USING ITEM OPTION CHARACTERISTICS CURVE TO ANALYZE MISCONCEPTIONS ON CHEMICAL REACTIONS OF PRE-SERVICE CHEMISTRY TEACHERS IN INDONESIA

HILMAN QUDRATUDDARSI

FACULTY OF EDUCATION
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

USING ITEM OPTION CHARACTERISTICS CURVE TO ANALYZE MISCONCEPTIONS ON CHEMICAL REACTIONS OF PRE-SERVICE CHEMISTRY TEACHERS IN INDONESIA

HILMAN QUDRATUDDARSI

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF EDUCATION (SCIENCE EDUCATION)

FACULTY OF EDUCATION
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate : Hilman Qudratuddarsi
Matric No. : POC170001
Name of Degree : Master of Science Education
Title of Dissertation

USING ITEM OPTION CHARACTERISTICS CURVE TO ANALYZE MISCONCEPTIONS ON CHEMICAL REACTIONS OF PRE-SERVICE CHEMISTRY TEACHERS IN INDONESIA

Field of Study: Science Education instrument

I do solemnly and sincerely declare that:

1) I am the role author/writer of this work;
2) This work is original;
3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any axcerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorshiphave been acknowledged in this work;
4) I do not have any actual knowledge not do T ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (―UM") , who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
6) I am fully aware that if in the course of making this Work I have infringed any copyright wheter intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM;

Candidate Signature Date: 13 November 2020

Subscribed and solemnly declared before

Witness Signature Date: 18/11/2020

ACKNOWLEDGEMENT

I would like to express my gratitude to God because of His grace and I was blessed by the strength and good health to complete this thesis. I also would like to thanks to my supervisor Dr Renuka V Sathasivam and Dr Hutkemri, for their endless support, guidance, and intellectual support from the first day until I completed this thesis writing. They always welcome to help me even we are in different country whenever I have any confusion on my writing. Although we cannot meet in person due to covid 19, they can still guide me virtually to WhatsApp or Google Meet along the journey.

I am thankful to the lectures and the readers at the University of Malaya for reading my dissertation, commenting on my views, and help me understand and enrich my ideas. I would say special thanks to Dr Hidayah Mohd Fadzil, Dr Rose Amnah Abd Rauf, Dr Leong Kwan Eu and all lectures in the education faculty for endless support to complete my dissertations.

Moreover, I would like say thanks to my family, my parents, my brother, my sister, and my girl Susi Puspitasari for the moral and funding support throughout my journey. I also very grateful for my country, Indonesia for supporting me through 2-years funding supports through LPDP scholarship.

Last but not least, my friends in the university such as kak Yaqin, kak Saras, kak Suci, Norizan, Thilaga, Pan pan, abang Ryan, my housemate Fariz, My LPDP friends and all of them. It was very amazing to know that all of you were ready to listen all my sadness and struggle along my journey.

USING ITEM OPTION CHARACTERISTICS CURVE TO ANALYZE MISCONCEPTIONS ON CHEMICAL REACTIONS OF PRE-SERVICE CHEMISTRY TEACHERS IN INDONESIA

ABSTRACT

Chemical reactions are an important component when studying chemistry but it is difficult to master them. The presence of three-level of representations (macro, sub-micro and symbolic) in chemistry add to this difficulty and as a result, misconceptions do take place. Misconceptions interfere with cognitive development of students and prevent them from having meaningful learning experiences. Thus, identifying accurately these misconceptions are paramount. These misconceptions are analysed using two-tier multiple-choice questions (2TMC). Some researchers argue its limiting diagnostic power and introduced three-tier and four-tier multiple-choice questions (3TMC and 4TMC). However, 3TMC and 4TMC have its own challenges in term of time and resources and is not very helpful in helping teachers to diagnose students' misconceptions. To counter this, researchers tried to enhance the diagnostic power of ordinary multiple-choice questions. This study aims to propose the use of item option characteristics curve (IOCC) to analyze strongest distractors to show misconceptions and promote 2TMC diagnostic power. The study was conducted by adopting the Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI) and 185 Indonesian pre-service chemistry teachers participated in the study. The data was analysed using SPSS and Winstep software to create IOCC. The findings revealed, both IOOC and alternative answers for analysis using traditional method, showed comparable results, implying that using IOCC can detect pre-service chemistry teachers' misconceptions. In addition, IOCC analysis also highlighted aspects such as the unexpected curve after 0 logit of student measure, the inconsistency of alternative answers or distractor with the highest probability, and provided additional information to judge the strongest distractor to show misconceptions. The findings implied that distractor analysis using IOCC reveals information that are more detailed and richer, therefore has the ability enhance diagnostic power of 2TMC. This method is simpler and it allows teachers and students to diagnose their misconception to improve the quality of teaching and learning.

