and symbolic representations is closely related to the construction of students' conceptual understanding. If students do not have proper understanding of the redox reaction, then students may difficult to explain the redox reaction using representation.

Acquiring chemical knowledge without clear conceptual understanding can be confusing for student, because simultaneously they are required to be able to explain the concept using all three levels of representation (Chandrasegaran et al., 2007). Students must explain reaction changes at the macroscopic level, then explain changes at the particulate level, and the last explain using symbols and formulas (Gabel, 1999). For example, the redox reaction A. When iron (Fe) is dipped in CuSO₄ solution, macroscopically, the color changes reddish. The next level is students must be able to interpret the "macro" changes to the particle level (using a submicroscopic diagram), and then symbolically write down the redox reaction equation, into: $\text{Fe}_{(s)} + \text{Cu}^{2+}_{(aq)} \rightarrow \text{Fe}^{2+}_{(aq)} + \text{Cu}_{(s)}$ (symbolic). Students competency to explain the redox reaction A, by interrelating the three levels of representation, can only be achieved when students are facilitated to experience direct learning with the real world (macro), then submicro and symbolic explanations are gradually introduced. Therefore getting knowledge by students becomes more structured, systematic and constructive that stored in the long-term memory of students (Tsaparlis, 2009).

Consider the following picture:

Question:

Q1. Figure (a) show iron metal, and (b) is CuSO_4 solution with blue color. If iron metal dipped into CuSO_4 solution, what is the change that will be occur ?

- Iron metal does not dissolve in CuSO_4 solution
- Iron metal dissolved like sucrose dissolved in water
- Iron changed to reddish color.
- CuSO_4 solution does not undergo any change

Q2. Which one of the following is the reason of your answer:

- Iron metal undergo oxidation
- Iron metal does not react with CuSO_4 solution
- Iron metal is easier to gain electron.
- Iron metal dissolve in CuSO_4 solution

Figure 1. Item 1mA(Q1/Q2)

2. Item design

The second step is developing item design. Item designs that are often used to evaluate students' conceptual understanding are interviews and open tests. But these type of tests take a long time, and are not comfortable to be used by classroom teachers, especially to diagnose students' conceptual understanding in large numbers (Lu & Bi, 2016; Mintzes, Wandersee, & Novak, 1999). Multiple choice is item design that have been chosen, because easy and effective (Wilson, 2008). This form has often been debated and criticized, because it is considered to evaluate only the facts or terminate definitions. Therefore it is not possible to be used to diagnose students' conceptual understanding (Klassen, 2006). However, in previous studies, Treagust (1988) has answered the criticism and succeeded in developing a two-tier multiple choice diagnostic instrument, to evaluate students' conceptual understanding.

Based on Treagust's (1988) proposal, Chandrasegaran et al., (2007) developed two-tier multiple choice items to evaluate students' ability to describe and explain chemical reactions, using various levels of representation. Furthermore, this form is used by Wilson (2008) to develop a item response model, as a cognitive diagnostic instrument to evaluate the progress of student learning outcomes. Similarly, Tüysüz (2009) developed a two-tier multiple choice item to evaluate students' chemical understanding of separation of matter concepts. Briggs (2009) with an ordered multiple choice (OMC) item format, developing an item response model to evaluate the level of progress of student learning outcomes. The idea of OMC, adapted by Hadenfeldt et al., (2013) to develop a diagnostic instrument that can evaluate sequencing of student learning outcomes, in studying certain concepts. Based on previous study, Lu & Bi (2016) developed a two-tier multiple choice diagnostic instrument to evaluate the level of students' conceptual understanding of electrolytes concepts.

An interesting study is reported by Herrmann-Abell & DeBoer (2011), who uses multiple-choice items with distractor-based answer choices. The effect distractor is used to add to the strength of diagnostic items, so can be able to diagnose the structure of students' chemical misconceptions. The item design developed for this study is a two-level distractor-based multiple choice item, adapted from Treagust (1988), Chandrasegaran et al., (2007), Wilson (2008), Herrmann-Abell & DeBoer (2011), Hadenfeldt et al., (2013), and Lu & Bi (2016). Instruments diagnostically are used to reveal: the level of conceptual understanding of students, their knowledge, how students use that knowledge, and to find student misconceptions. Each item contains two questions, namely questions in the first tier (Q1) and in the second tier (Q2). Q1 to evaluate students' knowledge, and Q2 evaluate students' reasoning about concept that being measured on each item. Both Q1 and Q2, are designed with four options (answer choices). One of them is the right answer, and the other three are answer choices that are designed as a distractor. The distractors is answers that contain misconceptions and as if correct answer. This answer is identified from interviews and written tests that have been done. Studies by Sadler (1998), Herrmann-Abell & DeBoer (2011), Hadenfeldt et al., (2013), (Lu & Bi, 2016) has shown that the distractor choices that are designed based on student misconceptions can improve the diagnostic strength of multiple choice items. The choice of answers developed for this research has been designed based on distractor.

3. Misconceptions as a distractor

Commented [TR10]: rephrase

In general, the drawback of multiple choice items is that there is an opportunity for students to answer questions by guessing. This limitation can be minimized by provide a distractor on each item. Students with low ability may choose distractor option, because it meet their thinking. An example of the distractor option on item 1mA is shown in Figure 1.

It contains two interrelated questions, written 1mA (Q1 / Q2). This item measures students' ability to identify color changes of solution in the redox reaction, through macroscopic representation. Suppose a student chooses C option for Q1 and A for Q2. Then the answer is written as 1mA (C / A), and is the correct answer. Low ability students will choose the answer that they understand. Students' opportunity to guess the correct answer, in Q1 and Q2 are 0.25, respectively. This means by entering the distractor's choices can minimize guess answers and at the same time can be diagnosed students misconception patterns.

4. Data Collection

The third step is data collection using a two-tier multiple choice instrument, which is given to 1150 tenth grade students from eight senior high schools in the northern region of Sulawesi, Indonesia. Students are chosen randomly, and agree to be voluntarily respondents. All students have learned redox reaction concept according to the Chemistry Curriculum Standards in Indonesia. Process standards and learning outcomes have been elaborated completely in this curriculum. Chemistry teachers are required to develop and use problem solving-based instructional strategies. This strategy is used to encourage students to design and make redox reaction experiment, practice to solve problems, explain the process of electron transfer, in the reaction between iron metal dipped in a solution of CuSO₄ and H₂SO₄, discuss redox reaction equation, and determine elements that are undergoing a changes in oxidation number. Based on this curriculum, it can be understood that by learning process students may form their conceptual understand of redox reactions in a gradual, constructive, systemic and interrelated manner. Likewise, it can be explained when students have many misconceptions, indicating that students do not experience a whole and meaningful learning process. If these assumptions can be proven empirically, the instruments developed in this study have good predictive validity.

Data have been collected for three months, in each school. Students respond manually through a written answer sheet. The test time is 45 minutes and was supervised by teacher. All students are told to answer all questions in the instrument. After testing, all instrument texts are collected directly by the researcher. There are no instrument texts that are not collected.

5. Measurement Model

The fourth step is use the Rasch analysis model approach to calibrate the difficulty level of the item and students' ability level on the same interval scale. Raw data is inputted in Excel format in the form of dichotomous data, then converted into interval data, using WINSTEPS version 3.75 model software (Linacre, 2012; Bond and Fox, 2015). Raw data input considers the item model that containing two questions Q1 / Q2. Student responses on each item in the first tier (Q1) and in the second tier (Q2) were analyzed separately.

The Rasch model, combine an algorithm 'i' (as the results of probabilistic expectations of items) and students 'n', which are mathematically stated by Bond and Fox (2015), as: $P_{ni}(X_{ni} = 1 / (\beta_n, \delta_i)) = (e^{(\beta_n - \delta_i)}) / (1 + (\beta_n - \delta_i))$, where: $P_{ni}(X_{ni} = 1 / (\beta_n, \delta_i))$ is the probability of students "n" in items "i" to get a correct answer ($x = 1$); with student ability β_n , and difficulty level of items δ_i . The above equation can be further simplified by entering the logarithmic function and making it: so that the probability of success is: the probability of success equals to the ability of the student minus the difficulty level of the item. It is important to underline that students ability and items are stated at the same and independent intervals. The level of students' abilities and difficulty level of items is measured in units of logarithms called odds or logs, which can vary from -00 to +00 (Herrmann-Abell & De Boer, 2011; Sumintono & Widhiarso, 2015).

Rasch models can evaluate the level of understanding, reasoning of students and the difficulty level of items. To test efficiency of the instrument indirectly is by comparing the distribution of difficulty items with the level of understanding and reasoning of students. This is different from the classical test theory approach, which cannot diagnose the level of understanding of students, because the difficulty level of items is only based on raw data (Lu & Bi, 2016). Student with high ability will be able to answer items with a lower level of difficulty.

RESULTS AND DISCUSSION

1. Reliability and Validity of measures

1.1 Reliability

Reliability explains how far measurements produce information that is consistent in revealing latent traits. Can the TMDI instrument measure the unidimensional nature of latency, namely students' conceptual understanding in explaining the redox reaction (Sumintono & Widhiarso, 2015). The measurement reliability of the person item is based on the person separation index and item separation index. The Separation index can also be converted to Cronbach's equivalent value from 0-1. Table 2 shows a summary of fit statistics, which are compatible with the Rasch dichotomous model.

Tabel 2. Summary of fit statistics

	Person (N=1150)	Item (N=30)
Reliabilitas	0,58	0,99
Infit MNSQ	1.00	1,00
Outfit MNSQ	1.03	1.03
Separation index	1.18	10,14
Cronbach Alpha	0,61	-

From Table 2, it is known that the person separation index is 1.18-which is equivalent to the reliability value of person 0.58. This value indicates that the reliability of the person is relatively weak. This means that the consistency of students' responses to TMDI test items is relatively weak- Cronbach Alpha Coefficient value of 0.61 indicates that the interaction between 1150 students with 30 items is fairly adequate. However, this fact will not affect the teacher's decision to develop students' abilities to be better (Wei, Liu, Wang, & Wang, 2012; Lu & Bi, 2016). The item separation index value (10.14) is equivalent to the reliability value (0.99). This means that the consistency of the item is classified as very good (special), or the item meets the requirements of unidimensionality. This indicates that the item is able to define latent variables very well. The infit and outfit values of most items are within the acceptable range for multiple choice tests (Abell & DeBoer, 2011; Bond and Fox, 2015).

1.2 Validity

Validity describe how well measurements are carried out according to what should be measured. The validity of the TMDI test was tested by the suitability analysis of the Rasch model item. If the test is able to measure four levels of a student's ability to explain the redox reaction, then this test has good construct validity (Linacre, 2012). The results of the analysis are presented in Table 3.

Table 3. Item Fit Statistics^a

Item	Measure	Model SE	INFIT		OUTFIT		PTMEA Corr
			MNSQ	ZSTD	MNSQ	ZSTD	
1maA(Q1)	-0,73	0,06	0,93	-3,9	0,92	-4,0	0,41
1maA(Q2)	-1,00	0,06	0,94	-3,0	0,91	-3,5	0,40
2baA(Q1)	-1,02	0,06	0,84	-7,7	0,81	-7,8	0,53
2baA(Q2)	-0,56	0,06	0,95	-2,9	0,93	-3,5	0,39
3kaAa(Q1)	0,18	0,06	1,01	0,6	1,05	1,7	0,25
3kaAa(Q2)	0,93	0,07	1,07	1,7	1,12	2,1	0,13
4kaAb(Q1)	-1,43	0,07	0,89	-4,1	0,80	-5,7	0,47
4kaAb(Q2)	-0,15	0,06	0,90	-5,4	0,89	-4,8	0,45
5mbB(Q1)	0,80	0,07	1,04	1,0	1,07	1,4	0,19
5mbB(Q2)	-0,71	0,06	0,96	-2,4	0,95	-2,6	0,37
6kaBb(Q1)	-0,51	0,06	0,88	-6,8	0,86	-7,3	0,48
6kaBb(Q2)	-0,44	0,06	0,88	-6,8	0,87	-6,6	0,48
7akBa(Q1)	0,46	0,07	1,15	5,0	1,23	5,5	0,03
7akBa(Q2)	0,99	0,08	1,16	3,6	1,36	5,6	-0,05
8abB(Q1)	0,17	0,06	1,04	1,8	1,07	2,1	0,22
8abB(Q2)	0,52	0,07	1,04	1,5	1,05	1,2	0,20
9abBa(Q1)	-0,40	0,06	0,88	-7,2	0,88	-5,9	0,48
9abBa(Q2)	1,27	0,08	1,13	2,5	1,41	5,2	-0,05
10sbBsa(Q1)	0,25	0,07	0,85	-6,4	0,83	-5,3	0,50
10sbBsa(Q2)	-0,35	0,06	0,96	-2,1	0,95	-2,5	0,36
11maC(Q1)	1,21	0,08	1,16	3,0	1,33	4,5	-0,05
11maC(Q2)	-0,88	0,06	0,89	-6,0	0,86	-6,5	0,48
12sbC(Q1)	0,97	0,08	1,06	1,4	1,11	1,9	0,14
12sbC(Q2)	-0,34	0,06	1,22	9,9	1,22	9,9	-0,01
13D(Q1)	0,31	0,07	1,05	2,0	1,08	2,3	0,19
13D(Q2)	-0,34	0,06	1,12	6,2	1,13	6,0	0,13
14E(Q1)	0,10	0,06	0,94	-2,5	0,93	-2,5	0,37
14E(Q2)	0,14	0,06	0,94	-2,6	0,95	-1,6	0,36
15F(Q1)	0,22	0,07	0,96	-1,8	0,97	-0,8	0,34
15F(Q2)	0,37	0,07	1,18	6,3	1,20	5,2	0,01

^aN = 1150. MNSQ values of 0.70–1.30 indicate the acceptable fit range; infit (weighted) or outfit (unweighted) values outside this range indicate poor fit of the data to the Rasch model.

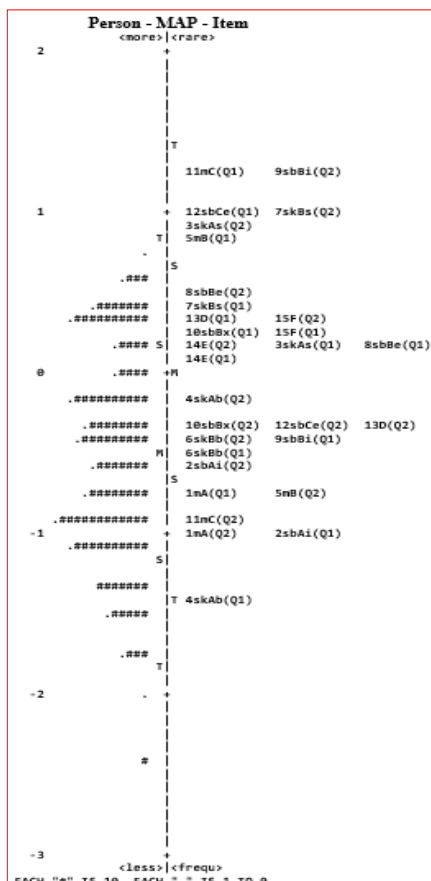


Figure. 2 Item-person map for the 30 items (Q1 and Q2)

Quality of item fit is to examine whether the test item is functioning normally or not. Items that are not fit indicate student misconception on that item. The indicator used is the value of Mean Square Residual (MNSQ). Outfit MNSQ and infit MNSQ are sensitive chisquare to detect outlier responses. This value is to show how much impact the item incompatibility has. Outlier responses are often correct answers to items that are difficult by students with low ability due to guessing, or wrong answers to items that are easier by students who are more capable, due to carelessness.

The ideal value expected from MNSQ is 1.0. In addition, another indicator is point measure correlation (PTMEA Corr). PTMEA Corr is a score and person measure score correlation, whose value must be positive and not close to zero (Bond and Fox, 2015). The criteria used is the means-square outfit (MNSQ) value: $0.5 < y < 1.5$; z-standard outfit: $-2.0 < Z < +2.0$, and PTMEA Corr: $0.4 < x < 0.8$. If the three criteria are not fulfill, the items are not good enough therefore need further investigation (Boone, Yale, & Staver, 2014). Based on Table 3, it is known that no item does not meet the three criteria. However, there are four items, namely 12sbC

(Q2), 11mC (Q1), 9skBi (Q2), 7skBs (Q2)) which are negative, and item 15F (Q2) which is close to zero. The five items require further investigation

After testing the suitability of the item, we examine level of difficulty of the item with the level of ability of students, using "Wright Map" (person-item map), is presented in Figure 2. This map shows the distribution of students' abilities (on the left side) and the distribution of difficulty items (on the right side) vertical lines. Distribution of low-ability students and low difficulty items are at the bottom of the map. Distribution of high-ability students and items with high difficulty levels are at the top of the map. The item's horizontal line mean (0.0 logit) is higher than the person mean (-0.52 logit). All items can reach the entire scale of the student's logit ability. However, there are six items that are too difficult for students, above the 1 SD scale (standard deviation). They are 5mB/Q1 (0.80), 11mC/Q1 (1.21), 12sbC/Q1 (0.97), 3skAs/Q2 (0.93), 7skBs/Q2 (0.99), and 9skBi/Q2 (1.27). These items need to be adjusted and examined further.

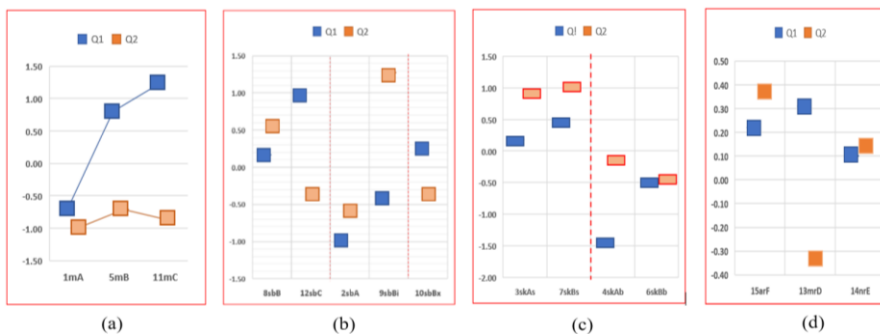


Figure 3. Comparison of the size of items in the first tier (Q1) and in the second tier (Q2)

2. Level of conceptual understanding

The level of conceptual understanding of students is distinguished by the level of knowledge (Q1) and the level of reasoning (Q2) in each construct studied (item). This difference is explained based on the comparison of item size (measures items). Students' misconceptions are examined by testing response patterns and option probability curve patterns. The response pattern test is estimated from the results of the distractor test, which is based on the acquisition of the average ability of the item. If the average ability value rises, the effect of the distractor works. This means that students give the answers are not by guessing. However if the average ability value drops, the effect of the distractor does not work meaning that the item is easy to guess. Student's answers are not representing the knowledge they have learned, because students guess the answer. The decrease in average ability value indicates that there are low-ability students who are able to answer items correctly and / or there are high-ability students who answer incorrectly on items that should be answered correctly, and this is not favorable for measurement (Smiley, 2015).

The option probability curve is used to explore the pattern of student responses. Curves that correspond to the probability of the correct answer will usually increase, and curve due to distractor will usually decrease with the increase in students' abilities (Haladyna, 2004). In general, for items that are driven by a distractor, the curve obtained does not correspond to the probability curve of the correct answer (Sadler, 1998). Therefore, each student's answer are analyzed separately, to represent data more accurately and provide additional information about the aspects of misconception students have.

The probability choice answer curve provides information on the possibility of students choosing on each answer option provided, as a function of the level of knowledge and the level of reasoning in the construct being measured. In this case, there are four curves, where one curve represents one answer. It means that students who choose one answer choice, have certain conceptual understanding of the construct being measured. Low-ability students related to measured constructs will be interested in certain misconceptions, and high-ability students will be attracted to other misconceptions (Herrmann-Abell & DeBoer, 2011). In addition, the probability curve is presented through visual images, to show the distribution of the correct answers and those experiencing misconceptions in the whole of the spectrum of students' conceptual understanding. Thus it can be identified which form of the curve is appropriate, or is not appropriate, as an indication of the misconception of an item. Forms of curves that are not appropriate, can also show a certain

structure of misconceptions, with patterns that are repetitive, consistent and will disappear sequentially, along with increasing student abilities. Diagnosis of misconception structures in an accurate way, allows teachers to develop more effective learning strategies. In this article, an example of how Rasch modeling and the probability option curve can be used to reveal the structure of students' misconceptions in explaining the types of redox reactions.

2.1 Level of macroscopic reasoning

The level of macroscopic reasoning is the students' ability to describe the color change of the solution due to A, B, and C the redox reaction processes, that was evaluated with items 1mA, 5mB, and 11mC. In the first tier (Q1) students are asked to determine what changes they know, if reactions A, B and C occur. In the second tier (Q2) students are asked to state the reason for the answer in the first tier. The difference in the level of knowledge and reasoning of students on each item is determined by the size of the item. The larger the item size, the higher the difficulty level of the item, the lower the level of knowledge and reasoning of students.

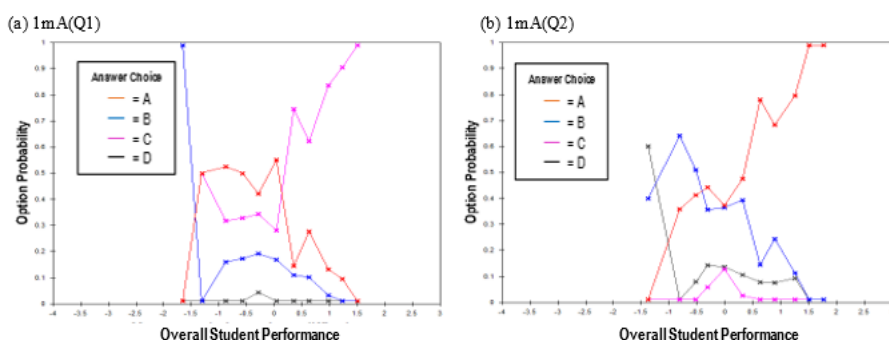


Figure 4 Option Probability Curve (a) item 1mA(Q1); (b) item 1mA(Q2)

Figure 3a shows that the size of items in the first tier (Q1) is in the range of -0.55 logit to 1.25 logit, while in the second tier (Q2), no more than -0.50 logit. It means, empirically the level of student knowledge is lower than the level of reasoning in identifying the changes colors of the solution of A, B and C redox reaction at the macroscopic representation level. To explain this fact, -option probability curve is used for item 1mA (Figure 4).

For curve 4 (a), low-ability students (between -1.5 to 0.5 logit) tend to choose the distractor answer option, namely A (iron will not dissolve in CuSO_4), and B (Iron dissolved in CuSO_4 , such as as sugar dissolves in water). It is seen that the peak of the A curve is more dominant than B. Students with moderate ability levels (more than 0.5 logit) tend to choose the right answer C (the solution turns reddish). This pattern of misconception shows evidence of the lack of acquisition of student knowledge about the redox reaction, and develops logical misconceptions, iron is insoluble in CuSO_4 , due to the hard properties of iron. Through the 4c curve, it is known that students with ability levels (-1.5 to 1.5 logit) are more dominant choosing the answer distractor B option (Iron does not react with CuSO_4 solution). Students with an ability level of more than 0.5 logit, tend to choose the correct answer option A (iron undergoing an oxidation reaction). It appears that there is the same student response pattern, stating that iron is insoluble in CuSO_4 . This pattern of misconception can be understood, because students with limited levels of knowledge, incomplete learning experiences, tend to choose to what is physically visible. This is an alternative framework of students (Johnstone, 1991; Lu & Bi, 2016; Taber, 2013), understanding that contains misconceptions. This result support the studies reported by Chandrasegaran et al (2007) and Herrmann-Abell & DeBoer (2011). This result also describe that macroscopic factual knowledge and students' rational reasoning are out of sync. It is possible that students never make direct observations through experiments to identifying the color changes in the solution of the redox reaction process.

2.2 Submicroscopic reasoning level

The level of submicroscopic reasoning is related to student ability to explain the redox reaction at the particle level. Based on item size (Figure 3b), interesting facts were found, first, in the first tier question (Q1), the item size 12sbCe (0.97) is greater than 8sbBe (0.17), while in the second tier (Q2), item size 12sbCe (-0.24) is smaller than 8sbBe (0.52). This means that the level of student knowledge is lower than the level of reasoning, in explaining sub-microscopically the electron transfer process in the redox B and C reactions. Second, in the first tier (Q1) item size is 2sbAi (-1.02) and 9sbBi (-0.40) very low, so in the second tier (Q2) items 2sbAi (-0.56) are low. That is, the level of students' knowledge and reasoning in expressing the net ion reaction equation through the interpretation of the submicroscopic diagram of the redox A and B reactions, is very low. Unless the item size is 9sbBi (1.27), which is very high, it's hard to explain because this item's distractor option doesn't work. Third, the size of items is 10sbBx in the first tier (0.25) while in the second tier (0.35). This means that the level of student knowledge is lower than the level of reasoning, in determining the submicroscopic diagram of the reaction of Zink metal dipped in CuSO_4 solution, if the reaction is complete (irreversible). These three facts prove that submicroscopically, students do not have good knowledge and reasoning. There is a gap between the level of knowledge and the level of reasoning. Conceptual understanding of students tends to be inconsistent and partial.

To explain how students develop their understanding at the submicroscopic level, it can be exemplified through the analysis of item response patterns 2sbAi. This item tests the ability of students to interpret the submicroscopic diagram, the redox A reaction, which is iron metal dipped in a solution of CuSO_4 (Figure 5). Based on Figure 5, in the first tier question (Q1), students are asked to write net ion reaction equations, as in the second tier (Q2), students are asked to give the reasons of answers in the first tier. The option probability curve form for item 2sbAi is shown in Figure 6.

Curve 6a, shows that students with low ability levels (smaller 0.0 logit than the ability of students as a whole) tend to choose the answer distractor C ($\text{Fe}^{2+}_{(\text{s})} + \text{Cu}_{(\text{aq})} \rightarrow 2\text{Fe}_{(\text{aq})} + \text{Cu}_{(\text{s})}$) and B ($\text{Fe}_{(\text{s})} + \text{Cu}_{(\text{aq})} \rightarrow \text{Fe}^{2+}_{(\text{aq})} + \text{Cu}_{(\text{s})}$). C and B are answers that contain misconceptions. Students with a level of ability greater than 0.0 logit, choose the correct answer A ($\text{Fe}_{(\text{s})} + \text{Cu}^{2+}_{(\text{aq})} \rightarrow \text{Fe}^{2+}_{(\text{aq})} + \text{Cu}_{(\text{s})}$). On the other hand, on the 6b curve, students with low ability levels (less than -0.5 logit) tend to choose C option (reaction of iron and CuSO_4 is a redox reaction). Students with abilities ranging from -1.5 to 1.0 logit, tend to choose A (iron accepts electrons from CuSO_4). C and A contains misconceptions. Students with abilities above -0.5 tend to choose the correct answer B (Cu^{2+} ions are reduced to Cu).

Student response patterns on curves 6a and 6b indicate that students have low ability at submicroscopic reasoning. Although the diagram of electron transfer has been given, it cannot help students, because they are more likely to take to what they understand about oxidation and reduction. This is also found in item 9sbi, students are more difficult to express the net ion reaction equation of B reaction than A. In fact, these two items qualitatively have the same degree.

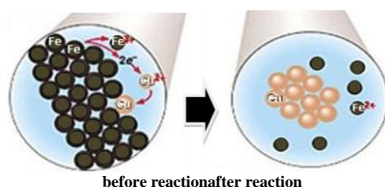


Figure 5. Submicroscopic Diagram of Transfer electron process between iron metal and CuSO_4 Solution

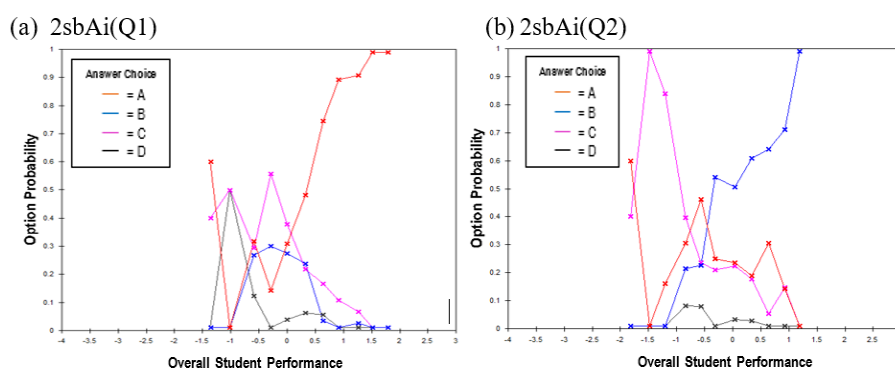


Figure 6. Option probability Curve Kurva: (a). item 2sbAi(Q1), and (b) item 2sbAi(Q2)

2.3 Level of symbolic reasoning

The level of symbolic reasoning is evaluated based on students' ability to determine: (a) species that act as oxidizing and reducing agents (items 3skAs and 7skBs); and (b) elements undergoing oxidation number changes (items 4skAb and 6skBb). Based on item size comparisons (Figure 3c), interesting findings can be explained, as follows: first, item size Q1 (3skAs: 0.18 and 7skBs: 0.46) is greater than item Q1 (4skAb: -1.43 and 6skBb: -0.44). That means, students are more difficult to determine which species act as oxidizing and reducing agents rather than determining the elements that undergo changes in oxidation numbers. Second, in terms of item size Q2, the highest difficulties for students are 7skBs (0.99), 3skAs (0.93), 4skAb (-0.15), 6skB (-0.44) respectively. This means students are more difficult to give a reason in determining the oxidizing / reducing species rather than determining the element that undergoes oxidation number changes. This fact is interesting, because the understanding of the concept of oxidation number and determination of species which act as oxidizers and reducing agents, are interrelated. To elaborate this problem, student response patterns was analyzed, and was exemplified through items 6BB and 7KBs (Figure 7).

(a) 6skBb	(b) 7skBs
Q1. Identify oxidation number change of Zn, in reaction between Zinc metal dipped in CuSO_4 solution	Q1. Identify reducing agent in reaction between Zinc metal dipped in CuSO_4 solution
A. 0 to -2. B. +1 to +2. C. +2 to 0. D. 0 to +2	A. Zinc powder B. CuSO_4 solution
Q2. Reason of your	C. Copper D. ZnSO_4 solution
A. $\text{Zn}^0_{(s)} + \text{Cu}^{+2}\text{SO}_4^{+2}_{(aq)} \rightarrow \text{Zn}^{+2}\text{SO}_4^{+2}_{(aq)} + \text{Cu}^0_{(s)}$	Q2. which one of the following is the reason of your answer!
B. $\text{Zn}^0_{(s)} + \text{Cu}^{+2}\text{SO}_4^{+2}_{(aq)} \rightarrow \text{Zn}^{+2}\text{SO}_4^{+2}_{(aq)} + \text{Cu}^0_{(s)}$	A. Zinc atom undergo oxidation and becomes Zn^{2+} ion
C. $\text{Zn}^0_{(s)} + \text{Cu}^{+2}\text{SO}_4^{+2}_{(aq)} \rightarrow \text{Zn}^{+2}\text{SO}_4^{+2}_{(aq)} + \text{Cu}^0_{(s)}$	B. Cu^{2+} Ion undergo reduction and becomes copper atom
D. $\text{Zn}^0_{(s)} + \text{Cu}^{+2}\text{SO}_4^{+2}_{(aq)} \rightarrow \text{Zn}^{+2}\text{SO}_4^{+2}_{(aq)} + \text{Cu}^0_{(s)}$	C. Reaction between Zn and Copper (II) sulphate produce copper deposit
	D. Copper (II) sulphate is product of Zinc that undergo reduction

Figure 7. (a) Item 6skBb(Q1/Q2), (b) Item 7skBs(Q1/Q2)

Questions in the first tier (Q1) item 6skBb, basically testing the ability of students to write equilibrium reaction equations of Zinc metal dipped in CuSO_4 , and determine how much Zn oxidation number is in the

reaction. In the second tier (Q2), students are asked to determine the oxidation number of Zn in the reaction equation. The option probability curve of this item is shown in Figure 8. Curve 8 (a), shows that students with the lowest ability (smaller than -1.0 logit) tend to choose the misconceptions answer A, B and C. The highest probability of choice is A, then decreases with increasing probability of high-ability students (more than 0.5 logit), who choose the right answer D. It means, students' misconceptions, especially in understanding the changes in oxidation number of Zn (from 0 to -2), Zn experiences reduction.

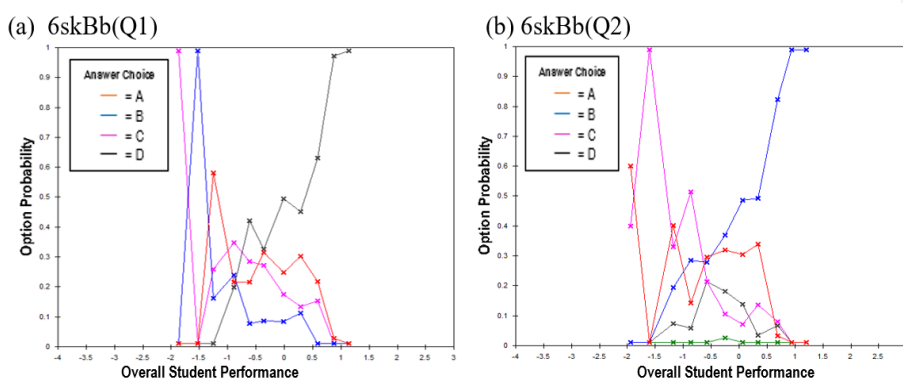


Figure 8. Option probability curve (a). item 6skBb(Q1), and (b).item 6skBb(Q2)

Curve 8(b), confirms this misconception, where low-ability students (small from -0.5 logit), tend to choose C and A, while the probability of students choosing the right answer option B, above ability from -0.5, 5 logit. This fact shows the students' weak understanding regarding the concept of oxidation number. To test this fact, the evaluation continued with the ability of students to determine which species acted as reducing agents in the reaction of Zinc and CuSO_4 , through item 7skBs, presented in Figure 7 (b).

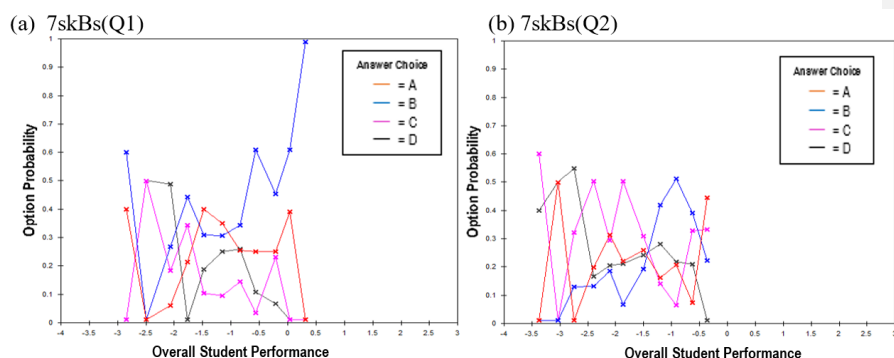


Figure 9. Option probability curve (a). item 7skBs(Q1), and (b). item 7skBs(Q2)

Based on the probability curve item 7skBs (Q1) in Figure 9 (a), it is known that the choice of answer B, which contains misconceptions, is the answer choice with the highest probability chosen by the lowest ability students (-3.0 logit) to 0.5 logit. This curve rises to reach a peak of 1.0 logit. Students argue that the reducing agent in the reaction between Zn and CuSO₄ is a solution of CuSO₄. The probability of students choosing the right answer A (reducing agent is Zn powder) was chosen by low-ability students, -3.0 logit to 0.3 logit. However, the curve actually shows unusual shape, have two peaks then decreasing. The first peak, at the level of students' ability is smaller than -1.0 logit, and the second peak, at the level of students' ability is greater than 0.0 logit. On the other hand, the 7skBs (Q2) curve in Figure 9 (b) shows that there is no answer of students that reach option probability values up to 1.0 logit. The average of student ability is only able to reach a value of 0.6 logit option probability, from a scale of -3.5 logit to -0.4 logit. These interesting facts can be understood, because they support the previous facts, namely students understanding about the concepts of reduction, oxidation, and oxidation numbers is weak. Students are not able to determine which species act as reducing agents in the reaction of Zinc and CuSO₄.

2.4 Macro, submicro and symbolic levels of reasoning

Macro, submicro and symbolic reasoning levels are evaluated based on students' ability to interpret and convert macro, submicro and symbolic representations, to explain three types of reactions, namely D, E and F. The letter "D" is a symbol of an item that presents a reaction to iron metal; "E" is a symbol of non-redox reaction (i.e., silver nitrate and potassium chloride solution); and F is the auto-redox reaction symbol. These three types of reactions are evaluated, to confirm students' ability to relate their knowledge and reasoning to the three levels of representation.

Figure 3d show the comparisons of sizes 13D, 14E, and 15F. In the first tier (Q1), the difficulty level of the item is 13D > 15F > 14E. That means, students are more difficult to interpret rusting iron reactions, rather than auto-redox reactions and non-redox reactions. In the second tier (Q2), the difficulty level of the item is 15F > 14E > 13D. This means that students are more difficult to explain the auto redox reaction process, rather than a non-redox reaction and a rusting iron reaction. This fact is really interesting, especially item 13D, because between Q1 and Q2 there is a very wide gap. This item asks students to analyze a diagram that illustrates the redox reaction of rusting iron, and then converts to equivalent chemical reaction equation. The result is that students do not have difficulty giving logical reasons related to the iron rusting process with the reaction equation (Q2), but they are very difficult to convert submicro diagrams into reaction equations (Q1). The results of the response pattern analysis found that 34% of students stated the reaction of iron rusting is $4\text{Fe} + 3\text{O}_2 \rightarrow 2\text{Fe}_2\text{O}_3$ and 48% of students stated that the metal was reduced and oxygen was oxidation. In fact, the correct reaction is: $4\text{Fe} + 3\text{O}_2 \rightarrow 2\text{Fe}_2\text{O}_3$, because iron metal undergoes oxidation, oxygen undergoes reduction. This evidence support previous findings that students' understanding of oxidation, reduction and oxidation numbers concepts is very weak. Then in item 14E, students were asked to analyze the submicroscopic diagram of the reaction process of a solution of silver nitrate and potassium chloride solution which produced white deposits: $\text{AgNO}_3 + \text{KCl} \rightarrow \text{AgCl} + \text{KNO}_3$, and asked whether the reaction is redox reactions or not. The results showed that the size of items 14E / Q1 and 14E / Q2 is relatively not much different. That means, students are able to understand and have logical reasons for determining non redox reaction. However, there are still 25% of students stating that the reaction of $\text{AgNO}_3 + \text{KCl} \rightarrow \text{AgCl} + \text{KNO}_3$ is a reduction reaction, and 33% of students have reason that there is a change in oxidation number and electrons transfer in the reaction. This fact makes it clear that students are not able to distinguish redox and not redox, due to confusion in determining which species act as oxidizers and reducing agents.

In item 15F, students were asked to analyze submicroscopic diagrams which illustrated the reaction of bleach powder mixed with HCl ($\text{CaOCl}_2 + \text{HCl} \rightarrow \text{CaCl}_2 + \text{Cl}_2 + \text{H}_2\text{O}$), then mixed with water, and produced hydrochloric acid and hypochlorous acid ($\text{Cl}_2 + \text{H}_2\text{O} \rightarrow \text{HCl} + \text{HClO}$). The results of the analysis show that the size of the item 15F/Q2 is more difficult than 15F/Q1. That means, students are very difficult to give opinions regarding the auto-redox reaction diagram, but can determine the auto redox reaction equation. 44% of students are sure that the reaction: $\text{Cl}_2 + \text{H}_2\text{O} \rightarrow \text{HCl} + \text{HClO}$ is a redox reaction, and 30% of students have reason that the reaction occurs because there is electrons transfer. The correct answer is that the reaction ($\text{Cl}_2 + \text{H}_2\text{O} \rightarrow \text{HCl} + \text{HClO}$) is an autoredox reaction, because of the oxidation and reduction reactions of the same

substance. This result is evidence that student misconceptions tend to be consistent, related to the understanding and reasoning of the concepts of oxidators and reducing agents, changes in oxidation numbers.

CONCLUSION

The diagnostic instrument developed has construct validity, which can be used to evaluate the level of students' conceptual understanding, including knowledge and reasoning, and diagnosing students' misconceptions in explaining the redox reaction. At the level of macroscopic reasoning, a repetitive student response was found, which states that iron is insoluble in CuSO_4 . For low-ability students, this pattern of misconception can be understood, because the limited level of knowledge and reasoning makes students tend to stick to what is physically visible. This misconception may develop from learning experiences that do not go through the process of direct observation or through experimentation. In the level of submicroscopic and symbolic reasoning, there is a gap between the level of knowledge and the level of reasoning of students. This gap is detected again at the level of macro, submicro and symbolic reasoning. Because the acquisition of student knowledge is not complete, it is difficult for students to explain the redox reaction macroscopically, submicroscopically and symbolically. Students have difficulty translating representational diagrams, and converting them into reaction equations. The difficulties lead students to develop misconceptions that are based on their intuitive reasoning, by choosing a choice of distractor answers that are misconceptions. The choice of this answer is detected repeatedly, and tends to come from the same understanding structure, namely the weak of students understanding of reduction, oxidation and oxidation number concept. In turn, all available evidence shows that the progress of the learning outcomes of the students' in redox reactions concepts, overall tends to be low.

Combining the item development process qualitatively with Rasch analysis models, is relatively effective to depth evaluation of the progress of student learning outcomes, reveals students' ways of thinking that are misconceptions, and know what students have not understood. Differences in item size and option probability curves illustrate that the prevalence of the most dominant misconceptions is low ability students. But for certain cases, the prevalence even increases, because the misconception factors tend to be resistant, as seen in item 9sbBe (Q2). The availability of this type of information makes it possible for teachers to classify student understanding based on the level of misconception, so that it is useful in developing instructional strategies. In addition, with the option probability curve on each item, it can be identified an unusual curve, which indicates a problem with the item.

The multiple choice distractor items in this study were developed by adapting some of the results of previous interviews and studies, which were aligned with the targets outcome of chemical learning in Indonesia. The hope is that further research can be carried out, for the development of other material. Although this study does not answer the question of why students experience misconceptions, it is possible for similar items to be developed to help teachers diagnose students' ways of thinking and understanding patterns, so that chemical learning becomes more effective and meaningful.

ACKNOWLEDGMENTS

The highest appreciation and thanks are conveyed to the Directorate of Research and Community Service, the Directorate General of Research and Development, the Indonesian Ministry of Research, Technology and Higher Education, which has funded this research.