Keywords: chemical reactions, cognitive development, item option characteristics curve, misconceptions**,** pre-service chemistry teachers

MENGGUNAKAN *ITEM OPTION CHARATERISTICS CURVE* UNTUK MENGANALISIS MISKONSEPSI TINDAK BALAS KIMIA BAGI GURU KIMIA PRA PERKHIDMATAN DI INDONESIA

ABSTRAK

Tindak balas kimia adalah komponen penting dalam matapelajaran kimia tetapi sukar untuk mempelajarinya. Kewujudan tiga peringkat perwakilan (makro, sub-mikro dan simbolik) dalam kimia menambah kesukaran ini dan akibatnya, miskonsepsi berlaku. Miskonsepsi mengganggu perkembangan kognitif pelajar dan menghalang mereka daripada mengalami pengalaman pembelajaran yang bermakna. Oleh itu, mengenal pasti secara tepat miskonsepsi ini adalah penting. Miskonsepsi dapat dianalisis dengan menggunakan soalan aneka pilihan dua peringkat (2TMC). Sesetengah penyelidik berpendapat kuasa diagnostik adalah terhad dan memperkenalkan soalan aneka pilihan tiga peringkat dan empat (3TMC dan 4TMC). Walau bagaimanapun, 3TMC dan 4TMC mempunyai cabaran tersendiri dari segi masa dan sumber dan tidak begitu membantu guru mendiagnosis miskonsepsi pelajar. Untuk mengatasi masalah ini, penyelidik cuba meningkatkan kuasa diagnostik soalan aneka pilihan. Kajian ini bertujuan untuk mencadangkan penggunaan *item option characteristics curve* (IOCC) untuk menganalisis distraktor yang paling kuat untuk menunjukkan miskonsepsi dan meningkatkan kuasa diagnostik 2TMC. Kajian ini dilakukan dengan mengadaptasi instrumen *Representational Systems and Chemical Reactions Diagnostic* (RSCRDI) dan 185 guru kimia pra-perkhidmatan Indonesia terlibat dalam kajian ini. Data dianalisis menggunakan perisian SPSS dan Winstep untuk membina IOCC. Dapatan kajian menunjukkan, kedua-dua IOOC dan analisis jawapan alternatif menggunakan kaedah tradisional, memberi maklumat yang agak sama, menunjukkan bahawa IOCC dapat mengesan miskonsepsi guru kimia pra-perkhidmatan. Di samping itu, analisis IOCC juga menunjukkan aspek seperti lengkung yang tidak dijangka selepas 0 logit ukuran pelajar, ketidakselarasan jawapan alternatif atau distraktor dengan kebarangkalian tertinggi, dan memberikan maklumat tambahan tentang kuasa distraktor yang paling kuat untuk mengesan miskonsepsi. Penemuan ini menunjukkan bahawa analisis distraktor menggunakan IOCC mendedahkan maklumat yang lebih terperinci dan kaya, oleh itu mempunyai keupayaan meningkatkan kuasa diagnostik 2TMC. Kaedah ini lebih mudah dan ia boleh digunakan oleh guru dan pelajar untuk mendiagnosis miskonsepsi untuk meningkatkan kualiti pengajaran dan pembelajaran.