REFERENCES

- 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>
- Boone, W. J., Yale, M. S., & Staver, J. R. (2014). *Rasch Analysis in the Human Sciences*. Rasch Analysis in the Human Sciences. Springer Dordrecht Heidelberg London New York Library: Springer Dordrecht Heidelberg London New York Library. <https://doi.org/10.1007/978-94-007-6857-4>
- Briggs, D. (2009). The psychometric modeling of ordered multiple-choice item responses for diagnostic assessment with a learning progression. *Learning Progressions in Science (LeaPS) Conference, June 2009, Iowa City, IA*, (June).
- 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.
- Chi, S., Wang, Z., Luo, M., Yang, Y., & Huang, M. (2018). Student progression on chemical symbol

- representation abilities at different grade levels (Grades 10–12) across gender. *Chemistry Education Research and Practice*, 19(4), 1055–1064.
<https://doi.org/10.1039/c8rp00010g>
- Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The Perspectives of Chemists. *Science Education*, 93(1), 56–85.
<https://doi.org/10.1002/sce.20292>
- 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. (1999). Improving Teaching and Learning through Chemistry Education Research: A Look to the Future. *Journal of Chemical Education*, 76(4), 548.
<https://doi.org/10.1021/ed076p548>
- Hadenfeldt, J. C., Bernholt, S., Liu, X., Neumann, K., & Parchmann, I. (2013). Using Ordered Multiple-Choice Items To Assess Students' Understanding of the Structure and Composition of Matter. *Journal of Chemical Education*, 90(12), 1602–1608.
<https://doi.org/10.1021/ed3006192>
- Haladyna, T. M. (2004). Developing and Validating Multiple-choice Test Items (3rd Editio, p. 320). New Jersey: Lawrence Erlbaum Associates.
<https://doi.org/10.4324/9780203825945>
- 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>
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2016). Silabus Mata Pelajaran Kimia. In *Silabus Mata Pelajaran Sekolah Menengah Atas/Madrasah Aliyah (SMA/MA)*.
- Klassen, S. (2006). Contextual assessment in science education: Background, issues, and policy. *Science Education*, 90(5), 820–851. <https://doi.org/10.1002/sce.20150>
- Linacre, J. M. (2012). *A User's Guide to WINSTEPS® MINISTEP Rasch-Model Computer Programs Program Manual 3.75.0*. Winsteps.com.
<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>
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (Eds.). (1999). *Assessing Science Understanding: A Human Constructivist View*. Academic Press, San Diego, California : Elsevier 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
- National Research Council. (2001). *Knowing what students know: The Science and Design of Educational Assessment*. National Academy Press (Vol. 19).
<https://doi.org/10.17226/10019>
- 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/ijel.vol2no1.3>
- Rahayu, S. (2017). Promoting the 21st century scientific literacy skills through innovative chemistry instruction (Vol. 020025, p. 020025). Published by AIP Publishing. <https://doi.org/10.1063/1.5016018>

- 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.
- Smiley, J. (2015). Classical test theory or Rasch- A personal account from a novice user. *Shiken*, 19(1), 16–29.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Penerbit Trim Komunikata.
- Taber, K. S. (2009). Challenging Misconceptions in the Chemistry Classroom: Resources to Support Teachers. *Educació Química EduQ*, 4, 13–20.
<https://doi.org/10.2346/20.2003.02.27>
- Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, 14(2), 156–168. <https://doi.org/10.1039/c3rp00012e>
- 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*. Routledge Taylor & Francis Group (Third Edit, Vol. 44). New York and London: Routledge Taylor & Francis Group.
<https://doi.org/10.1088/1751-8113/44/8/085201>
- Tsaparlis, G. (2009). Linking the Macro with the Submicro Levels of Chemistry: Demonstrations and Eksperiments that can Contribute to Active/Meaningful/Conceptual Learning. In S. A. Devetak, Iztok and Glazar (Ed.), *Learning with Understanding in the Chemistry Classroom* (pp. 41–61). Springer Dordrecht Heidelberg London New York Library.
<https://doi.org/10.1007/978-94-007-4366-3>
- 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. (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). Rotterdam: Sense Publishers.
<https://doi.org/10.1007/978-94-6091-824-7>



Lukman Laliyo <lukman.laliyo019@gmail.com>

[IJLTER] Editor Decision

15 messages

IJLTER .ORG <editor@ijlter.org>

Sun, Sep 22, 2019 at 10:31 AM

To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Cc: "Deasy N. Botutihe" <transbahasa.ilmiah@gmail.com>, Citra Panigoro <transbahasatb@gmail.com>

Lukman Abdul Rauf Laliyo:

We have reached a decision regarding your submission to International Journal of Learning, Teaching and Educational Research, "THE DEVELOPMENT OF TWO-TIER INSTRUMENT BASED ON DISTRACTOR TO ASSESS CONCEPTUAL UNDERSTANDING LEVEL AND STUDENT MISCONCEPTIONS IN EXPLAINING REDOX REACTIONS".

Our decision is to: accept the paper if the requested changes are made - esp. the whole paper must be proof-read to improve the English language

IJLTER .ORG
editor@ijlter.org

Reviewer B:

Paper length::
Ok

Originality::
Acceptable

Scope of paper::
Relevant to IJLTER

Related work::
Acceptable

Language::
Appropriate and good language

References::
Quite wide and commensurate with the work presented.

Additional comments along the following lines: originality, literature review, methodology, evaluation of results, research implications, quality of communication, etc.:

The limit of the job can be the excessive specificity but it is however interesting.

Reviewer C:

Paper length::
Quite long

Originality::
Acceptable

Scope of paper::
Highly relevant

Related work::
Acceptable

Language::

Verify and correct the English text, because are some errors, i.e. probabilitycurve, conclusion (write „conclusions") and so on.

References::

The bibliography corresponds to the theme covered.

Additional comments along the following lines: originality, literature review, methodology, evaluation of results, research implications, quality of communication, etc.:

Rewrite the introduction, focusing on the purpose and objectives of the research, on the research approach. Then, after the introduction, enter a new section „Theoretical framework of the research".
Verify and write the chemical formulas correctly; for example, CuSO₄ and H₂SO₄.

Reviewer D:

Paper length::

Originality::

Scope of paper::

Related work::

Language::

References::

Additional comments along the following lines: originality, literature review, methodology, evaluation of results, research implications, quality of communication, etc.:
see attachment 5889

International Journal of Learning, Teaching and Educational Research
<http://ijlter.org/index.php/ijlter>



1650-5889-1-RV(1).docx
2996K

IJLTER ORG <ijlter.org@gmail.com>

Sun, Sep 22, 2019 at 10:32 AM

To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Cc: "Deasy N. Botutihe" <transbahasa.ilmiah@gmail.com>, Citra Panigoro <transbahasatb@gmail.com>

Dear Authors,

The decision for your paper is: **accept the paper if the requested changes are made**

Editor's Note

1. Format the paper strictly according to the template (attached).
2. Follow APA style strictly for the references. References must strictly follow APA otherwise publication will be delayed.

<http://www.apastyle.org/>

http://student.ucol.ac.nz/library/onlineresources/Documents/APA_guide_2015.pdf

An APA Manual is also attached for your convenience.

3. Check if references are mentioned in the text (and vice-versa). The paper must have at least 20 references. Failure to correct the references will delay the review process.

Also include recent papers (2015-2019) in the references (if need be).

Any non-English words must have their equivalent inside square brackets [].

4. Address all the concerns of the reviewers, as far as appropriate.

5. Mention the doi of each reference where available.

a. Go to <http://search.crossref.org/>

b. Copy and paste the title of the paper in the search box

c. Press the Enter key.

d. Copy and paste the doi back to the paper (into the reference section) if one is available. DOI may not be available for some of the references. This is fine.

e. There are two ways you can write the doi (1) doi:10.1037/a0028240 or (2) <http://dx.doi.org/10.1037/a0028240>

6. Prepare a separate word document to indicate the changes that were made as a result of each reviewer's comments. This is compulsory. The revise paper will not be considered without this.

7. Paper must be at least 5000 words (inclusive of references).

Failure to make the requested changes will significantly delay the publication of the paper.
If you have any queries, please let us know.

Prof. Antonio Sprock
CE

 **2019_IJLTER_Journal_Template.docx**

[Quoted text hidden]

--

Editorial Office

International Journal of Learning, Teaching and Educational Research

ISSN: 1694-2116 (Online)

ISSN: 1694-2493 (Print)

Website: <http://ijlter.org/>


Email: ijlter.org@gmail.com

 **APA_Guide_2017.pdf**
918K

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: transbahasa.go@gmail.com

Mon, Sep 23, 2019 at 2:29 PM

[Quoted text hidden]

 **1650-5889-1-RV(1).docx**
2996K

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: transbahasa.go@gmail.com

Mon, Sep 23, 2019 at 2:30 PM

Ini pak Lukman punya hasil review, sir.

----- Forwarded message -----

From: **IJLTER ORG** <ijlter.org@gmail.com>

Date: Sun, Sep 22, 2019, 10:32

Subject: Re: [IJLTER] Editor Decision

To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Cc: Deasy N. Botutihe <transbahasa.ilmiah@gmail.com>, Citra Panigoro <transbahasatb@gmail.com>

[Quoted text hidden]

 **APA_Guide_2017.pdf**
918K

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: transbahasaid@gmail.com

Wed, Oct 2, 2019 at 11:25 AM

----- Forwarded message -----

From: **IJLTER ORG** <ijlter.org@gmail.com>

Date: Sun, Sep 22, 2019 at 10:32 AM

Subject: Re: [IJLTER] Editor Decision


To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Cc: Deasy N. Botutihe <transbahasa.ilmiah@gmail.com>, Citra Panigoro <transbahasatb@gmail.com>

[Quoted text hidden]

2 attachments

 **APA_Guide_2017.pdf**
918K

 **1650-5889-1-RV(1).docx**
3309K

IJLTER ORG <ijlter.org@gmail.com>

Sat, Oct 5, 2019 at 7:53 PM

To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>

Cc: "Deasy N. Botutihe" <transbahasa.ilmiah@gmail.com>, Citra Panigoro <transbahasatb@gmail.com>

Dear Authors,

Kindly let us know when you will be submitting the revised paper.

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com>

Sat, Oct 5, 2019 at 8:02 PM

To: IJLTER ORG <ijlter.org@gmail.com>

Dear Editor,

I am planning to resubmit the revised manuscript on Monday, 7 October 2019. Hopefully it will not too late to submit it

Thank you.

Regards,

Lukman Laliyo

[Quoted text hidden]

IJLTER ORG <ijlter.org@gmail.com>

Sat, Oct 5, 2019 at 8:03 PM

To: Lukman Laliyo <lukman.laliyo019@gmail.com>

It is ok. Thank you for the prompt update. It is better to take your time to submit a good manuscript.

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com>

Sat, Oct 5, 2019 at 8:08 PM

To: IJLTER ORG <ijlter.org@gmail.com>

Thank you for your understanding.

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com>

Mon, Oct 7, 2019 at 2:33 PM

To: IJLTER ORG <ijlter.org@gmail.com>

Dear Editor,

Please kindly find the documents attached as requested.


Thank you!

Regards,

Lukman Laliyo

[Quoted text hidden]

2 attachments

 **1650-5889-1-RV.docx**
2851K



Correction Sheet.docx
13K

IJLTER ORG <ijlter.org@gmail.com>
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Tue, Oct 8, 2019 at 11:10 AM

Dear Dr Laliyo,

Thank you for making the requested changes and for submitting the final paper. It is now ready for publication after a final assessment has been made.

Our copy-editing team will do the final formatting before publication.

You can now make the payment of \$350 using the following link:
<https://www.paypal.me/LondonConsultingLtd/350>

Your paper will be published in the **September 2019** issue within 5 business days of payment confirmation.

Prof. Antonio Sprock
CE

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: transbahasa.go@gmail.com

Tue, Oct 8, 2019 at 1:30 PM

Halo Sir,

Berikut diteruskan e-mail dari Editor terkait artikelnnya pak Lukman.

Terima kasih.

Best,
Mimi

----- Forwarded message -----

From: **IJLTER ORG** <ijlter.org@gmail.com>
Date: Tue, Oct 8, 2019 at 11:11 AM
Subject: Re: [IJLTER] Editor Decision
To: Lukman Laliyo <lukman.laliyo019@gmail.com>
[Quoted text hidden]

IJLTER ORG <ijlter.org@gmail.com>
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Sun, Oct 13, 2019 at 1:49 PM

Dear Dr Laliyo,

Thank you for making the requested changes and for submitting the final paper. It is now ready for publication after a final assessment has been made.

Our copy-editing team will do the final formatting before publication.

You can now make the payment of \$350 using the following link:
<https://www.paypal.me/LondonConsultingLtd/350>

Your paper will be published in the **September 2019** issue within 5 business days of payment confirmation.

Prof. Antonio Sprock
CE

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: IJLTER ORG <ijlter.org@gmail.com>

Sun, Oct 13, 2019 at 2:08 PM

[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: IJLTER ORG <ijlter.org@gmail.com>

Sun, Oct 13, 2019 at 2:10 PM

Dear Editor,

I have paid the publication fee five day ago. Please find the attachment as proof of payment you received.

Thank you!

[Quoted text hidden]



Screenshot_2019-10-13-14-07-39-95.png
198K



Lukman Laliyo <lukman.laliyo019@gmail.com>

Fwd: You've sent a payment

5 messages

Transbahasa <transbahasa.go@gmail.com>
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Tue, Oct 8, 2019 at 1:58 PM

Payment for Lukman Laliyo's article to IJLTER

From: "service@intl.paypal.com" <service@intl.paypal.com>
Date: 8 October 2019 1.56.21. PM GMT+8
To: Novriyanto Napu <novrique@gmail.com>
Subject: You've sent a payment

Dear Novriyanto Napu,



You've sent 350,00 \$ USD to
London Consulting Ltd.

[Transaction Details](#)

Transaction ID: 55L72034P5586200Y

8 October 2019

Payment amount sent

350,00 \$ USD

PAID WITH

MASTER_CARD 9834

5.163.229 Rp IDR

You paid

350,00 \$ USD

Conversion rate:

1 IDR = 0 USD

Shipping Address

Jl. Jend. Sudirman No.6
Gorontalo Gorontalo 96128
Indonesia

PayPal Buyer Protection



Your eligible purchases can be safeguarded by our
Buyer Protection.



[Help & Contact](#) | [Security](#) | [Apps](#)



PayPal is committed to preventing fraudulent emails. Emails from PayPal will always contain your full name. [Learn to identify phishing](#)

Please don't reply to this email. To get in touch with us, click [Help & Contact](#).

Copyright © 1999-2019 PayPal. All rights reserved.

Consumer advisory – PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the terms and conditions carefully.

PayPal PPC000956:1.4.5:476dc8f9d4689

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: IJLTER ORG <ijlter.org@gmail.com>

Tue, Oct 8, 2019 at 2:06 PM

Dear Editor,

Please kindly find the payment confirmation forwarded regarding my submission to International Journal of Learning, Teaching and Educational Research, "THE DEVELOPMENT OF TWO-TIER INSTRUMENT BASED ON DISTRACTOR TO ASSESS CONCEPTUAL UNDERSTANDING LEVEL AND STUDENT MISCONCEPTIONS IN EXPLAINING REDOX REACTIONS".

Thank you!

Regards,
Lukman Laliyo
[Quoted text hidden]

Lukman Laliyo <lukman.laliyo019@gmail.com>
To: Transbahasa <transbahasa.go@gmail.com>

Tue, Oct 8, 2019 at 2:07 PM

Received and forwarded.

Thank you, Sir!
[Quoted text hidden]

IJLTER ORG <ijlter.org@gmail.com>
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Tue, Oct 8, 2019 at 4:04 PM

Dear Dr Laliyo,

We have received your payment. Thank you.
Your paper is being prepared for publication. You will be kept informed.
[Quoted text hidden]

--

Editorial Office

International Journal of Learning, Teaching and Educational Research

ISSN: 1694-2116 (Online)

ISSN: 1694-2493 (Print)

Website: <http://ijlter.org/>

Email: ijlter.org@gmail.com

IJLTER ORG <ijlter.org@gmail.com>
To: Lukman Laliyo <lukman.laliyo019@gmail.com>

Sun, Oct 13, 2019 at 2:30 PM

Indeed your payment has been well received.
Sincere apologies for this mistake.
Your paper will be available online within 24 hours. You will be kept informed.
[Quoted text hidden]



Lukman Laliyo <lukman.laliyo019@gmail.com>

[IJLTER] The Development of Two-Tier Instrument Based On Distractor to Assess Conceptual Understanding Level and Student Misconceptions in Explaining Redox Reactions

2 messages

IJLTER .ORG <editor@ijlter.org>

Tue, Oct 15, 2019 at 2:52 AM

To: Lukman Abdul Rauf Laliyo <lukman.laliyo019@gmail.com>, "Deasy N. Botutihe" <transbahasa.ilmiah@gmail.com>, Citra Panigoro <transbahasatb@gmail.com>

Dear Authors,

Your paper has been published. You may check it at:

<http://ijlter.org/index.php/ijlter/article/view/1650/pdf>

International Journal of Learning, Teaching and Educational Research

<http://ijlter.org/index.php/ijlter>

Lukman Laliyo <lukman.laliyo019@gmail.com>

Tue, Oct 15, 2019 at 8:48 AM

To: transbahasaid@gmail.com

Dear Admin,

Please open the site forwarded. Also, download and print out the pdf. Soon I am going to send the cover page.

Thank you!

Cheers!

Mimi

[Quoted text hidden]

The Development of Two-Tier Instrument Based On Distractor to Assess Conceptual Understanding Level and Student Misconceptions in Explaining Redox Reactions

Lukman Abdul Rauf Laliyo
Universitas Negeri Gorontalo
Gorontalo, Indonesia

Deasy Natalia Botutihe and Citra Panigoro
Universitas Negeri Gorontalo
Gorontalo, Indonesia

Abstract. Fifteen distractor two-level multiple choice items were developed as diagnostic instruments to evaluate the level of conceptual understanding and structure of students' misconceptions in explaining redox reactions. Questions at the first tier (Q1) assess the level of knowledge, and questions at the second tier (Q2) assess the level of reasoning of students. This instrument was given to 1150 participants. The participants were 11th grade students, from eight senior high schools, in the Eastern part of Indonesia. The collected data was analyzed using the Rasch model approach. The results of this study provide diagnostic and summative information on the progressiveness of student learning outcomes, as well as evidence of empirical validity and reliability of measurement. In addition, by comparing the size of items Q1 with Q2, it was found that the level of student knowledge is not always proportional to the level of reasoning, even in some cases, the level of knowledge is lower than the level of reasoning, and vice versa. The results of the investigation using the option probability curve; it was revealed that there were students' misconceptions and inconsistencies about the concepts of reduction, oxidation and oxidation numbers. This result confirms why students have difficulty interpreting and converting redox reaction equations.

Keywords: two-tier instrument; distractor; understanding; reasoning; redox reaction

Introduction

The framework of this research is based on constructivism approaches, which assume that students actively build their own conceptual understanding. In this

context, teachers play an important role in facilitating students to develop their learning creativity, to construct their knowledge and conceptual understanding as optimally as possible. Early concepts that have been understood by students will determine the learning outcomes (Rahayu, 2017). However, this effort is not easy for students. Many studies have showed that students must study hard to understand chemical concepts deeply and meaningfully (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013). Students can find it difficult to understand chemical concepts due to their complex and abstract characteristics (Johnstone, 1991; Taber, 2013; Gabel, 1999). Therefore, students develop their own conceptions, which tend to be different from the scientific community presumption. Coherent but wrong conceptual structures that have been firmly embedded in the minds of students, especially related to daily experience are misconception (Chandrasegaran, Treagust, & Mocerino, 2007; Johnstone, 2006; Taber, 2009; Taber, 2013).

One of the core concept within Chemistry curriculum in middle school in Indonesia is redox reaction. Chemical education experts in Indonesia have extensively reported that students find it difficult to understand this concept. Students find it difficult to show good quality of conceptual understanding through utilization of two tiers distractor-based instrument. This study was aimed at evaluating the conceptual level of understanding of the students in describing the redox reaction, which focuses on two important aspects, knowledge and reasoning. Evaluation of knowledge relates to measurement of students' mastery toward the content of oxidation, reduction, and the shifts in oxidation number concepts. Whereas, reasoning evaluation deals with the measurement of students' ability in providing reasons/feedbacks to back up their understanding. The good quality of students' conceptual understanding is when students were able to know and describe their knowledge properly. The main focus of this study was aimed at describing two issues, (1) how effective is the two-tier distractor-based instrument to measure level of conceptual understanding and diagnose the structure of students' misconception? And (2) how is the description of conceptual level of understanding and the structure of students' misconception in describing the redox reaction? Therefore, this quantitative research is non-experimental in combination with a qualitative descriptive study. It is not be manipulated nor regulated the process or learning material.

Theoretical Framework

Scientific education researchers have developed many types of instruments to diagnose student misconceptions. One of them that is often used today is a two-tier multiple choice diagnostic instrument (Treagust, 1988; Chandrasegaran et al., 2007; Tüysüz, 2009; Femintasari, 2015). This instrument cannot only evaluate conceptual understanding, but also can diagnose student misconceptions. Qualitatively, this diagnostic instrument is relatively effective in providing information regarding the way of students think and who have misconceptions. However this instrument tends to have limitations, such as the weak internal consistency. Because the instrument is not in the same measurement dimension, then it is difficult to use for summative measurement (Lu & Bi, 2016).

Misconceptions studies tend to have interesting progress as researchers have new concepts of this issue. Research findings showed that the fallacies are often resistant and still persist even after formal learning. This new concept of misperception has been evolving since the early 2000s, along with the progression of knowledge and reasoning, which varied according to the level of understanding of students (Aktan, 2013). Two-level diagnostic instruments can diagnose misconceptions, but cannot measure the level of conceptual understanding and evaluate the development of students' knowledge and reasoning. As a result the information produced is relatively limited for teachers for making decisions (Wilson, 2008).

Furthermore, Rasch model has been introduced as a measurement method that can integrate the diagnostic evaluation approach of students' misconception with summative evaluation of students (Liu, 2012). Although there are still studies that develop formative evaluations to investigate the progression of student learning outcomes (Claesgens, Scalise, Wilson, & Stacy, 2009; Hadenfeldt et al., 2013), several other studies have developed integrative diagnostic and summative approaches (Hoe & Subramaniam, 2016). In fact, there are studies have developed instruments based on Rasch modeling, which integrate the diagnostic evaluation approach to misconception with summative evaluation in chemistry learning (Chi, Wang, Luo, Yang, & Huang, 2018).

This study was designed using a Rasch-based instrument, in the form of a two-level distractor-based multiple choice item. Multiple choice items are often criticize as it tests only the facts and due to its inability to diagnose student's understanding (Klassen, 2006). These downsides are solved using the power of diagnostic selected items with two stratified questions (Treagust, 1988), and utilizing answer distractor choices on each item (Sadler, 1998; Herrmann-Abell & DeBoer, 2011). The aim of this research was to develop a distractor-based two-tier multiple-choice diagnostic instrument (TMDI), to evaluate the level of conceptual understanding and structure of students' misconceptions in explaining the redox reaction.

Method

There are three important components that are interrelated in the progression of measurement instruments: cognition, observations and interpretations. Cognition refers to theories or constructs about how students' understanding develops related to what they learn. Observations refer to student performance based on the type of problem solving or task and situation at the time of measurement. Interpretations refer to the results of data analysis, in the form of a statistical model as a summary of students' understanding patterns (NRC, 2001). These three components become an important part of the development stage of the instrument proposed by Wilson (2008), which includes developmental progress variables, item models, outcome space, and measurement models. The first stage is developmental progress variable, with regard to the developmental evaluation variables and progressiveness of student

learning outcomes that only focus on one characteristic to be measured. The second stage, the items model relating to the form of items that designed to obtain suitability between the measurement and various understanding diagnoses. Each item is designed to provide a response that is diagnosing student understanding, at least one level for one construct map. The third stage, the outcome space, is student learning outcomes that are categorized for all items related to certain progress variables. The fourth stage, the measurement model, in this study uses the Rasch model.

The diagnostic instruments developed in this study were adapted from the framework of developing two-tier diagnostic instruments reported by Treagust (1988) and Chandrasegaran, Treagust, & Mocerino (2007). Distractor item design were adapted from Herrmann-Abell & DeBoer (2011) and Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann (2013), so that it can be used to diagnosis what students do, what don't they know, the misconceptions and alternative ideas that they have understood. The data obtained were analyzed using the Rasch model approach, through the four steps described below.

Description of students' conceptual understanding

The first step is to describe students' conceptual understanding. This requires a construct map that provides the substantive definition and the qualitative level of the construct, according to the complexity of the construct of the variable being measured (Wilson, 2009). The progress of students' understanding on how and what they understand is related to the construct of definition map (Wilson, 2012). The map definitions of students' conceptual understanding in explaining the redox reaction was elaborated from the Chemistry Curriculum of High Schools in Indonesia. Learning redox reaction concept intended to improve students' ability to identify redox reactions, oxidation numbers and able to analyze the redox reaction based on changes in oxidation numbers, obtained from experimental data or through experiments (Ministry of Education and Culture, 2016). The progressiveness of student learning outcomes is the comprehensive and interrelated students' conceptual understanding in explaining the redox reaction. The word "explain" in the definition, implies that students must be able to interpret and develop their own reasoning, in which used to solve problems related to redox reactions. A map of the definition of level of conceptual understanding is shown in Table 1.

Table 1. The level of students' conceptual understanding in explaining the redox reaction

Representation Level	Conceptual Understanding of Redox Reaction	Item
Macro, submicro, symbolic	9. Students can determine non-redox reaction by analyzing submicroscopic diagram.	14E
	8. Students can convert submicroscopic diagrams of iron rusting reaction to symbolic chemical	13D

	reaction equation	
	7. Students can determine autoredox reaction by analyzing submicroscopic diagram that illustrated by reaction between bleaching powder and HCl and dissolve in water.	15F
Symbolic	6. Students can determine elements that undergoing oxidation state change in A and B redox reaction	4skAb, 6skBb
	5. Students can determine substances that act as oxidator and reductor in A and redox reaction.	3skAs, 7skBs
Submicroscopic	4. Students can determine submicroscopic diagram of irreversible reaction between Zn dan CuSO ₄ .	10sbBx
	3. Students can determine net ion equation by submicroscopic diagram of A dan B redox reaction.	2sbAi, 9sbBi
	2. Students can interpret submicroscopic diagram of transfer elctron in B and C redox reaction.	8sbBe, 12sbCe
Macroscopic	1. Students can identify color change of solution due to redox process:	
	A. Iron metal dipped in CuSO ₄ solution	1mA
	B. Zink metal dipped in CuSO ₄ solution	5mB
	C.. Iron metal dipped in H ₂ SO ₄ solution	11mC

Table 1 presents nine levels of conceptual understanding of students, in interpreting and converting redox reactions A, B and C, through macro, submicro, and symbolic representations. The purpose using representations are (1) Macroscopic representation to measure students' abilities in describing real and visible redox reactions, according to everyday student experience, or at least experience when observing change in redox reactions (such as discoloration of solutions, gas formation, deposits) through laboratory experiment. (2) Submicroscopic representation to measure their ability to interpret and explain redox reactions at the particulate level, using images of atoms, molecules and ions. (3) Symbolic representation to measure students' abilities using chemical symbols, formulas and equations, molecular structure, and diagrams. (4) Through macro, submicro and symbolic representations, can be measured students' ability to explain redox reactions based on the relationship of the three levels of representation. Therefore the ability of students to interpret and convert macro, submicro and symbolic representations is closely related to the construction of students' conceptual understanding. If students do not have

proper understanding of the redox reaction, then students may find it difficult to explain the redox reaction using representation.

Acquiring chemical knowledge without clear conceptual understanding can be confusing for student, because simultaneously they are required to be able to explain the concept using all three levels of representation (Chandrasegaran et al., 2007). Students must explain reaction changes at the macroscopic level, then explain changes at the particulate level, and the last explain using symbols and formulas (Gabel, 1999). For example, the redox reaction A. When iron (Fe) is dipped in CuSO_4 solution, macroscopically, the color changes reddish. The next level is students must be able to interpret the “macro” changes to the particle level (using a submicroscopic diagram), and then symbolically write down the redox reaction equation, into: $\text{Fe(s)} + \text{Cu}^{2+}(\text{aq}) \rightarrow \text{Fe}^{2+}(\text{aq}) + \text{Cu(s)}$ (symbolic). Students competency to explain the redox reaction A, by interrelating the three levels of representation, can only be achieved when students are facilitated to experience direct learning with the real world (macro), then submicro and symbolic explanations are gradually introduced. Therefore getting knowledge by students becomes more structured, systematic and constructive that stored in the long-term memory of students (Tsapalis, 2009).

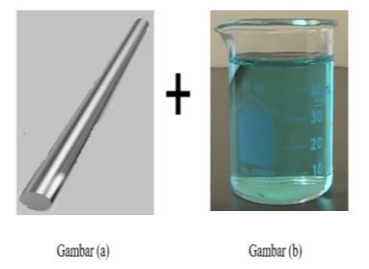
<p>Consider the following picture:</p> <div style="text-align: center; margin: 10px 0;">  </div> <div style="display: flex; justify-content: space-around; margin-top: 5px;"> Gambar (a) Gambar (b) </div>	<p>Question:</p> <p>Q1. Figure (a) show iron metal, and (b) is CuSO_4 solution with blue color. If iron metal dipped into CuSO_4 solution, what is the change that will be occur?</p> <ul style="list-style-type: none"> A. Iron metal does not dissolve in CuSO_4 solution B. Iron metal dissolved like sucrose dissolved in water C. Iron changed to reddish color. D. CuSO_4 solution does not undergo any change <p>Q2. Which one of the following is the reason of your answer:</p> <ul style="list-style-type: none"> A. Iron metal undergo oxidation B. Iron metal does not react with CuSO_4 solution C. Iron metal is easier to gain electron. D. Iron metal dissolve in CuSO_4 solution
---	---

Figure 1. Item 1mA (Q1/Q2)

Item Design

The third step is data collection using a two-tier multiple choice instrument, which is given to 1150 tenth grade students from eight senior high schools in the northern region of Sulawesi, Indonesia. Students are chosen randomly, and agree to be voluntarily respondents. All students have learned redox reaction concept according to the Chemistry Curriculum Standards in Indonesia. Process standards and learning outcomes have been elaborated completely in this curriculum. Chemistry teachers are required to develop and use problem solving-based instructional strategies. This strategy is used to encourage students to design and make redox reaction experiment, practice to solve problems, explain the process of electron transfer, in the reaction between iron metal dipped in a solution of CuSO_4 and H_2SO_4 , discuss redox reaction equation, and determine elements that are undergoing a changes in oxidation number. Based on this curriculum, it can be understood that by learning process students may form their conceptual understand of redox reactions in a gradual, constructive, systemic and interrelated manner. Likewise, it can be explained when students have many misconceptions, indicating that students do not experience a whole and meaningful learning process. If these assumptions can be proven empirically, the instruments developed in this study have good predictive validity.

Data have been collected for three months, in each school. Students respond manually through a written answer sheet. The test time is 45 minutes and was supervised by teacher. All students are told to answer all questions in the instrument. After testing, all instrument texts are collected directly by the researcher. There are no instrument texts that are not collected.

Measurement Model

The fourth step is use the Rasch analysis model approach to calibrate the difficulty level of the item and students' ability level on the same interval scale. Raw data is inputted in Excel format in the form of dichotomous data, then converted into interval data, using WINSTEPS version 3.75 model software (Linacre, 2012; Bond and Fox, 2015). Raw data input considers the item model that containing two questions Q1 / Q2. Student responses on each item in the first tier (Q1) and in the second tier (Q2) were analyzed separately.

The Rasch model, combine an algorithm 'i' (as the results of probabilistic expectations of items) and students 'n', which are mathematically stated by Bond and Fox (2015), as: $P_{ni} (X_{ni}=1/(\beta_n, \delta_i)) = (e^{((\beta_n - \delta_i))}) / (1 + (\beta_n - \delta_i))$, where: $P_{ni} (X_{ni}=1/(\beta_n, \delta_i))$ is the probability of students "n" in items "i" to get a correct answer ($x = 1$); with student ability β_n , and difficulty level of items δ_i . The above equation can be further simplified by entering the logarithmic function and making it: so that the probability of success is: the probability of success equals to the ability of the student minus the difficulty level of the item. It is important to underline that students' ability and items are stated at the same and independent intervals. The level of students' abilities and difficulty level of items is measured in units of logarithms called odds or logs, which can vary from -00 to +00 (Herrmann-Abell & De Boer, 2011; Sumintono & Widhiarso, 2015).

Rasch models can evaluate the level of understanding, reasoning of students and the difficulty level of items. To test efficiency of the instrument indirectly is by comparing the distribution of difficulty items with the level of understanding and reasoning of students. This is different from the classical test theory approach, which cannot diagnose the level of understanding of students, because the difficulty level of items is only based on raw data (Lu & Bi, 2016). Student with high ability will be able to answer items with a lower level of difficulty.

Result and Discussion

Reliability

Reliability explains how far measurements produce information that is consistent in revealing latent traits. Can the TMDI instrument measure the unidimensional nature of latency, namely students' conceptual understanding in explaining the redox reaction (Sumintono & Widhiarso, 2015). The measurement reliability of the person item is based on the person separation index and item separation index. The Separation index can also be converted to Cronbach's equivalent value from 0-1. Table 2 shows a summary of fit statistics, which are compatible with the Rasch dichotomous model.

Tabel 2. Summary of fit statistics

	Person (N=1150)	Item (N=30)
Reliabilitas	0,58	0,99
Infit MNSQ	1.00	1,00
Outfit MNSQ	1.03	1.03
Separation index	1.18	10,14
Cronbach Alpha	0,61	-

From Table 2, it is known that the person separation index is 1.18 which is equivalent to the reliability value of person 0.58. This value indicates that the reliability of the person is relatively weak. This means that the consistency of students' responses to TMDI test items is relatively weak. Cronbach Alpha Coefficient value of 0.61 indicates that the interaction between 1150 students with 30 items is fairly adequate. However, this fact will not affect the teacher's decision to develop students' abilities to be better (Wei, Liu, Wang, & Wang, 2012; Lu & Bi, 2016). The item separation index value (10.14) is equivalent to the reliability value (0.99). This means that the consistency of the item is classified as very good (special), or the item meets the requirements of unidimensionality. This indicates that the item is able to define latent variables very well. The infit and outfit values of most items are within the acceptable range for multiple choice tests (Abell & DeBoer, 2011; Bond and Fox, 2015).

Validity

Validity describe how well measurements are carried out according to what should be measured. The validity of the TMDI test was tested by the suitability analysis of the Rasch model item. If the test is able to measure four levels of a student's ability to explain the redox reaction, then this test has good construct validity (Linacre, 2012). The results of the analysis are presented in Table 3.

Table 3. Item Fit Statistics^a

Item	Measure	Model SE	INFIT		OUTFIT		PTMEA Corr
			MNSQ	ZSTD	MNSQ	ZSTD	
1mA(Q1)	-0,73	0,06	0,93	-3,9	0,92	-4,0	0,41
1mA(Q2)	-1,00	0,06	0,94	-3,0	0,91	-3,5	0,40
2sbA(Q1)	-1,02	0,06	0,84	-7,7	0,81	-7,8	0,53
2sbA(Q2)	-0,56	0,06	0,95	-2,9	0,93	-3,5	0,39
3skAs(Q1)	0,18	0,06	1,01	0,6	1,05	1,7	0,25
3skAs(Q2)	0,93	0,07	1,07	1,7	1,12	2,1	0,13
4skAb(Q1)	-1,43	0,07	0,89	-4,1	0,80	-5,7	0,47
4skAb(Q2)	-0,15	0,06	0,90	-5,4	0,89	-4,8	0,45
5mB(Q1)	0,80	0,07	1,04	1,0	1,07	1,4	0,19
5mB(Q2)	-0,71	0,06	0,96	-2,4	0,95	-2,6	0,37
6skBb(Q1)	-0,51	0,06	0,88	-6,8	0,86	-7,3	0,48
6skBb(Q2)	-0,44	0,06	0,88	-6,8	0,87	-6,6	0,48
7skBs(Q1)	0,46	0,07	1,15	5,0	1,23	5,5	0,03
7skBs(Q2)	0,99	0,08	1,16	3,6	1,36	5,6	-0,05
8sbB(Q1)	0,17	0,06	1,04	1,8	1,07	2,1	0,22
8sbB(Q2)	0,52	0,07	1,04	1,5	1,05	1,2	0,20
9sbBi(Q1)	-0,40	0,06	0,88	-7,2	0,88	-5,9	0,48
9sbBi(Q2)	1,27	0,08	1,13	2,5	1,41	5,2	-0,05
10sbBx(Q1)	0,25	0,07	0,85	-6,4	0,83	-5,3	0,50
10sbBx(Q2)	-0,35	0,06	0,96	-2,1	0,95	-2,5	0,36
11mC(Q1)	1,21	0,08	1,16	3,0	1,33	4,5	-0,05
11mC(Q2)	-0,88	0,06	0,89	-6,0	0,86	-6,5	0,48
12sbC(Q1)	0,97	0,08	1,06	1,4	1,11	1,9	0,14
12sbC(Q2)	-0,34	0,06	1,22	9,9	1,22	9,9	-0,01
13D(Q1)	0,31	0,07	1,05	2,0	1,08	2,3	0,19
13D(Q2)	-0,34	0,06	1,12	6,2	1,13	6,0	0,13
14E(Q1)	0,10	0,06	0,94	-2,5	0,93	-2,5	0,37
14E(Q2)	0,14	0,06	0,94	-2,6	0,95	-1,6	0,36
15F(Q1)	0,22	0,07	0,96	-1,8	0,97	-0,8	0,34
15F(Q2)	0,37	0,07	1,18	6,3	1,20	5,2	0,01

^aN = 1150. MNSQ values of 0.70–1.30 indicate the acceptable fit range; infit (weighted) or outfit (unweighted) values outside this range indicate poor fit of the data to the Rasch model.

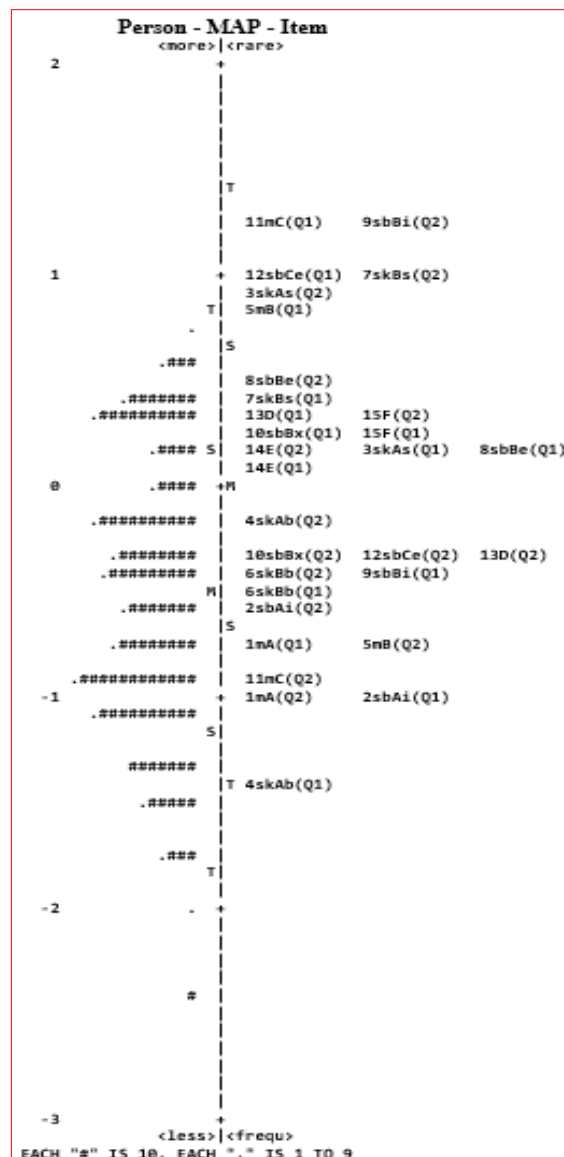


Figure. 2 Item-person map for the 30 items (Q1 and Q2)

Quality of item fit is to examine whether the test item is functioning normally or not. Items that are not fit indicate student misconception on that item. The indicator used is the value of Mean Square Residual (MNSQ). Outfit MNSQ and infit MNSQ are sensitive chi-square to detect outlier responses. This value is to show how much impact the item incompatibility has. Outlier responses are often correct answers to items that are difficult by students with low ability due to guessing, or wrong answers to items that are easier by students who are more capable, due to carelessness.

The ideal value expected from MNSQ is 1.0. In addition, another indicator is point measure correlation (PTMEA Corr). PTMEA Corr is a score and person measure score correlation, whose value must be positive and not close to zero (Bond and Fox, 2015). The criteria used is the means-square outfit (MNSQ) value: $0.5 < y < 1.5$; z-standard outfit: $-2.0 < Z < +2.0$, and PTMEA Corr: $0.4 < x < 0.8$. If the three criteria are not fulfill, the items are not good enough therefore need

further investigation (Boone, Yale, & Staver, 2014). Based on Table 3, it is known that no item does not meet the three criteria. However, there are four items, namely 12sbC (Q2), 11mC (Q1), 9skBi (Q2), 7skBs (Q2)) which are negative, and item 15F (Q2) which is close to zero. The five items require further investigation

After testing the suitability of the item, we examine level of difficulty of the item with the level of ability of students, using "Wright Map" (person-item map), is presented in Figure 2. This map shows the distribution of students' abilities (on the left side) and the distribution of difficulty items (on the right side) vertical lines. Distribution of low-ability students and low difficulty items are at the bottom of the map. Distribution of high-ability students and items with high difficulty levels are at the top of the map. The item's horizontal line mean (0.0 logit) is higher than the person mean (-0.52 logit). All items can reach the entire scale of the student's logit ability. However, there are six items that are too difficult for students, above the 1 SD scale (standard deviation). They are 5mB/Q1 (0.80), 11mC/Q1 (1.21), 12sbC/Q1 (0.97), 3skAs/Q2 (0.93), 7skBs/Q2 (0.99), and 9skBi/Q2 (1.27). These items need to be adjusted and examined further.

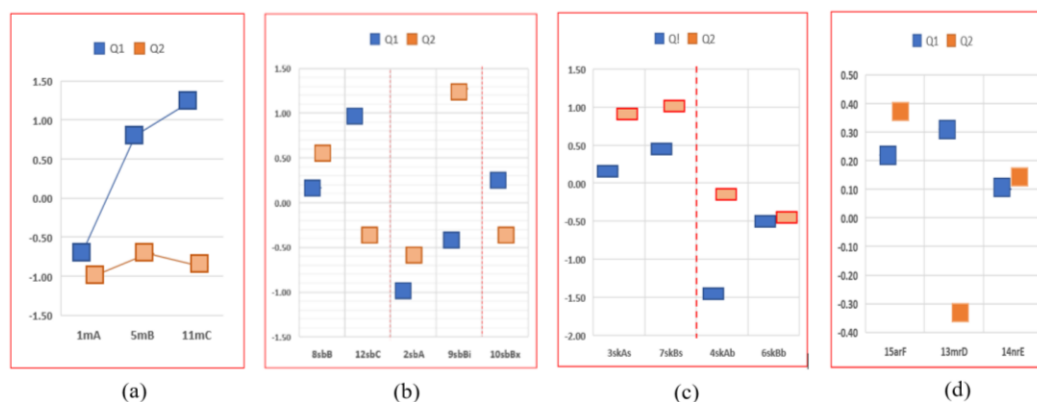


Figure 3. Comparison of the size of items in the first tier (Q1) and in the second tier (Q2)

Level of conceptual understanding

The level of conceptual understanding of students is distinguished by the level of knowledge (Q1) and the level of reasoning (Q2) in each construct studied (item). This difference is explained based on the comparison of item size (measures items). Students' misconceptions are examined by testing response patterns and option probability curve patterns. The response pattern test is estimated from the results of the distractor test, which is based on the acquisition of the average ability of the item. If the average ability value rises, the effect of the distractor works. This means that students give the answers are not by guessing. However if the average ability value drops, the effect of the distractor does not work meaning that the item is easy to guess. Student's answers are not representing the knowledge they have learned, because students guess the answer. The decrease in average ability value indicates that there are low-ability students who are able to answer items correctly and / or there are high-ability students who answer incorrectly on items that should be answered correctly, and this is not favorable for measurement (Smiley, 2015).