Kata kunci: tindak balas kimia, *item option characteristics curve*, miskonsepsi

TABLE OF CONTENTS

CHAPTER 5 : DISCUSSION AND CONCLUSION

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

HCl	:	Hydrochloric Acid
Fe	:	Iron
$FeCl_2$	:	Iron(II) Chloride
H_2	:	Hydrogen gas
CH_3COOH	:	Acetic acid
CH_3COO^-	:	Acetic ion
H^+	:	Hydrogen ion
Fe^{2+}	:	Iron(II) ions
Cu^{2+}	:	Copper ion
Zn^{2+}	:	Zinc ion
NaI	:	Sodium iodide
KI	:	Potassium iodide
H_2SO_4	:	Sulfuric acid
KOH	:	Potassium hydroxide
Na^+	:	Sodium ion
NO_3^-	:	Nitrate ion
$NaNO_3$	:	Sodium Nitrate
SO_4^{2-}	:	Sulfuric ion
$FeSO_4$	:	Iron(II) sulfate
Mg^{2+}	:	Magnesium ion

LIST OF ABBREVIATIONS

SC	:	Scientific Conception
MSC	:	Misconception
LK	:	Lack of Knowledge
CRI	:	Confidence rating index
P-tier	:	Phenomenon tier
R-tier	:	Reasoning tier
Item P1-P15	:	Item Phenomenon Number 1-15
Item R1-R15	:	Item Reasoning number 1-15
KTSP	:	Kurikulum Tingkat Satuan Pendidikan
RSCRDI	:	Representational systems and chemical reactions diagnostic instrument
ANOVA	:	Analysis of Variance
SPSS	:	Statistical Package for Social Science
MDDMC	:	Misconception Distractor-Driven Multiple-Choice Questions
MANOVA	:	Multivariate Analysis of Variance
ICC	:	Item Characteristics Curve
IOCC	:	Item Option Characteristics Curve
MCQs	:	Multiple Choice Questions
MNSQ	:	Mean Square
ZSTD	:	Z-standard
Pt Mean Corr	:	Point Measure Correlation
M	:	Mean
SD	:	Standard Deviation

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

The chemical reaction topic is considered paramount in chemistry as prerequisite of subsequent topics in chemistry (Bain, 2017; Bain & Towns, 2018). Thus, sound conceptual understanding of chemical reactions can influence the study of other topics (Ozmen & Ayas, 2003; Salta & Tzougraki, 2011). Without understanding concepts of chemical reactions such as how to write correct chemical reactions, how to balance them or the symbol of compounds may lead to students having difficulties mastering other topics.

Chemical reaction, based on the curriculum of chemistry in Indonesia, is in the various topics such as the atom, chemical bonding, electrolyte, reduction-oxidation, stoichiometry (first year), thermochemistry, the rate of reaction, chemical equilibrium, acid-base, buffer solution (second year) and electrochemistry, corrosion, properties of the metal (third year) (Ministry of Education and Culture, 2016). In Indonesia, high school students (year 10-12) study chemistry and could be taken as a major at the university level. Chemistry is taught as part of an integrated science subject at elementary school and junior high school level (Muttaqin, Wittek, Heyse, & van Duijn, 2019).

Studying chemical reactions require students to explicate concepts and phenomenon at three levels of representations (Chandrasegaran, Treagust, & Mocerino, 2011; Santos & Arroio, 2016). The first representation begins at what students see in the surrounding (macroscopic representation: observable, touchable and smellable) (Chandrasegaran, Treagust, & Mocerino, 2007; Johnstone, 2000;

Taber, 2013), the next level is what they can explain about the event from chemical perspectives (sub-microscopic level: atoms, molecules, ions, and structures) (Johnstone, 2000; Rappoport & Ashkenazi, 2008) and the last level is how they communicate using chemical formulae or chemical equation (symbolic representation: symbols, formulae, equations, molarity, mathematical manipulation, and graphs) (Taber, 2013; Touli, Talbi, & Radid, 2015). The three-level of representation (macroscopic, submicroscopic and symbolic) can be recognized as Johnstone‛s triagonal approach or chemistry triplet (Johnstone, 2000; Taber, 2013; Talanquer, 2011).