The option probability curve is used to explore the pattern of student responses. Curves that correspond to the probability of the correct answer will usually increase, and curve due to distractor will usually decrease with the increase in students' abilities (Haladyna, 2004). In general, for items that are driven by a distractor, the curve obtained does not correspond to the probability curve of the correct answer (Sadler, 1998). Therefore, each student's answer are analyzed separately, to represent data more accurately and provide additional information about the aspects of misconception students have.

The probability choice answer curve provides information on the possibility of students choosing on each answer option provided, as a function of the level of knowledge and the level of reasoning in the construct being measured. In this case, there are four curves, where one curve represents one answer. It means that students who choose one answer choice, have certain conceptual understanding of the construct being measured. Low-ability students related to measured constructs will be interested in certain misconceptions, and high-ability students will be attracted to other misconceptions (Herrmann-Abell & DeBoer, 2011). In addition, the probability curve is presented through visual images, to show the distribution of the correct answers and those experiencing misconceptions in the whole of the spectrum of students' conceptual understanding. Thus it can be identified which form of the curve is appropriate, or is not appropriate, as an indication of the misconception of an item. Forms of curves that are not appropriate, can also show a certain structure of misconceptions, with patterns that are repetitive, consistent and will disappear sequentially, along with increasing student abilities. Diagnosis of misconception structures in an accurate way, allows teachers to develop more effective learning strategies. In this article, an example of how Rasch modeling and the probability option curve can be used to reveal the structure of students' misconceptions in explaining the types of redox reactions.

Level of macroscopic reasoning

The level of macroscopic reasoning is the students' ability to describe the color change of the solution due to A, B, and C the redox reaction processes, that was evaluated with items 1mA, 5mB, and 11mC. In the first tier (Q1) students are asked to determine what changes they know, if reactions A, B and C occur. In the second tier (Q2) students are asked to state the reason for the answer in the first tier. The difference in the level of knowledge and reasoning of students on each item is determined by the size of the item. The larger the item size, the higher the difficulty level of the item, the lower the level of knowledge and reasoning of students.

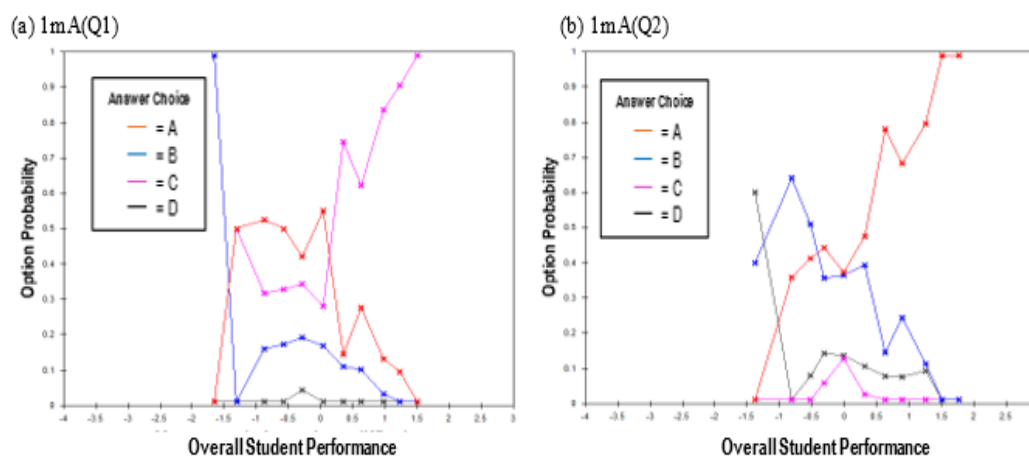


Figure 4: Option Probability Curve (a) item 1mA(Q1); (b) item 1mA(Q2)

Figure 3a shows that the size of items in the first tier (Q1) is in the range of -0.55 logit to 1.25 logit, while in the second tier (Q2), no more than -0.50 logit. It means, empirically the level of student knowledge is lower than the level of reasoning in identifying the changes colors of the solution of A, B and C redox reaction at the macroscopic representation level. To explain this fact, option probability curve is used for item 1mA (Figure 4).

For curve 4 (a), low-ability students (between -1.5 to 0.5 logit) tend to choose the distractor answer option, namely A (iron will not dissolve in CuSO_4), and B (Iron dissolved in CuSO_4 , such as as sugar dissolves in water). It is seen that the peak of the A curve is more dominant than B. Students with moderate ability levels (more than 0.5 logit) tend to choose the right answer C (the solution turns reddish). This pattern of misconception shows evidence of the lack of acquisition of student knowledge about the redox reaction, and develops logical misconceptions, iron is insoluble in CuSO_4 , due to the hard properties of iron. Through the 4c curve, it is known that students with ability levels (-1.5 to 1.5 logit) are more dominant choosing the answer distractor B option (Iron does not react with CuSO_4 solution). Students with an ability level of more than 0.5 logit, tend to choose the correct answer option A (iron undergoing an oxidation reaction). It appears that there is the same student response pattern, stating that iron is insoluble in CuSO_4 . This pattern of misconception can be understood, because students with limited levels of knowledge, incomplete learning experiences, tend to choose to what is physically visible. This is an alternative framework of students (Johnstone, 1991; Lu & Bi, 2016; Taber, 2013), understanding that contains misconceptions. This result support the studies reported by Chandrasegaran et al (2007) and Herrmann-Abell & DeBoer (2011). This result also describe that macroscopic factual knowledge and students' rational reasoning are out of sync. It is possible that students never make direct observations through experiments to identifying the color changes in the solution of the redox reaction process.

Submicroscopic reasoning level

The level of submicroscopic reasoning is related to student ability to explain the redox reaction at the particle level. Based on item size (Figure 3b), interesting

facts were found, first, in the first tier question (Q1), the item size 12sbCe (0.97) is greater than 8sbBe (0.17), while in the second tier (Q2), item size 12sbCe (-0.24) is smaller than 8sbBe (0.52). This means that the level of student knowledge is lower than the level of reasoning, in explaining sub-microscopically the electron transfer process in the redox B and C reactions. Second, in the first tier (Q1) item size is 2sbAi (-1.02) and 9sbBi (-0.40) very low, so in the second tier (Q2) items 2sbAi (-0.56) are low. That is, the level of students' knowledge and reasoning in expressing the net ion reaction equation through the interpretation of the submicroscopic diagram of the redox A and B reactions, is very low. Unless the item size is 9sbBi (1.27), which is very high, it's hard to explain because this item's distractor option doesn't work. Third, the size of items is 10sbBx in the first tier (0.25) while in the second tier (0.35). This means that the level of student knowledge is lower than the level of reasoning, in determining the submicroscopic diagram of the reaction of Zinc metal dipped in CuSO₄ solution, if the reaction is complete (irreversible). These three facts prove that submicroscopically, students do not have good knowledge and reasoning. There is a gap between the level of knowledge and the level of reasoning. Conceptual understanding of students tends to be inconsistent and partial.

To explain how students develop their understanding at the submicroscopic level, it can be exemplified through the analysis of item response patterns 2sbAi. This item tests the ability of students to interpret the submicroscopic diagram, the redox A reaction, which is iron metal dipped in a solution of CuSO₄ (Figure 5). Based on Figure 5, in the first tier question (Q1), students are asked to write net ion reaction equations, as in the second tier (Q2), students are asked to give the reasons of answers in the first tier. The option probability curve form for item 2sbAi is shown in Figure 6.

Curve 6a, shows that students with low ability levels (smaller 0.0 logit than the ability of students as a whole) tend to choose the answer distractor C ($\text{Fe}^{2+}(\text{s}) + \text{Cu}(\text{aq}) \rightarrow 2\text{Fe}(\text{aq}) + \text{Cu}(\text{s})$) and B ($\text{Fe}(\text{s}) + \text{Cu}(\text{aq}) \rightarrow \text{Fe}^{2+}(\text{aq}) + \text{Cu}(\text{s})$). C and B are answers that contain misconceptions. Students with a level of ability greater than 0.0 logit, choose the correct answer A ($\text{Fe}(\text{s}) + \text{Cu}^{2+}(\text{aq}) \rightarrow \text{Fe}^{2+}(\text{aq}) + \text{Cu}(\text{s})$). On the other hand, on the 6b curve, students with low ability levels (less than -0.5 logit) tend to choose C option (reaction of iron and CuSO₄ is a redox reaction). Students with abilities ranging from -1.5 to 1.0 logit, tend to choose A (iron accepts electrons from CuSO₄). C and A contains misconceptions. Students with abilities above -0.5 tend to choose the correct answer B (Cu^{2+} ions are reduced to Cu).

Student response patterns on curves 6a and 6b indicate that students have low ability at submicroscopic reasoning. Although the diagram of electron transfer has been given, it cannot help students, because they are more likely to take to what they understand about oxidation and reduction. This is also found in item 9sbi, students are more difficult to express the net ion reaction equation of B reaction than A. In fact, these two items qualitatively have the same degree.

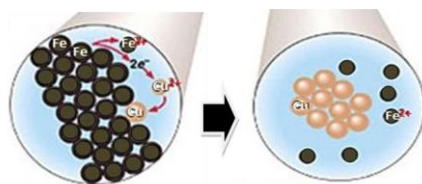


Figure 5. Submicroscopic Diagram of Transfer electron process between iron metal and CuSO_4 Solution

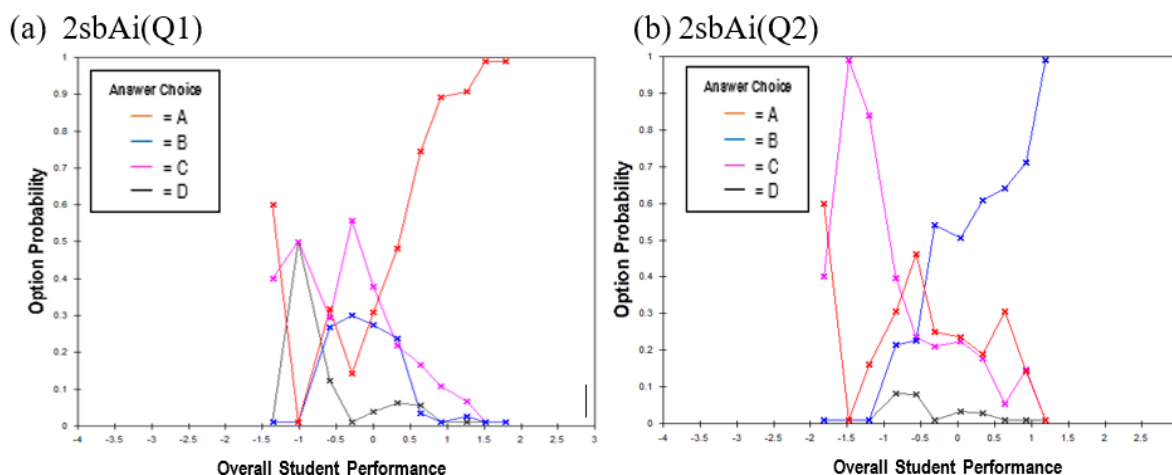


Figure 6. Option porbability Curve Kurva: (a). item 2sbAi(Q1), and (b) item 2sbAi(Q2)

Level of symbolic reasoning

The level of symbolic reasoning is evaluated based on students' ability to determine: (a) species that act as oxidizing and reducing agents (items 3skAs and 7skBs); and (b) elements undergoing oxidation number changes (items 4skAb and 6skBb). Based on item size comparisons (Figure 3c), interesting findings can be explained, as follows: first, item size Q1 (3skAs: 0.18 and 7skBs: 0.46) is greater than item Q1 (4skAb: -1.43 and 6skBb: -0.44). That means, students are more difficult to determine which species act as oxidizeing and reducing agents rather than determining the elements that undergo changes in oxidation numbers. Second, in terms of item size Q2, the highest difficulties for students are 7skBs (0.99), 3skAs (0.93), 4skAb (-0.15), 6skB (-0.44) respectively. This means students are more difficult to give a reason in determining the oxidizing / reducing species rather than determining the element that undergoes oxidation number changes. This fact is interesting, because the understanding of the concept of oxidation number and determination of species which act as oxidizers and reducing agents, are interrelated. To elaborate this problem, student response patterns was analyzed, and was exemplified through items 6BB and 7KBs (Figure 7).

(a) 6skBb	(b) 7skBs
Q1. Identify oxidation number change of Zn, in reaction between Zinc metal dipped in CuSO ₄ solution	Q1. Identify reducing agent in reaction between Zinc metal dipped in CuSO ₄ solution
A. 0 to -2. B. +1 to +2. C. +2 to 0. D. 0 to +2	A. Zinc powder B. CuSO ₄ solution
Q2. which one of the following is the reason of your answer!	C. Copper D. ZnSO ₄ solution
A. $\overset{0}{\text{Zn}}_{(s)} + \overset{+2}{\text{Cu}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_4(aq) \rightarrow \overset{-2}{\text{Zn}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_4(aq) + \overset{0}{\text{Cu}}_{(s)}$ B. $\overset{0}{\text{Zn}}_{(s)} + \overset{+2}{\text{Cu}}\overset{+6}{\text{S}}\overset{-2}{\text{O}}_4(aq) \rightarrow \overset{+2}{\text{Zn}}\overset{+6}{\text{S}}\overset{-2}{\text{O}}_4(aq) + \overset{0}{\text{Cu}}_{(s)}$ C. $\overset{0}{\text{Zn}}_{(s)} + \overset{+6}{\text{Cu}}\overset{+2}{\text{S}}\overset{-4}{\text{O}}_4(aq) \rightarrow \overset{-2}{\text{Zn}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_4(aq) + \overset{0}{\text{Cu}}_{(s)}$ D. $\overset{0}{\text{Zn}}_{(s)} + \overset{+2}{\text{Cu}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_4(aq) \rightarrow \overset{0}{\text{Zn}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_4(aq) + \overset{0}{\text{Cu}}_{(s)}$	Q2. Which one of the following is the reason of your answer! A. Zinc atom undergo oxidation and becomes Zn ²⁺ ion B. Cu ²⁺ Ion undergo reduction and becomes copper atom C. Reaction between Zn and Copper (II) sulphate produce copper deposit D. Copper (II) sulphate is product of Zinc that undergo reduction

Figure 7. (a) Item 6skBb(Q1/Q2), (b) Item 7skBs(Q1/Q2)

Questions in the first tier (Q1) item 6skBb, basically testing the ability of students to write equilibrium reaction equations of Zinc metal dipped in CuSO₄, and determine how much Zn oxidation number is in the reaction. In the second tier (Q2), students are asked to determine the oxidation number of Zn in the reaction equation. The option probability curve of this item is shown in Figure 8. Curve 8 (a), shows that students with the lowest ability (smaller than -1.0 logit) tend to choose the misconceptions answer A, B and C. The highest probability of choice is A, then decreases with increasing probability of high-ability students (more than 0.5 logit), who choose the right answer D. It means, students' misconceptions, especially in understanding the changes in oxidation number of Zn (from 0 to -2), Zn experiences reduction.

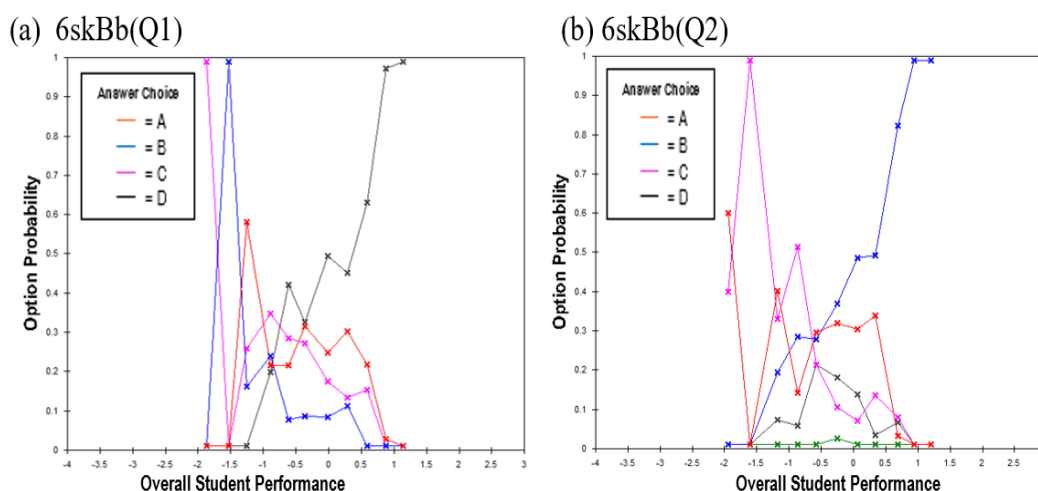


Figure 8. Option probability curve (a). item 6skBb(Q1), and (b). item 6skBb(Q2)

Curve 8(b), confirms this misconception, where low-ability students (small from -0.5 logit), tend to choose C and A, while the probability of students choosing the right answer option B, above ability from -0.5 logit. This fact shows the students' weak understanding regarding the concept of oxidation number. To test this fact, the evaluation continued with the ability of students to determine

which species acted as reducing agents in the reaction of Zinc and CuSO_4 , through item 7skBs, presented in Figure 7 (b).

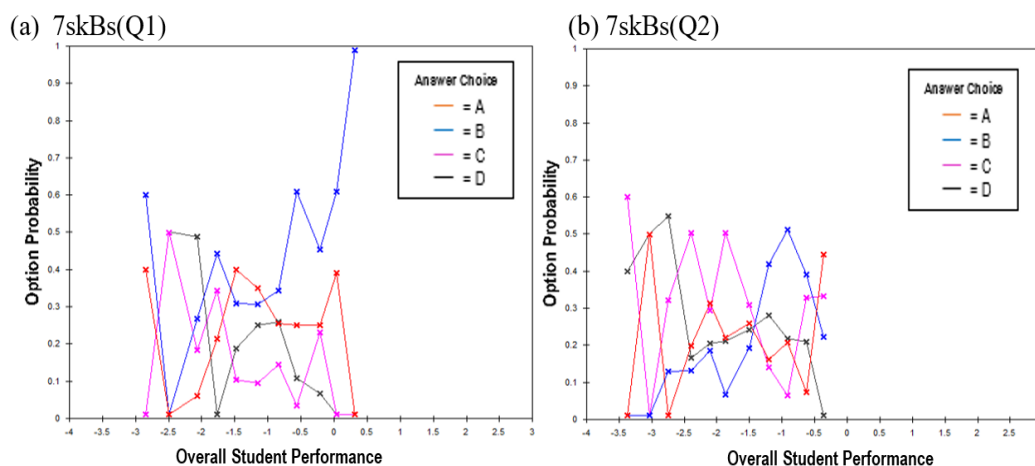


Figure 9. Option probability curve (a). item 7skBs(Q1), and (b). item 7skBs(Q2)

Based on the probability curve item 7skBs (Q1) in Figure 9 (a), it is known that the choice of answer B, which contains misconceptions, is the answer choice with the highest probability chosen by the lowest ability students (-3.0 logit) to 0.5 logit. This curve rises to reach a peak of 1.0 logit. Students argue that the reducing agent in the reaction between Zn and CuSO_4 is a solution of CuSO_4 . The probability of students choosing the right answer A (reducing agent is Zn powder) was chosen by low-ability students, -3.0 logit to 0.3 logit. However, the curve actually shows unusual shape, have two peaks then decreasing. The first peak, at the level of students' ability is smaller than -1.0 logit, and the second peak, at the level of students' ability is greater than 0.0 logit. On the other hand, the 7skBs (Q2) curve in Figure 9 (b) shows that there is no answer of students that reach option probability values up to 1.0 logit. The average of student ability is only able to reach a value of 0.6 logit option probability, from a scale of -3.5 logit to -0.4 logit. These interesting facts can be understood, because they support the previous facts, namely students understanding about the concepts of reduction, oxidation, and oxidation numbers is weak. Students are not able to determine which species act as reducing agents in the reaction of Zinc and CuSO_4 .

Macro, submicro and symbolic levels of reasoning

Macro, submicro and symbolic reasoning levels are evaluated based on students' ability to interpret and convert macro, submicro and symbolic representations, to explain three types of reactions, namely D, E and F. The letter "D" is a symbol of an item that presents a reaction to iron metal; "E" is a symbol of non-redox reaction (i.e., silver nitrate and potassium chloride solution); and F is the auto-redox reaction symbol. These three types of reactions are evaluated, to confirm students' ability to relate their knowledge and reasoning to the three levels of representation.

Figure 3d show the comparisons of sizes 13D, 14E, and 15F. In the first tier (Q1), the difficulty level of the item is $13D > 15F > 14E$. That means, students are more difficult to interpret rusting iron reactions, rather than auto-redox reactions and

non-redox reactions. In the second tier (Q2), the difficulty level of the item is $15F > 14E > 13D$. This means that students are more difficult to explain the auto redox reaction process, rather than a non-redox reaction and a rusting iron reaction. This fact is really interesting, especially item 13D, because between Q1 and Q2 there is a very wide gap. This item asks students to analyze a diagram that illustrates the redox reaction of rusting iron, and then converts to equivalent chemical reaction equation. The result is that students do not have difficulty giving logical reasons related to the iron rusting process with the reaction equation (Q2), but they are very difficult to convert submicro diagrams into reaction equations (Q1). The results of the response pattern analysis found that 34% of students stated the reaction of iron rusting is $4Fe + 3O_3 \rightarrow 2Fe_2O_3$ and 48% of students stated that the metal was reduced and oxygen was oxidation. In fact, the correct reaction is: $4Fe + 3O_2 \rightarrow 2Fe_2O_3$, because iron metal undergoes oxidation, oxygen undergoes reduction. This evidence support previous findings that students' understanding of oxidation, reduction and oxidation numbers concepts is very weak. Then in item 14E, students were asked to analyze the submicroscopic diagram of the reaction process of a solution of silver nitrate and potassium chloride solution which produced white deposits: $AgNO_3 + KCl \rightarrow AgCl + KNO_3$, and asked whether the reaction is redox reactions or not. The results showed that the size of items 14E / Q1 and 14E / Q2 is relatively not much different. That means, students are able to understand and have logical reasons for determining non redox reaction. However, there are still 25% of students stating that the reaction of $AgNO_3 + KCl \rightarrow AgCl + KNO_3$ is a reduction reaction, and 33% of students have reason that there is a change in oxidation number and electrons transfer in the reaction. This fact makes it clear that students are not able to distinguish redox and not redox, due to confusion in determining which species act as oxidizers and reducing agents.

In item 15F, students were asked to analyze submicroscopic diagrams which illustrated the reaction of bleach powder mixed with HCl ($CaOCl_2 + HCl \rightarrow CaCl_2 + Cl_2 + H_2O$), then mixed with water, and produced hydrochloric acid and hypochloric acid ($Cl_2 + H_2O \rightarrow HCl + HClO$) The results of the analysis show that the size of the item 15F/Q2 is more difficult than 15F/Q1. That means, students are very difficult to give opinions regarding the auto-redox reaction diagram, but can determine the auto redox reaction equation. 44% of students are sure that the reaction: $Cl_2 + H_2O \rightarrow HCl + HClO$ is a redox reaction, and 30% of students have reason that the reaction occurs because there is electrons transfer. The correct answer is that the reaction ($Cl_2 + H_2O \rightarrow HCl + HClO$) is an autoredox reaction, because of the oxidation and reduction reactions of the same substance. This result is evidence that student misconceptions tend to be consistent, related to the understanding and reasoning of the concepts of oxidators and reducing agents, changes in oxidation numbers.

Conclusion

The diagnostic instrument developed has construct validity, which can be used to evaluate the level of students 'conceptual understanding, including knowledge and reasoning, and diagnosing students' misconceptions in explaining the redox reaction. At the level of macroscopic reasoning, a repetitive student response was found, which states that iron is insoluble in $CuSO_4$. For

low-ability students, this pattern of misconception can be understood, because the limited level of knowledge and reasoning makes students tend to stick to what is physically visible. This misconception may develop from learning experiences that do not go through the process of direct observation or through experimentation. In the level of submicroscopic and symbolic reasoning, there is a gap between the level of knowledge and the level of reasoning of students. This gap is detected again at the level of macro, submicro and symbolic reasoning. Because the acquisition of student knowledge is not complete, it is difficult for students to explain the redox reaction macroscopically, submicroscopically and symbolically. Students have difficulty translating representational diagrams, and converting them into reaction equations. The difficulties lead students to develop misconceptions that are based on their intuitive reasoning, by choosing a choice of distractor answers that are misconceptions. The choice of this answer is detected repeatedly, and tends to come from the same understanding structure, namely the weak of students understanding of reduction, oxidation and oxidation number concept. In turn, all available evidence shows that the progress of the learning outcomes of the students' in redox reactions concepts, overall tends to be low.

Combining the item development process qualitatively with Rasch analysis models, is relatively effective to depth evaluation of the progress of student learning outcomes, reveals students' ways of thinking that are misconceptions, and know what students have not understood. Differences in item size and option probability curves illustrate that the prevalence of the most dominant misconceptions is low ability students. But for certain cases, the prevalence even increases, because the misconception factors tend to be resistant, as seen in item 9sbBe (Q2). The availability of this type of information makes it possible for teachers to classify student understanding based on the level of misconception, so that it is useful in developing instructional strategies. In addition, with the option probability curve on each item, it can be identified an unusual curve, which indicates a problem with the item.

The multiple choice distractor items in this study were developed by adapting some of the results of previous interviews and studies, which were aligned with the targets outcome of chemical learning in Indonesia. The hope is that further research can be carried out, for the development of other material. Although this study does not answer the question of why students experience misconceptions, it is possible for similar items to be developed to help teachers diagnose students' ways of thinking and understanding patterns, so that chemical learning becomes more effective and meaningful.

References

- 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. doi:10.1088/0143-0807/34/1/33
- Boone, W. J., Yale, M. S., & Staver, J. R. (2014). *Rasch Analysis in the Human Sciences*. Rasch Analysis in the Human Sciences. Springer Dordrecht Heidelberg London New York Library: Springer Dordrecht Heidelberg London New York Library. doi:10.1007/978-94-007-6857-4

- Briggs, D. (2009). The psychometric modeling of ordered multiple-choice item responses for diagnostic assessment with a learning progression. *Learning Progressions in Science (LeaPS) Conference, June 2009, Iowa City, IA, (June)*.
- 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.
- Chi, S., Wang, Z., Luo, M., Yang, Y., & Huang, M. (2018). Student progression on chemical symbol representation abilities at different grade levels (Grades 10–12) across gender. *Chemistry Education Research and Practice*, 19(4), 1055–1064. doi:10.1039/c8rp00010g
- Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The Perspectives of Chemists. *Science Education*, 93(1), 56–85. doi:10.1002/sce.20292
- 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. (1999). Improving Teaching and Learning through Chemistry Education Research: A Look to the Future. *Journal of Chemical Education*, 76(4), 548. doi:10.1021/ed076p548
- 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. doi:10.1021/ed3006192
- Haladyna, T. M. (2004). Developing and Validating Multiple-choice Test Items (3rd Editio, p. 320). New Jersey: Lawrence Erlbaum Associates. doi:10.4324/9780203825945
- 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. doi:10.1039/c1rp90023d
- Hoe, K. Y., & Subramaniam, R. (2016). On the prevalence of alternative conceptions on acid-base among secondary students: Insights from cognitive and confidence measures. *Chemistry Education Research and Practice*, 17(2), 263–282. doi:10.1039/c5rp00146c
- Johnstone, A. H. (1991). Why is science difficult to chemistry 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. doi:10.1039/b5rp90021b
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2016). Silabus Mata Pelajaran Kimia. In *Silabus Mata Pelajaran Sekolah Menengah Atas/Madrasah Aliyah (SMA/MA)*.
- Klassen, S. (2006). Contextual assessment in science education: Background, issues, and policy. *Science Education*, 90(5), 820–851. doi:10.1002/sce.20150
- Linacre, J. M. (2012). *A User's Guide to WINSTEPS® MINISTEP Rasch-Model Computer Programs Program Manual 3.75.0*. Winsteps.com. doi: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. doi:10.1039/c6rp00137h
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (Eds.). (1999). *Assessing Science Understanding: A Human Constructivist View*. Academic Press, San Diego, California : Elsevier 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
- National Research Council. (2001). *Knowing what students know: The Science and Design of Educational Assessment*. National Academy Press (Vol. 19). doi:10.17226/10019
- 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. doi:10.22452/iojel.vol2no1.3
- Rahayu, S. (2017). Promoting the 21st century scientific literacy skills through innovative chemistry instruction (Vol. 020025, p. 020025). Published by AIP Publishing. doi:10.1063/1.5016018
- 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.
- Smiley, J. (2015). Classical test theory or Rasch- A personal account from a novice user. *Shiken*, 19(1), 16–29.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Penerbit Trim Komunika.
- Taber, K. S. (2009). Challenging Misconceptions in the Chemistry Classroom: Resources to Support Teachers. *Educació Química EduQ*, 4, 13–20. doi:10.2346/20.2003.02.27
- Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, 14(2), 156–168. doi:10.1039/c3rp00012e
- 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. doi:10.1080/0950069880100204
- Trevor G. Bond and Christine M. Fox. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Routledge Taylor & Francis Group (Third Edit, Vol. 44). New York and London: Routledge Taylor & Francis Group. doi:10.1088/1751-8113/44/8/085201
- Tsaparlis, G. (2009). Linking the Macro with the Submicro Levels of Chemistry: Demonstrations and Eksperiments that can Contribute to Active/Meaningful/Conceptual Learning. In S. A. Devetak, Iztok and Glazar (Ed.), *Learning with Understanding in the Chemistry Classroom* (pp. 41–61). Springer Dordrecht Heidelberg London New York Library. doi:10.1007/978-94-007-4366-3
- 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. doi:10.1021/ed100852t
- Wilson, M. (2008). Cognitive Diagnosis Using Item Response Models. *Zeitschrift Für Psychologie / Journal of Psychology*, 216(2), 74–88. doi: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. doi: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*, 317–344. Rotterdam: Sense Publishers. doi:10.1007/978-94-6091-824-7

Correction Sheet

1. As suggested by reviewer C, the introduction section has been summarized and strengthened by focusing more on the aims and objectives as well as research approaches.
2. Still a suggestion from reviewer C, the “theoretical framework” has been added and separated in a new section.
3. Some spelling and chemical formula errors have been fixed.
4. The writing style has been adjusted to the template from reviewer D.

The Development of Two-Tier Instrument Based On Distractor to Assess Conceptual Understanding Level and Student Misconceptions in Explaining Redox Reactions

Lukman Abdul Rauf Laliyo
Universitas Negeri Gorontalo
Gorontalo, Indonesia

Deasy Natalia Botutihe and Citra Panigoro
Universitas Negeri Gorontalo
Gorontalo, Indonesia

Abstract. Fifteen distractor two-level multiple choice items were developed as diagnostic instruments to evaluate the level of conceptual understanding and structure of students' misconceptions in explaining redox reactions. Questions at the first tier (Q1) assess the level of knowledge, and questions at the second tier (Q2) assess the level of reasoning of students. This instrument was given to 1150 participants. The participants were 11th grade students, from eight senior high schools, in the Eastern part of Indonesia. The collected data was analyzed using the Rasch model approach. The results of this study provide diagnostic and summative information on the progressiveness of student learning outcomes, as well as evidence of empirical validity and reliability of measurement. In addition, by comparing the size of items Q1 with Q2, it was found that the level of student knowledge is not always proportional to the level of reasoning, even in some cases, the level of knowledge is lower than the level of reasoning, and vice versa. The results of the investigation using the option probability curve; it was revealed that there were students' misconceptions and inconsistencies about the concepts of reduction, oxidation and oxidation numbers. This result confirms why students have difficulty interpreting and converting redox reaction equations.

Keywords: two-tier instrument; distractor; understanding; reasoning; redox reaction.

Introduction

The framework of this research is based on constructivism approaches, which assume that students actively build their own conceptual understanding. In this

context, teachers play an important role in facilitating students to develop their learning creativity, to construct their knowledge and conceptual understanding as optimally as possible. Early concepts that have been understood by students will determine the learning outcomes (Rahayu, 2017). However, this effort is not easy for students. Many studies have showed that students must study hard to understand chemical concepts deeply and meaningfully (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013). Students can find it difficult to understand chemical concepts due to their complex and abstract characteristics (Johnstone, 1991; Taber, 2013; Gabel, 1999). Therefore, students develop their own conceptions, which tend to be different from the scientific community presumption. Coherent but wrong conceptual structures that have been firmly embedded in the minds of students, especially related to daily experience are misconception (Chandrasegaran, Treagust, & Mocerino, 2007; Johnstone, 2006; Taber, 2009; Taber, 2013).

One of the core concept within Chemistry curriculum in middle school in Indonesia is redox reaction. Chemical education experts in Indonesia have extensively reported that students find it difficult to understand this concept. Students find it difficult to show good quality of conceptual understanding through utilization of two tiers distractor-based instrument. This study was aimed at evaluating the conceptual level of understanding of the students in describing the redox reaction, which focuses on two important aspects, knowledge and reasoning. Evaluation of knowledge relates to measurement of students' mastery toward the content of oxidation, reduction, and the shifts in oxidation number concepts. Whereas, reasoning evaluation deals with the measurement of students' ability in providing reasons/feedbacks to back up their understanding. The good quality of students' conceptual understanding is when students were able to know and describe their knowledge properly. The main focus of this study was aimed at describing two issues, (1) how effective is the two-tier distractor-based instrument to measure level of conceptual understanding and diagnose the structure of students' misconception? And (2) how is the description of conceptual level of understanding and the structure of students' misconception in describing the redox reaction? Therefore, this quantitative research is non-experimental in combination with a qualitative descriptive study. It is not be manipulated nor regulated the process or learning material.

Theoretical Framework

Scientific education researchers have developed many types of instruments to diagnose student misconceptions. One of them that is often used today is a two-tier multiple choice diagnostic instrument (Treagust, 1988; Chandrasegaran et al., 2007; Tüysüz, 2009; Femintasari, 2015). This instrument cannot only evaluate conceptual understanding, but also can diagnose student misconceptions. Qualitatively, this diagnostic instrument is relatively effective in providing information regarding the way of students think and who have misconceptions. However this instrument tends to have limitations, such as the weak internal consistency. Because the instrument is not in the same measurement dimension, then it is difficult to use for summative measurement (Lu & Bi, 2016).

Misconceptions studies tend to have interesting progress as researchers have new concepts of this issue. Research findings showed that the fallacies are often resistant and still persist even after formal learning. This new concept of misperception has been evolving since the early 2000s, along with the progression of knowledge and reasoning, which varied according to the level of understanding of students (Aktan, 2013). Two-level diagnostic instruments can diagnose misconceptions, but cannot measure the level of conceptual understanding and evaluate the development of students' knowledge and reasoning. As a result the information produced is relatively limited for teachers for making decisions (Wilson, 2008).

Furthermore, Rasch model has been introduced as a measurement method that can integrate the diagnostic evaluation approach of students' misconception with summative evaluation of students (Liu, 2012). Although there are still studies that develop formative evaluations to investigate the progression of student learning outcomes (Claesgens, Scalise, Wilson, & Stacy, 2009; Hadenfeldt et al., 2013), several other studies have developed integrative diagnostic and summative approaches (Hoe & Subramaniam, 2016). In fact, there are studies have developed instruments based on Rasch modeling, which integrate the diagnostic evaluation approach to misconception with summative evaluation in chemistry learning (Chi, Wang, Luo, Yang, & Huang, 2018).

This study was designed using a Rasch-based instrument, in the form of a two-level distractor-based multiple choice item. Multiple choice items are often criticize as it tests only the facts and due to its inability to diagnose student's understanding (Klassen, 2006). These downsides are solved using the power of diagnostic selected items with two stratified questions (Treagust, 1988), and utilizing answer distractor choices on each item (Sadler, 1998; Herrmann-Abell & DeBoer, 2011). The aim of this research was to develop a distractor-based two-tier multiple-choice diagnostic instrument (TMDI), to evaluate the level of conceptual understanding and structure of students' misconceptions in explaining the redox reaction.

Method

There are three important components that are interrelated in the progression of measurement instruments: cognition, observations and interpretations. Cognition refers to theories or constructs about how students' understanding develops related to what they learn. Observations refer to student performance based on the type of problem solving or task and situation at the time of measurement. Interpretations refer to the results of data analysis, in the form of a statistical model as a summary of students' understanding patterns (NRC, 2001). These three components become an important part of the development stage of the instrument proposed by Wilson (2008), which includes developmental progress variables, item models, outcome space, and measurement models. The first stage is developmental progress variable, with regard to the developmental evaluation variables and progressiveness of student

learning outcomes that only focus on one characteristic to be measured. The second stage, the items model relating to the form of items that designed to obtain suitability between the measurement and various understanding diagnoses. Each item is designed to provide a response that is diagnosing student understanding, at least one level for one construct map. The third stage, the outcome space, is student learning outcomes that are categorized for all items related to certain progress variables. The fourth stage, the measurement model, in this study uses the Rasch model.

The diagnostic instruments developed in this study were adapted from the framework of developing two-tier diagnostic instruments reported by Treagust (1988) and Chandrasegaran, Treagust, & Mocerino (2007). Distractor item design were adapted from Herrmann-Abell & DeBoer (2011) and Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann (2013), so that it can be used to diagnosis what students do, what don't they know, the misconceptions and alternative ideas that they have understood. The data obtained were analyzed using the Rasch model approach, through the four steps described below.

Description of students' conceptual understanding

The first step is to describe students' conceptual understanding. This requires a construct map that provides the substantive definition and the qualitative level of the construct, according to the complexity of the construct of the variable being measured (Wilson, 2009). The progress of students' understanding on how and what they understand is related to the construct of definition map (Wilson, 2012). The map definitions of students' conceptual understanding in explaining the redox reaction was elaborated from the Chemistry Curriculum of High Schools in Indonesia. Learning redox reaction concept intended to improve students' ability to identify redox reactions, oxidation numbers and able to analyze the redox reaction based on changes in oxidation numbers, obtained from experimental data or through experiments (Ministry of Education and Culture, 2016). The progressiveness of student learning outcomes is the comprehensive and interrelated students' conceptual understanding in explaining the redox reaction. The word "explain" in the definition, implies that students must be able to interpret and develop their own reasoning, in which used to solve problems related to redox reactions. A map of the definition of level of conceptual understanding is shown in Table 1.

Table 1. The level of students' conceptual understanding in explaining the redox reaction

Representation Level	Conceptual Understanding of Redox Reaction	Item
Macro, submicro, symbolic	9. Students can determine non-redox reaction by analyzing submicroscopic diagram.	14E
	8. Students can convert submicroscopic diagrams of iron rusting reaction to symbolic chemical reaction equation	13D
	7. Students can determine autoredox reaction by analyzing submicroscopic diagram that illustrated by reaction between bleaching powder	15F

	and HCl and dissolve in water.	
Symbolic	6. Students can determine elements that undergoing oxidation state change in A and B redox reaction	4skAb, 6skBb
	5. Students can determine substances that act as oxidator and reductor in A and redox reaction.	3skAs, 7skBs
Submicroscopic	4. Students can determine submicroscopic diagram of irreversible reaction between Zn dan CuSO ₄ .	10sbBx
	3. Students can determine net ion equation by submicroscopic diagram of A dan B redox reaction.	2sbAi, 9sbBi
	2. Students can interpret submicroscopic diagram of transfer elctron in B and C redox reaction.	8sbBe, 12sbCe
Macroscopic	1. Students can identify color change of solution due to redox process:	
	A. Iron metal dipped in CuSO ₄ solution	1mA
	B. Zink metal dipped in CuSO ₄ solution	5mB
	C.. Iron metal dipped in H ₂ SO ₄ solution	11mC

Table 1 presents nine levels of conceptual understanding of students, in interpreting and converting redox reactions A, B and C, through macro, submicro, and symbolic representations. The purpose using representations are (1) Macroscopic representation to measure students' abilities in describing real and visible redox reactions, according to everyday student experience, or at least experience when observing change in redox reactions (such as discoloration of solutions, gas formation, deposits) through laboratory experiment. (2) Submicroscopic representation to measure their ability to interpret and explain redox reactions at the particulate level, using images of atoms, molecules and ions. (3) Symbolic representation to measure students' abilities using chemical symbols, formulas and equations, molecular structure, and diagrams. (4) Through macro, submicro and symbolic representations, can be measured students' ability to explain redox reactions based on the relationship of the three levels of representation. Therefore the ability of students to interpret and convert macro, submicro and symbolic representations is closely related to the construction of students' conceptual understanding. If students do not have proper understanding of the redox reaction, then students may difficult to explain the redox reaction using representation.

Acquiring chemical knowledge without clear conceptual understanding can be confusing for student, because simultaneously they are required to be able to explain the concept using all three levels of representation (Chandrasegaran et al., 2007). Students must explain reaction changes at the macroscopic level, then explain changes at the particulate level, and the last explain using symbols and formulas (Gabel, 1999). For example, the redox reaction A. When iron (Fe) is dipped in CuSO₄ solution, macroscopically, the color changes reddish. The next level is students must be able to interpret the "macro" changes to the particle level (using a submicroscopic diagram), and then symbolically write down the redox reaction equation, into: $\text{Fe(s)} + \text{Cu}^{2+}(\text{aq}) \rightarrow \text{Fe}^{2+}(\text{aq}) + \text{Cu(s)}$ (symbolic). Students competency to explain the redox reaction A, by interrelating the three levels of representation, can only be achieved when students are facilitated to experience direct learning with the real world (macro), then submicro and

symbolic explanations are gradually introduced. Therefore getting knowledge by students becomes more structured, systematic and constructive that stored in the long-term memory of students (Tsaparlis, 2009).

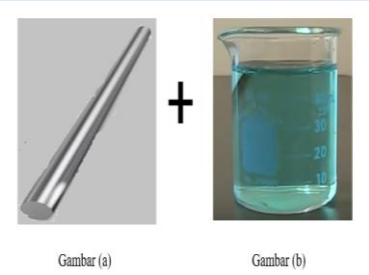
<p>Consider the following picture:</p> <div style="text-align: center;">  <p>Gambar (a) + Gambar (b)</p> </div>	<p>Question:</p> <p>Q1. Figure (a) show iron metal, and (b) is CuSO_4 solution with blue color. If iron metal dipped into CuSO_4 solution, what is the change that will be occur?</p> <p>A. Iron metal does not dissolve in CuSO_4 solution</p> <p>B. Iron metal dissolved like sucrose dissolved in water</p> <p>C. Iron changed to reddish color.</p> <p>D. CuSO_4 solution does not undergo any change</p> <p>Q2. Which one of the following is the reason of your answer:</p> <p>A. Iron metal undergo oxidation</p> <p>B. Iron metal does not react with CuSO_4 solution</p> <p>C. Iron metal is easier to gain electron.</p> <p>D. Iron metal dissolve in CuSO_4 solution</p>
--	---

Figure 1. Item 1mA (Q1/Q2)

Item Design

The third step is data collection using a two-tier multiple choice instrument, which is given to 1150 tenth grade students from eight senior high schools in the northern region of Sulawesi, Indonesia. Students are chosen randomly, and agree to be voluntarily respondents. All students have learned redox reaction concept according to the Chemistry Curriculum Standards in Indonesia. Process standards and learning outcomes have been elaborated completely in this curriculum. Chemistry teachers are required to develop and use problem solving-based instructional strategies. This strategy is used to encourage students to design and make redox reaction experiment, practice to solve problems, explain the process of electron transfer, in the reaction between iron metal dipped in a solution of CuSO_4 and H_2SO_4 , discuss redox reaction equation, and determine elements that are undergoing a changes in oxidation number. Based on this curriculum, it can be understood that by learning process students may form their conceptual understand of redox reactions in a gradual,

constructive, systemic and interrelated manner. Likewise, it can be explained when students have many misconceptions, indicating that students do not experience a whole and meaningful learning process. If these assumptions can be proven empirically, the instruments developed in this study have good predictive validity.