When students are required to describe chemical phenomena, students are not apt to connect these three levels of representation often leading to misconceptions (Chandrasegaran, Treagust, & Mocerino, 2009; Gurel, Eryilmaz, & McDermott, 2015; Kaltacki, 2012). Misconceptions are students‛ views of chemical phenomena contradicting with chemist‛s view (Arslan, Cigdemoglu, & Moseley, 2012; Kirbulut & Geban, 2014; Maskiewicz & Lineback, 2013; Romine, Todd, & Clark, 2016).

Often teachers perceive misconceptions either as obstacles or resources. If teachers assume it as an obstacle, they believe misconceptions hinder learning by preventing access to central scientific ideas, blocking the ability of students to understand concepts, and affecting how students acquire new knowledge (Larkin, 2012). Therefore, teachers find ways to detect these misconceptions so that they could be overcome, changed, replaced, avoided and eliminated (Hammer, 1996; Kaltakci & Eryilmaz, 2010; Ozmen & Ayas, 2003; Schultz et al., 2017). The detection is advantageous because its result can be the recommendation to modify delivery and measure learning gain (assessing teaching progress or effectiveness) if there is the administration of both pre-test and post-test (Hasan, Bagayoko, & Kelley, 1999; Liu, 2010; Yalcinkaya, Tastan, & Boz, 2009).

The second view perceives misconception as resources, student's ideas to foster deeper and more meaningful learning by guiding its preparation. As an asset, information on misconceptions can take an explicit role as the mediator to support formative assessment (Black & Wiliam, 1998; Black & William, 1998; Schultz et al., 2017). It refers to an assessment that is specifically intended to generate feedback on performance to improve and accelerate learning (Gikandi, Morrow, & Davis, 2011; Nicol & MacFarlane-Dick, 2006). Moreover, in the view of metacognition which defined by Kuhn (2000) as the enhance of metacognitive awareness of what students believe and how they know in the control of processing new information, students may explicitly compare their conceptions with other ideas when offering explanations, making arguments, and providing justifications (Larkin, 2012). From the brief explanation, it is concluded that misconception detection is vital for improvement in chemistry learning (Ozmen & Ayas, 2003).

Misconceptions can be experienced by in-service teachers, pre-service teachers, and students. However, the study on in-service teachers and pre-service teachers' misconceptions is still limited compared to high school students (Bayraktar, 2009; Mutlu & Sesen, 2016; Taskin, Bernholt, & Parchmann, 2015). This study, indeed, concentrates on pre-service teachers which will be predictors of success in education in the near future (Bayraktar, 2009; Rahman, Zamri, & Leong, 2018). The reveal of their misconception is vital to raise their awareness for improving their future teaching practice and can be a good beginning for the long-term purpose (Bain, 2017; Cox, Steegen, & De Cock, 2016). Some studies also concluded that the kind of student and teacher conception correlate strongly and tend to be similar (Svec, Boone, & Olmer, 1995; Trumper, 2000; Trumper & Gorsky, 1997).

In literature, various instruments have been introduced to analyse misconceptions (Schultz et al., 2017). It is proven by meta analysis study from Soeharto et al. (2019), Gurel et al., (2015) and Wandersee, Mintzes, and Novak (1994) by reviewing 111 articles (2015-2019) 237 articles (1980-2014) and 103 articles (before 1994) on diagnostics instrument with the trends as in Figure 1.1.