Data have been collected for three months, in each school. Students respond manually through a written answer sheet. The test time is 45 minutes and was supervised by teacher. All students are told to answer all questions in the instrument. After testing, all instrument texts are collected directly by the researcher. There are no instrument texts that are not collected.

Measurement Model

The fourth step is use the Rasch analysis model approach to calibrate the difficulty level of the item and students' ability level on the same interval scale. Raw data is inputted in Excel format in the form of dichotomous data, then converted into interval data, using WINSTEPS version 3.75 model software (Linacre, 2012; Bond and Fox, 2015). Raw data input considers the item model that containing two questions Q1 / Q2. Student responses on each item in the first tier (Q1) and in the second tier (Q2) were analyzed separately.

The Rasch model, combine an algorithm 'i' (as the results of probabilistic expectations of items) and students 'n', which are mathematically stated by Bond and Fox (2015), as: $P_{ni} = \frac{1}{1 + e^{-(\beta_n - \delta_i)}}$, where: P_{ni} ($X_{ni} = 1 / (\beta_n, \delta_i)$) is the probability of students "n" in items "i" to get a correct answer ($x = 1$); with student ability β_n , and difficulty level of items δ_i . The above equation can be further simplified by entering the logarithmic function and making it: so that the probability of success is: the probability of success equals to the ability of the student minus the difficulty level of the item. It is important to underline that students' ability and items are stated at the same and independent intervals. The level of students' abilities and difficulty level of items is measured in units of logarithms called odds or logs, which can vary from -00 to +00 (Herrmann-Abell & De Boer, 2011; Sumintono & Widhiarso, 2015).

Rasch models can evaluate the level of understanding, reasoning of students and the difficulty level of items. To test efficiency of the instrument indirectly is by comparing the distribution of difficulty items with the level of understanding and reasoning of students. This is different from the classical test theory approach, which cannot diagnose the level of understanding of students, because the difficulty level of items is only based on raw data (Lu & Bi, 2016). Student with high ability will be able to answer items with a lower level of difficulty.

Result and Discussion

Reliability

Reliability explains how far measurements produce information that is consistent in revealing latent traits. Can the TMDI instrument measure the unidimensional nature of latency, namely students' conceptual understanding in explaining the redox reaction (Sumintono & Widhiarso, 2015). The measurement reliability of the person item is based on the person separation index and item separation index. The Separation index can also be converted to Cronbach's equivalent value from 0-1. Table 2 shows a summary of fit statistics, which are compatible with the Rasch dichotomous model.

Tabel 2. Summary of fit statistics

	Person (N=1150)	Item (N=30)
Reliabilitas	0,58	0,99
Infit MNSQ	1.00	1,00
Outfit MNSQ	1.03	1.03
Separation index	1.18	10,14
Cronbach Alpha	0,61	-

From Table 2, it is known that the person separation index is 1.18 which is equivalent to the reliability value of person 0.58. This value indicates that the reliability of the person is relatively weak. This means that the consistency of students' responses to TMDI test items is relatively weak. Cronbach Alpha Coefficient value of 0.61 indicates that the interaction between 1150 students with 30 items is fairly adequate. However, this fact will not affect the teacher's decision to develop students' abilities to be better (Wei, Liu, Wang, & Wang, 2012; Lu & Bi, 2016). The item separation index value (10.14) is equivalent to the reliability value (0.99). This means that the consistency of the item is classified as very good (special), or the item meets the requirements of unidimensionality. This indicates that the item is able to define latent variables very well. The infit and outfit values of most items are within the acceptable range for multiple choice tests (Abell & DeBoer, 2011; Bond and Fox, 2015).

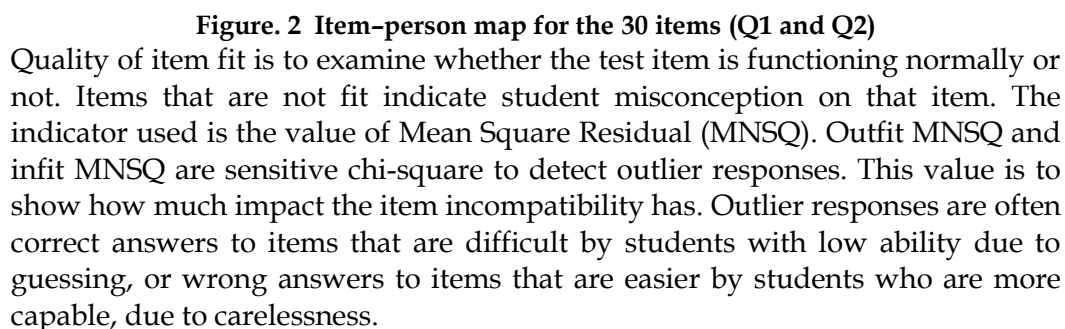
Validity

Validity describe how well measurements are carried out according to what should be measured. The validity of the TMDI test was tested by the suitability analysis of the Rasch model item. If the test is able to measure four levels of a student's ability to explain the redox reaction, then this test has good construct validity (Linacre, 2012). The results of the analysis are presented in Table 3.

Table 3. Item Fit Statistics^a

Item	Measure	Model SE	INFIT		OUTFIT		PTMEA
			MNSQ	ZSTD	MNSQ	ZSTD	Corr
1mA(Q1)	-0,73	0,06	0,93	-3,9	0,92	-4,0	0,41
1mA(Q2)	-1,00	0,06	0,94	-3,0	0,91	-3,5	0,40
2sbA(Q1)	-1,02	0,06	0,84	-7,7	0,81	-7,8	0,53
2sbA(Q2)	-0,56	0,06	0,95	-2,9	0,93	-3,5	0,39
3skAs(Q1)	0,18	0,06	1,01	0,6	1,05	1,7	0,25
3skAs(Q2)	0,93	0,07	1,07	1,7	1,12	2,1	0,13
4skAb(Q1)	-1,43	0,07	0,89	-4,1	0,80	-5,7	0,47
4skAb(Q2)	-0,15	0,06	0,90	-5,4	0,89	-4,8	0,45
5mB(Q1)	0,80	0,07	1,04	1,0	1,07	1,4	0,19
5mB(Q2)	-0,71	0,06	0,96	-2,4	0,95	-2,6	0,37
6skBb(Q1)	-0,51	0,06	0,88	-6,8	0,86	-7,3	0,48
6skBb(Q2)	-0,44	0,06	0,88	-6,8	0,87	-6,6	0,48
7skBs(Q1)	0,46	0,07	1,15	5,0	1,23	5,5	0,03
7skBs(Q2)	0,99	0,08	1,16	3,6	1,36	5,6	-0,05
8sbB(Q1)	0,17	0,06	1,04	1,8	1,07	2,1	0,22
8sbB(Q2)	0,52	0,07	1,04	1,5	1,05	1,2	0,20
9sbBi(Q1)	-0,40	0,06	0,88	-7,2	0,88	-5,9	0,48
9sbBi(Q2)	1,27	0,08	1,13	2,5	1,41	5,2	-0,05
10sbBx(Q1)	0,25	0,07	0,85	-6,4	0,83	-5,3	0,50
10sbBx(Q2)	-0,35	0,06	0,96	-2,1	0,95	-2,5	0,36
11mC(Q1)	1,21	0,08	1,16	3,0	1,33	4,5	-0,05
11mC(Q2)	-0,88	0,06	0,89	-6,0	0,86	-6,5	0,48
12sbC(Q1)	0,97	0,08	1,06	1,4	1,11	1,9	0,14
12sbC(Q2)	-0,34	0,06	1,22	9,9	1,22	9,9	-0,01
13D(Q1)	0,31	0,07	1,05	2,0	1,08	2,3	0,19
13D(Q2)	-0,34	0,06	1,12	6,2	1,13	6,0	0,13
14E(Q1)	0,10	0,06	0,94	-2,5	0,93	-2,5	0,37
14E(Q2)	0,14	0,06	0,94	-2,6	0,95	-1,6	0,36
15F(Q1)	0,22	0,07	0,96	-1,8	0,97	-0,8	0,34
15F(Q2)	0,37	0,07	1,18	6,3	1,20	5,2	0,01

^aN = 1150. MNSQ values of 0.70–1.30 indicate the acceptable fit range; infit (weighted) or outfit (unweighted) values outside this range indicate poor fit of the data to the Rasch model. ☐



The ideal value expected from MNSQ is 1.0. In addition, another indicator is point measure correlation (PTMEA Corr). PTMEA Corr is a score and person measure score correlation, whose value must be positive and not close to zero (Bond and Fox, 2015). The criteria used is the means-square outfit (MNSQ) value: $0.5 < y < 1.5$; z-standard outfit: $-2.0 < Z < +2.0$, and PTMEA Corr: $0.4 < x < 0.8$. If the three criteria are not fulfill, the items are not good enough therefore need

further investigation (Boone, Yale, & Staver, 2014). Based on Table 3, it is known that no item does not meet the three criteria. However, there are four items, namely 12sbC (Q2), 11mC (Q1), 9skBi (Q2), 7skBs (Q2)) which are negative, and item 15F (Q2) which is close to zero. The five items require further investigation

After testing the suitability of the item, we examine level of difficulty of the item with the level of ability of students, using "Wright Map" (person-item map), is presented in Figure 2. This map shows the distribution of students' abilities (on the left side) and the distribution of difficulty items (on the right side) vertical lines. Distribution of low-ability students and low difficulty items are at the bottom of the map. Distribution of high-ability students and items with high difficulty levels are at the top of the map. The item's horizontal line mean (0.0 logit) is higher than the person mean (-0.52 logit). All items can reach the entire scale of the student's logit ability. However, there are six items that are too difficult for students, above the 1 SD scale (standard deviation). They are 5mB/Q1 (0.80), 11mC/Q1 (1.21), 12sbC/Q1 (0.97), 3skAs/Q2 (0.93), 7skBs/Q2 (0.99), and 9skBi/Q2 (1.27). These items need to be adjusted and examined further.

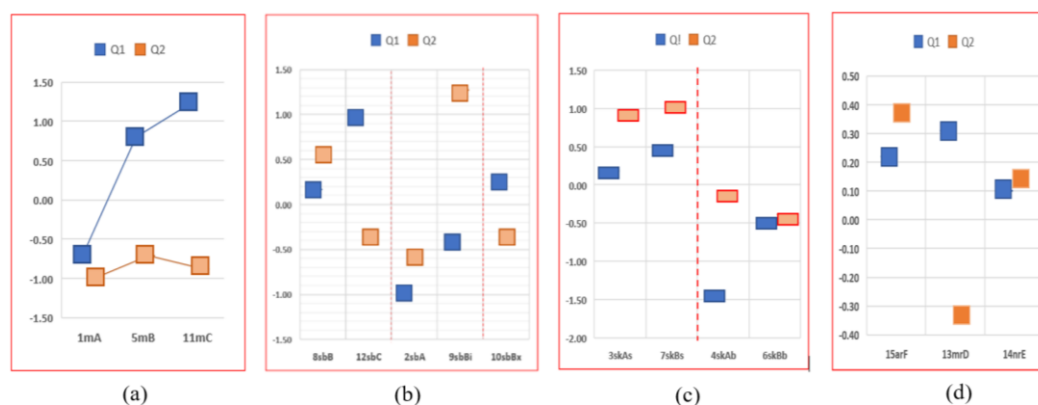


Figure 3. Comparison of the size of items in the first tier (Q1) and in the second tier (Q2)

Level of conceptual understanding

The level of conceptual understanding of students is distinguished by the level of knowledge (Q1) and the level of reasoning (Q2) in each construct studied (item). This difference is explained based on the comparison of item size (measures items). Students' misconceptions are examined by testing response patterns and option probability curve patterns. The response pattern test is estimated from the results of the distractor test, which is based on the acquisition of the average ability of the item. If the average ability value rises, the effect of the distractor works. This means that students give the answers are not by guessing. However if the average ability value drops, the effect of the distractor does not work meaning that the item is easy to guess. Student's answers are not representing the knowledge they have learned, because students guess the answer. The decrease in average ability value indicates that there are low-ability students who are able to answer items correctly and / or there are high-ability

students who answer incorrectly on items that should be answered correctly, and this is not favorable for measurement (Smiley, 2015).

The option probability curve is used to explore the pattern of student responses. Curves that correspond to the probability of the correct answer will usually increase, and curve due to distractor will usually decrease with the increase in students' abilities (Haladyna, 2004). In general, for items that are driven by a distractor, the curve obtained does not correspond to the probability curve of the correct answer (Sadler, 1998). Therefore, each student's answer are analyzed separately, to represent data more accurately and provide additional information about the aspects of misconception students have.

The probability choice answer curve provides information on the possibility of students choosing on each answer option provided, as a function of the level of knowledge and the level of reasoning in the construct being measured. In this case, there are four curves, where one curve represents one answer. It means that students who choose one answer choice, have certain conceptual understanding of the construct being measured. Low-ability students related to measured constructs will be interested in certain misconceptions, and high-ability students will be attracted to other misconceptions (Herrmann-Abell & DeBoer, 2011). In addition, the probability curve is presented through visual images, to show the distribution of the correct answers and those experiencing misconceptions in the whole of the spectrum of students' conceptual understanding. Thus it can be identified which form of the curve is appropriate, or is not appropriate, as an indication of the misconception of an item. Forms of curves that are not appropriate, can also show a certain structure of misconceptions, with patterns that are repetitive, consistent and will disappear sequentially, along with increasing student abilities. Diagnosis of misconception structures in an accurate way, allows teachers to develop more effective learning strategies. In this article, an example of how Rasch modeling and the probability option curve can be used to reveal the structure of students' misconceptions in explaining the types of redox reactions.

Level of macroscopic reasoning

The level of macroscopic reasoning is the students' ability to describe the color change of the solution due to A, B, and C the redox reaction processes, that was evaluated with items 1mA, 5mB, and 11mC. In the first tier (Q1) students are asked to determine what changes they know, if reactions A, B and C occur. In the second tier (Q2) students are asked to state the reason for the answer in the first tier. The difference in the level of knowledge and reasoning of students on each item is determined by the size of the item. The larger the item size, the higher the difficulty level of the item, the lower the level of knowledge and reasoning of students.

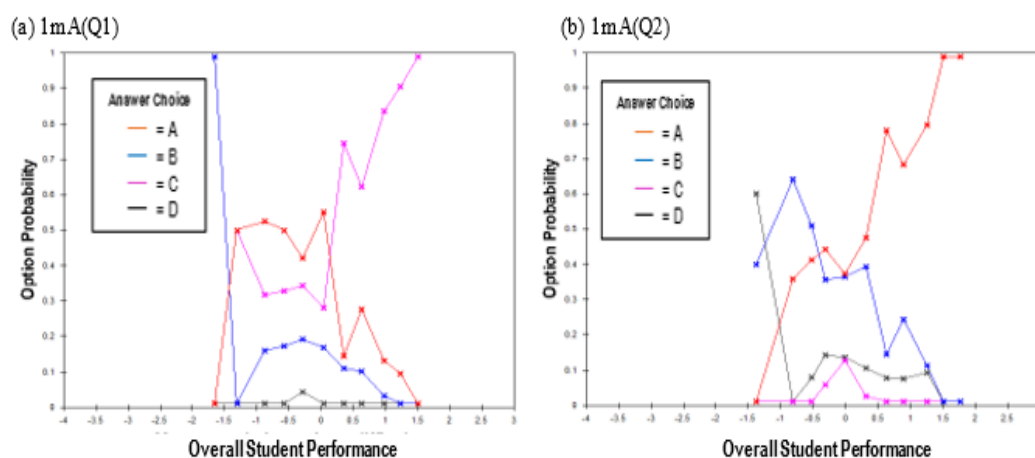


Figure 4: Option Probability Curve (a) item 1mA(Q1); (b) item 1mA(Q2)

Figure 3a shows that the size of items in the first tier (Q1) is in the range of -0.55 logit to 1.25 logit, while in the second tier (Q2), no more than -0.50 logit. It means, empirically the level of student knowledge is lower than the level of reasoning in identifying the changes colors of the solution of A, B and C redox reaction at the macroscopic representation level. To explain this fact, option probability curve is used for item 1mA (Figure 4).

For curve 4 (a), low-ability students (between -1.5 to 0.5 logit) tend to choose the distractor answer option, namely A (iron will not dissolve in CuSO_4), and B (Iron dissolved in CuSO_4 , such as as sugar dissolves in water). It is seen that the peak of the A curve is more dominant than B. Students with moderate ability levels (more than 0.5 logit) tend to choose the right answer C (the solution turns reddish). This pattern of misconception shows evidence of the lack of acquisition of student knowledge about the redox reaction, and develops logical misconceptions, iron is insoluble in CuSO_4 , due to the hard properties of iron. Through the 4c curve, it is known that students with ability levels (-1.5 to 1.5 logit) are more dominant choosing the answer distractor B option (Iron does not react with CuSO_4 solution). Students with an ability level of more than 0.5 logit, tend to choose the correct answer option A (iron undergoing an oxidation reaction). It appears that there is the same student response pattern, stating that iron is insoluble in CuSO_4 . This pattern of misconception can be understood, because students with limited levels of knowledge, incomplete learning experiences, tend to choose to what is physically visible. This is an alternative framework of students (Johnstone, 1991; Lu & Bi, 2016; Taber, 2013), understanding that contains misconceptions. This result support the studies reported by Chandrasegaran et al (2007) and Herrmann-Abell & DeBoer (2011). This result also describe that macroscopic factual knowledge and students' rational reasoning are out of sync. It is possible that students never make direct observations through experiments to identifying the color changes in the solution of the redox reaction process.

Submicroscopic reasoning level

The level of submicroscopic reasoning is related to student ability to explain the redox reaction at the particle level. Based on item size (Figure 3b), interesting facts were found, first, in the first tier question (Q1), the item size 12sbCe (0.97) is greater than 8sbBe (0.17), while in the second tier (Q2), item size 12sbCe (-0.24) is smaller than 8sbBe (0.52). This means that the level of student knowledge is lower than the level of reasoning, in explaining sub-microscopically the electron transfer process in the redox B and C reactions. Second, in the first tier (Q1) item size is 2sbAi (-1.02) and 9sbBi (-0.40) very low, so in the second tier (Q2) items 2sbAi (-0.56) are low. That is, the level of students' knowledge and reasoning in expressing the net ion reaction equation through the interpretation of the submicroscopic diagram of the redox A and B reactions, is very low. Unless the item size is 9sbBi (1.27), which is very high, it's hard to explain because this item's distractor option doesn't work. Third, the size of items is 10sbBx in the first tier (0.25) while in the second tier (0.35). This means that the level of student knowledge is lower than the level of reasoning, in determining the submicroscopic diagram of the reaction of Zink metal dipped in CuSO₄ solution, if the reaction is complete (irreversible). These three facts prove that submiroscopically, students do not have good knowledge and reasoning. There is a gap between the level of knowledge and the level of reasoning. Conceptual understanding of students tends to be inconsistent and partial.

To explain how students develop their understanding at the submicroscopic level, it can be exemplified through the analysis of item response patterns 2sbAi. This item tests the ability of students to interpret the submicroscopic diagram, the redox A reaction, which is iron metal dipped in a solution of CuSO₄ (Figure 5). Based on Figure 5, in the first tier question (Q1), students are asked to write net ion reaction equations, as in the second tier (Q2), students are asked to give the reasons of answers in the first tier. The option probability curve form for item 2sbAi is shown in Figure 6.

Curve 6a, shows that students with low ability levels (smaller 0.0 logit than the ability of students as a whole) tend to choose the answer distractor C ($\text{Fe}^{2+}(\text{s}) + \text{Cu}(\text{aq}) \rightarrow 2\text{Fe}(\text{aq}) + \text{Cu}(\text{s})$) and B ($\text{Fe}(\text{s}) + \text{Cu}(\text{aq}) \rightarrow \text{Fe}^{2+}(\text{aq}) + \text{Cu}(\text{s})$). C and B are answers that contain misconceptions. Students with a level of ability greater than 0.0 logit, choose the correct answer A ($\text{Fe}(\text{s}) + \text{Cu}^{2+}(\text{aq}) \rightarrow \text{Fe}^{2+}(\text{aq}) + \text{Cu}(\text{s})$). On the other hand, on the 6b curve, students with low ability levels (less than -0.5 logit) tend to choose C option (reaction of iron and CuSO₄ is a redox reaction). Students with abilities ranging from -1.5 to 1.0 logit, tend to choose A (iron accepts electrons from CuSO₄). C and A contains misconceptions. Students with abilities above -0.5 tend to choose the correct answer B (Cu^{2+} ions are reduced to Cu).

Student response patterns on curves 6a and 6b indicate that students have low ability at submicroscopic reasoning. Although the diagram of electron transfer has been given, it cannot help students, because they are more likely to take to what they understand about oxidation and reduction. This is also found in item 9sbi, students are more difficult to express the net ion reaction equation of B reaction than A. In fact, these two items qualitatively have the same degree.

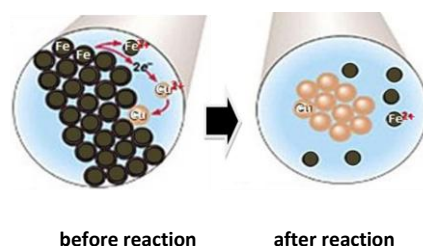


Figure 5. Submicroscopic Diagram of Transfer electron process between iron metal and CuSO₄ Solution

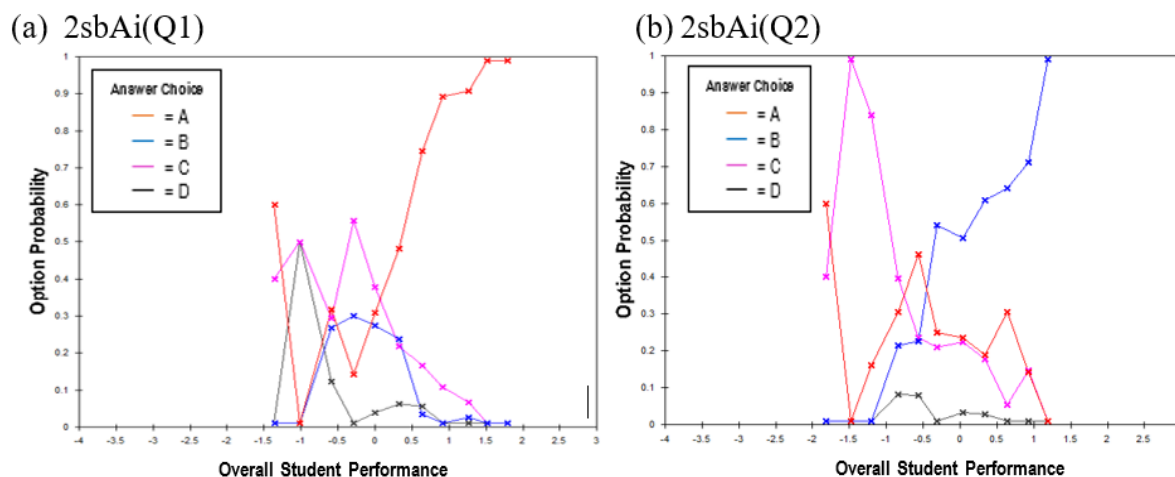


Figure 6. Option porbability Curve Kurva: (a). item 2sbAi(Q1), and (b) item 2sbAi(Q2)

Level of symbolic reasoning

The level of symbolic reasoning is evaluated based on students' ability to determine: (a) species that act as oxidizing and reducing agents (items 3skAs and 7skBs); and (b) elements undergoing oxidation number changes (items 4skAb and 6skBb). Based on item size comparisons (Figure 3c), interesting findings can be explained, as follows: first, item size Q1 (3skAs: 0.18 and 7skBs: 0.46) is greater than item Q1 (4skAb: -1.43 and 6skBb: -0.44). That means, students are more difficult to determine which species act as oxidizing and reducing agents rather than determining the elements that undergo changes in oxidation numbers. Second, in terms of item size Q2, the highest difficulties for students are 7skBs (0.99), 3skAs (0.93), 4skAb (-0.15), 6skB (-0.44) respectively. This means students are more difficult to give a reason in determining the oxidizing / reducing species rather than determining the element that undergoes oxidation number changes. This fact is interesting, because the understanding of the concept of oxidation number and determination of species which act as oxidizers and reducing agents, are interrelated. To elaborate this problem, student response patterns was analyzed, and was exemplified through items 6BB and 7KBs (Figure 7).

(a) 6skBb	(b) 7skBs
Q1. Identify oxidation number change of Zn, in reaction between Zinc metal dipped in CuSO ₄ solution	Q1. Identify reducing agent in reaction between Zinc metal dipped in CuSO ₄ solution
A. 0 to -2. B. +1 to +2. C. +2 to 0. D. 0 to +2	A. Zinc powder B. CuSO ₄ solution
Q2. Which one of the following is the reason of your answer!	C. Copper D. ZnSO ₄ solution
<p>A. $\overset{0}{\text{Zn}}_{(s)} + \overset{+2}{\text{Cu}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_{4(aq)} \rightarrow \overset{+2}{\text{Zn}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_{4(aq)} + \overset{0}{\text{Cu}}_{(s)}$</p> <p>B. $\overset{0}{\text{Zn}}_{(s)} + \overset{+2}{\text{Cu}}\overset{+6}{\text{S}}\overset{-2}{\text{O}}_{4(aq)} \rightarrow \overset{+2}{\text{Zn}}\overset{+6}{\text{S}}\overset{-2}{\text{O}}_{4(aq)} + \overset{0}{\text{Cu}}_{(s)}$</p> <p>C. $\overset{0}{\text{Zn}}_{(s)} + \overset{+6}{\text{Cu}}\overset{+2}{\text{S}}\overset{-2}{\text{O}}_{4(aq)} \rightarrow \overset{+2}{\text{Zn}}\overset{+6}{\text{S}}\overset{0}{\text{O}}_{4(aq)} + \overset{-2}{\text{Cu}}_{(s)}$</p> <p>D. $\overset{0}{\text{Zn}}_{(s)} + \overset{+2}{\text{Cu}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_{4(aq)} \rightarrow \overset{0}{\text{Zn}}\overset{+6}{\text{S}}\overset{-4}{\text{O}}_{4(aq)} + \overset{0}{\text{Cu}}_{(s)}$</p>	<p>Q2. Which one of the following is the reason of your answer!</p> <p>A. Zinc atom undergo oxidation and becomes Zn²⁺ ion</p> <p>B. Cu²⁺ Ion undergo reduction and becomes copper atom</p> <p>C. Reaction between Zn and Copper (II) sulphate produce copper deposit</p> <p>D. Copper (II) sulphate is product of Zinc that undergo reduction</p>

Figure 7. (a) Item 6skBb(Q1/Q2), (b) Item 7skBs(Q1/Q2)

Questions in the first tier (Q1) item 6skBb, basically testing the ability of students to write equilibrium reaction equations of Zinc metal dipped in CuSO₄, and determine how much Zn oxidation number is in the reaction. In the second tier (Q2), students are asked to determine the oxidation number of Zn in the reaction equation. The option probability curve of this item is shown in Figure 8. Curve 8 (a), shows that students with the lowest ability (smaller than -1.0 logit) tend to choose the misconceptions answer A, B and C. The highest probability of choice is A, then decreases with increasing probability of high-ability students (more than 0.5 logit), who choose the right answer D. It means, students' misconceptions, especially in understanding the changes in oxidation number of Zn (from 0 to -2), Zn experiences reduction.

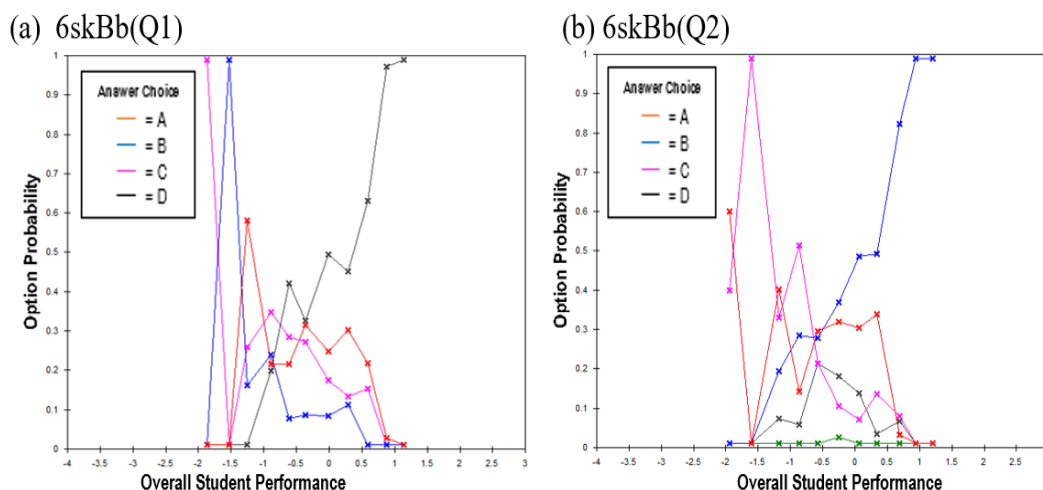


Figure 8. Option probability curve (a). item 6skBb(Q1), and (b). item 6skBb(Q2)

Curve 8(b), confirms this misconception, where low-ability students (small from -0.5 logit), tend to choose C and A, while the probability of students choosing the right answer option B, above ability from -0.5 logit. This fact shows the students' weak understanding regarding the concept of oxidation number. To test this fact, the evaluation continued with the ability of students to determine

which species acted as reducing agents in the reaction of Zinc and CuSO_4 , through item 7skBs, presented in Figure 7 (b).

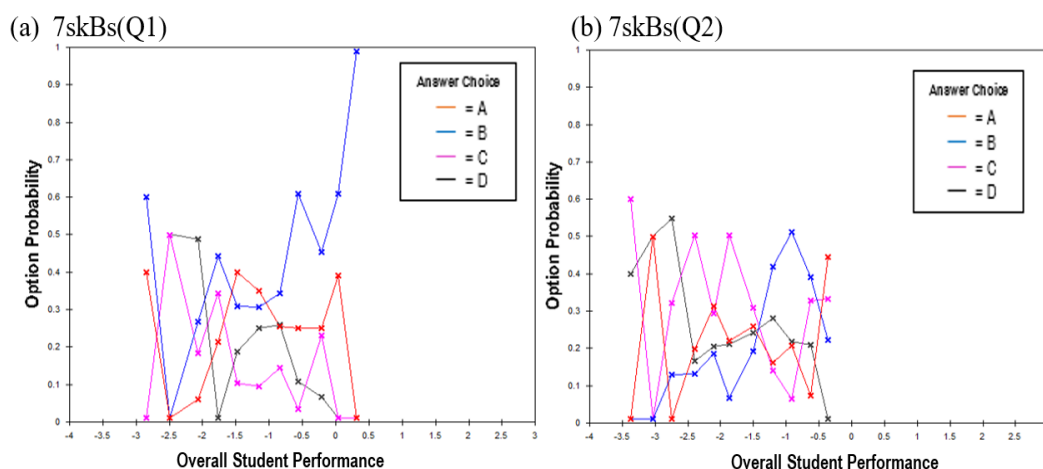


Figure 9. Option probability curve (a). item 7skBs(Q1), and (b). item 7skBs(Q2)

Based on the probability curve item 7skBs (Q1) in Figure 9 (a), it is known that the choice of answer B, which contains misconceptions, is the answer choice with the highest probability chosen by the lowest ability students (-3.0 logit) to 0.5 logit. This curve rises to reach a peak of 1.0 logit. Students argue that the reducing agent in the reaction between Zn and CuSO_4 is a solution of CuSO_4 . The probability of students choosing the right answer A (reducing agent is Zn powder) was chosen by low-ability students, -3.0 logit to 0.3 logit. However, the curve actually shows unusual shape, have two peaks then decreasing. The first peak, at the level of students' ability is smaller than -1.0 logit, and the second peak, at the level of students' ability is greater than 0.0 logit. On the other hand, the 7skBs (Q2) curve in Figure 9 (b) shows that there is no answer of students that reach option probability values up to 1.0 logit. The average of student ability is only able to reach a value of 0.6 logit option probability, from a scale of -3.5 logit to -0.4 logit. These interesting facts can be understood, because they support the previous facts, namely students understanding about the concepts of reduction, oxidation, and oxidation numbers is weak. Students are not able to determine which species act as reducing agents in the reaction of Zinc and CuSO_4 .

Macro, submicro and symbolic levels of reasoning

Macro, submicro and symbolic reasoning levels are evaluated based on students' ability to interpret and convert macro, submicro and symbolic representations, to explain three types of reactions, namely D, E and F. The letter "D" is a symbol of an item that presents a reaction to iron metal; "E" is a symbol of non-redox reaction (i.e., silver nitrate and potassium chloride solution); and F is the auto-redox reaction symbol. These three types of reactions are evaluated, to confirm students' ability to relate their knowledge and reasoning to the three levels of representation.

Figure 3d show the comparisons of sizes 13D, 14E, and 15F. In the first tier (Q1), the difficulty level of the item is $13D > 15F > 14E$. That means, students are more difficult to interpret rusting iron reactions, rather than auto-redox reactions and

non-redox reactions. In the second tier (Q2), the difficulty level of the item is $15F > 14E > 13D$. This means that students are more difficult to explain the auto redox reaction process, rather than a non-redox reaction and a rusting iron reaction. This fact is really interesting, especially item 13D, because between Q1 and Q2 there is a very wide gap. This item asks students to analyze a diagram that illustrates the redox reaction of rusting iron, and then converts to equivalent chemical reaction equation. The result is that students do not have difficulty giving logical reasons related to the iron rusting process with the reaction equation (Q2), but they are very difficult to convert submicro diagrams into reaction equations (Q1). The results of the response pattern analysis found that 34% of students stated the reaction of iron rusting is $4Fe + 3O_3 \rightarrow 2Fe_2O_3$ and 48% of students stated that the metal was reduced and oxygen was oxidation. In fact, the correct reaction is: $4Fe + 3O_2 \rightarrow 2Fe_2O_3$, because iron metal undergoes oxidation, oxygen undergoes reduction. This evidence support previous findings that students' understanding of oxidation, reduction and oxidation numbers concepts is very weak. Then in item 14E, students were asked to analyze the submicroscopic diagram of the reaction process of a solution of silver nitrate and potassium chloride solution which produced white deposits: $AgNO_3 + KCl \rightarrow AgCl + KNO_3$, and asked whether the reaction is redox reactions or not. The results showed that the size of items 14E / Q1 and 14E / Q2 is relatively not much different. That means, students are able to understand and have logical reasons for determining non redox reaction. However, there are still 25% of students stating that the reaction of $AgNO_3 + KCl \rightarrow AgCl + KNO_3$ is a reduction reaction, and 33% of students have reason that there is a change in oxidation number and electrons transfer in the reaction. This fact makes it clear that students are not able to distinguish redox and not redox, due to confusion in determining which species act as oxidizers and reducing agents.

In item 15F, students were asked to analyze submicroscopic diagrams which illustrated the reaction of bleach powder mixed with HCl ($CaOCl_2 + HCl \rightarrow CaCl_2 + Cl_2 + H_2O$), then mixed with water, and produced hydrochloric acid and hypochloric acid ($Cl_2 + H_2O \rightarrow HCl + HClO$) The results of the analysis show that the size of the item 15F/Q2 is more difficult than 15F/Q1. That means, students are very difficult to give opinions regarding the auto-redox reaction diagram, but can determine the auto redox reaction equation. 44% of students are sure that the reaction: $Cl_2 + H_2O \rightarrow HCl + HClO$ is a redox reaction, and 30% of students have reason that the reaction occurs because there is electrons transfer. The correct answer is that the reaction ($Cl_2 + H_2O \rightarrow HCl + HClO$) is an autoredox reaction, because of the oxidation and reduction reactions of the same substance. This result is evidence that student misconceptions tend to be consistent, related to the understanding and reasoning of the concepts of oxidators and reducing agents, changes in oxidation numbers.

Conclusion

The diagnostic instrument developed has construct validity, which can be used to evaluate the level of students 'conceptual understanding, including knowledge and reasoning, and diagnosing students' misconceptions in explaining the redox reaction. At the level of macroscopic reasoning, a repetitive

student response was found, which states that iron is insoluble in CuSO_4 . For low-ability students, this pattern of misconception can be understood, because the limited level of knowledge and reasoning makes students tend to stick to what is physically visible. This misconception may develop from learning experiences that do not go through the process of direct observation or through experimentation. In the level of submicroscopic and symbolic reasoning, there is a gap between the level of knowledge and the level of reasoning of students. This gap is detected again at the level of macro, submicro and symbolic reasoning. Because the acquisition of student knowledge is not complete, it is difficult for students to explain the redox reaction macroscopically, submicroscopically and symbolically. Students have difficulty translating representational diagrams, and converting them into reaction equations. The difficulties lead students to develop misconceptions that are based on their intuitive reasoning, by choosing a choice of distractor answers that are misconceptions. The choice of this answer is detected repeatedly, and tends to come from the same understanding structure, namely the weak of students understanding of reduction, oxidation and oxidation number concept. In turn, all available evidence shows that the progress of the learning outcomes of the students' in redox reactions concepts, overall tends to be low.

Combining the item development process qualitatively with Rasch analysis models, is relatively effective to depth evaluation of the progress of student learning outcomes, reveals students' ways of thinking that are misconceptions, and know what students have not understood. Differences in item size and option probability curves illustrate that the prevalence of the most dominant misconceptions is low ability students. But for certain cases, the prevalence even increases, because the misconception factors tend to be resistant, as seen in item 9sbBe (Q2). The availability of this type of information makes it possible for teachers to classify student understanding based on the level of misconception, so that it is useful in developing instructional strategies. In addition, with the option probability curve on each item, it can be identified an unusual curve, which indicates a problem with the item.

The multiple choice distractor items in this study were developed by adapting some of the results of previous interviews and studies, which were aligned with the targets outcome of chemical learning in Indonesia. The hope is that further research can be carried out, for the development of other material. Although this study does not answer the question of why students experience misconceptions, it is possible for similar items to be developed to help teachers diagnose students' ways of thinking and understanding patterns, so that chemical learning becomes more effective and meaningful.

References

- 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. doi:10.1088/0143-0807/34/1/33
- Boone, W. J., Yale, M. S., & Staver, J. R. (2014). *Rasch Analysis in the Human Sciences*. Rasch Analysis in the Human Sciences. Springer Dordrecht Heidelberg London New York Library: Springer Dordrecht Heidelberg London New York Library.

doi:10.1007/978-94-007-6857-4

- Briggs, D. (2009). The psychometric modeling of ordered multiple-choice item responses for diagnostic assessment with a learning progression. *Learning Progressions in Science (LeaPS) Conference, June 2009, Iowa City, IA, (June)*.
- 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.
- Chi, S., Wang, Z., Luo, M., Yang, Y., & Huang, M. (2018). Student progression on chemical symbol representation abilities at different grade levels (Grades 10–12) across gender. *Chemistry Education Research and Practice*, 19(4), 1055–1064. doi:10.1039/c8rp00010g
- Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The Perspectives of Chemists. *Science Education*, 93(1), 56–85. doi:10.1002/sce.20292
- 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. (1999). Improving Teaching and Learning through Chemistry Education Research: A Look to the Future. *Journal of Chemical Education*, 76(4), 548. doi:10.1021/ed076p548
- 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. doi:10.1021/ed3006192
- Haladyna, T. M. (2004). Developing and Validating Multiple-choice Test Items (3rd Editio, p. 320). New Jersey: Lawrence Erlbaum Associates. doi:10.4324/9780203825945
- 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. doi:10.1039/c1rp90023d
- Hoe, K. Y., & Subramaniam, R. (2016). On the prevalence of alternative conceptions on acid-base among secondary students: Insights from cognitive and confidence measures. *Chemistry Education Research and Practice*, 17(2), 263–282. doi:10.1039/c5rp00146c
- Johnstone, A. H. (1991). Why is science difficult to chemistry 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. doi:10.1039/b5rp90021b
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2016). Silabus Mata Pelajaran Kimia. In *Silabus Mata Pelajaran Sekolah Menengah Atas/Madrasah Aliyah (SMA/MA)*.
- Klassen, S. (2006). Contextual assessment in science education: Background, issues, and policy. *Science Education*, 90(5), 820–851. doi:10.1002/sce.20150
- Linacre, J. M. (2012). *A User's Guide to WINSTEP S® MINI STEP Rasch-Model Computer Programs Program Manual 3.75.0*. Winsteps.com. doi: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. doi:10.1039/c6rp00137h
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (Eds.). (1999). *Assessing Science Understanding: A Human Constructivist View*. Academic Press, San Diego, California : Elsevier 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
- National Research Council. (2001). *Knowing what students know: The Science and Design of Educational Assessment*. National Academy Press (Vol. 19). doi:10.17226/10019
- 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. doi:10.22452/iojel.vol2no1.3
- Rahayu, S. (2017). Promoting the 21st century scientific literacy skills through innovative chemistry instruction (Vol. 020025, p. 020025). Published by AIP Publishing. doi:10.1063/1.5016018
- 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.
- Smiley, J. (2015). Classical test theory or Rasch- A personal account from a novice user. *Shiken*, 19(1), 16-29.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Penerbit Trim Komunikata.
- Taber, K. S. (2009). Challenging Misconceptions in the Chemistry Classroom: Resources to Support Teachers. *Educació Química EduQ*, 4, 13-20. doi:10.2346/20.2003.02.27
- Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, 14(2), 156-168. doi:10.1039/c3rp00012e
- 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. doi:10.1080/0950069880100204
- Trevor G. Bond & Christine M. Fox. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Routledge Taylor & Francis Group (Third Edit, Vol. 44). New York and London: Routledge Taylor & Francis Group. doi:10.1088/1751-8113/44/8/085201
- Tsaparlis, G. (2009). Linking the Macro with the Submicro Levels of Chemistry: Demonstrations and Experiments that can Contribute to Active/Meaningful/Conceptual Learning. In S. A. Devetak, Iztok and Glazar (Ed.), *Learning with Understanding in the Chemistry Classroom* (pp. 41-61). Springer Dordrecht Heidelberg London New York Library. doi:10.1007/978-94-007-4366-3
- 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. doi:10.1021/ed100852t
- Wilson, M. (2008). Cognitive Diagnosis Using Item Response Models. *Zeitschrift Für Psychologie / Journal of Psychology*, 216(2), 74-88. doi: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.
doi: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*, 317–344. Rotterdam: Sense Publishers.
doi:10.1007/978-94-6091-824-7