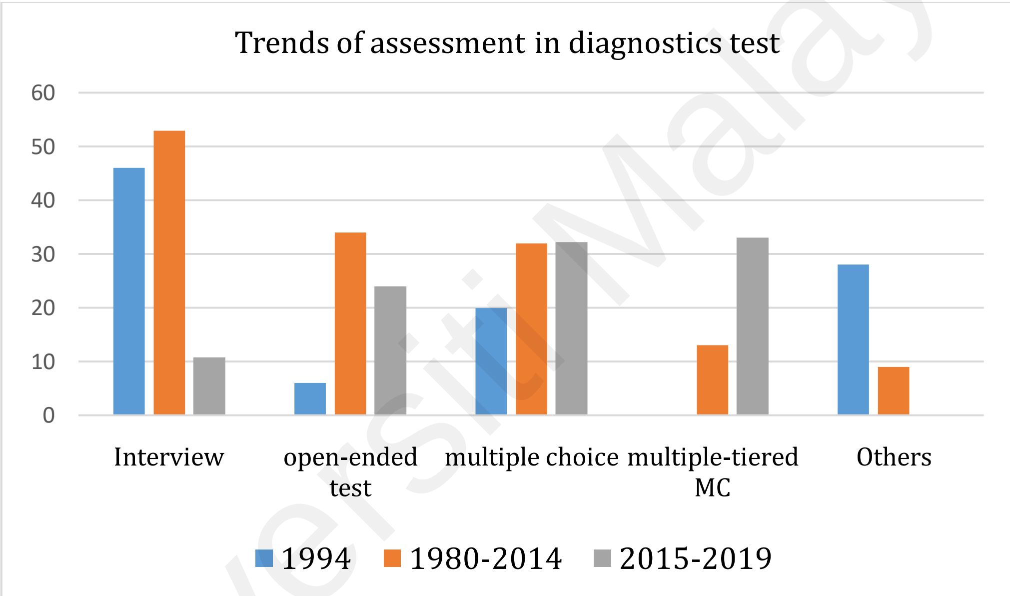


Figure 1.1 Trends of assessment in diagnostics test

Multiple-tiered instruments are popular because these instruments combine the benefit of data from subjective tests (interview or open-ended test) and multiple-choice questions (Lin, 2016). Multiple-tiered instruments can be divided into two-tier, three-tier and four-tier, and the most widely used based on meta-analysis studies is two-tier multiple-choice questions. Basically, chemistry concepts are available in the form of two tiers, while additional tier requires respondents to justify their confidence rating index on Likert-scale (Soeharto et al. 2019). Analysis of multiple-tiered is based on the selected distractor, and if students chose distractors incorrectly, those answers are their misconceptions. To analyse distractors at 2TMC, we have to look at the answers at phenomenon and reasoning tier. If any one of the answers are incorrect, that it can said that the student has misconception. The illustration is shown in Figure 1.2.

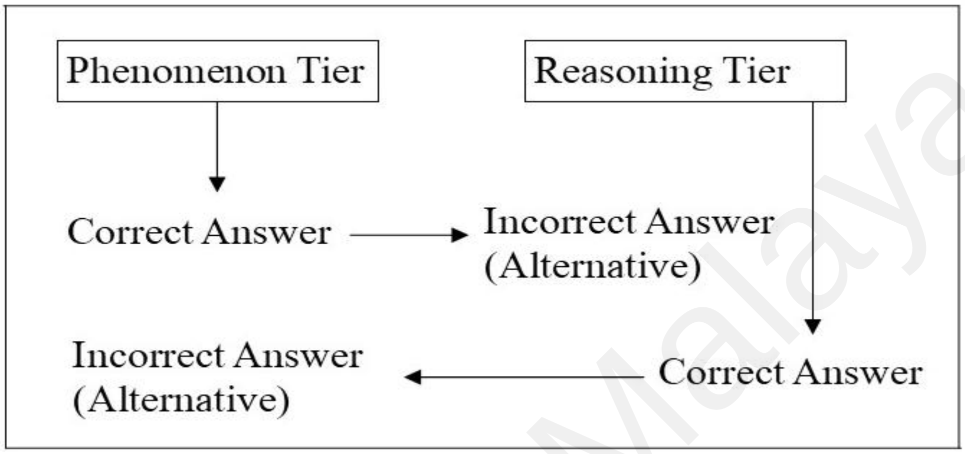


Figure 1.2 Analysis of two-tier multiple-choice questions

1.2 Statement of the Problem

Chemistry is considered as a difficult subject by students to study or by teachers to teach. The subject, by its very nature, is highly abstract, conceptual and followed by mathematical nature that requires high skills in the studying process (Ozmen, 2011; Sirhan, 2007; Wei, Liu, Wang, & Wang, 2012). Chemical reactions are considered as difficult when studying chemistry. These include university students having difficulties that are indicated by the mistake of students such as missing charge, unbalanced charge, and atoms when writing chemical equations (Naah & Sanger, 2012). Similarly, students have problems with the chemical reaction coefficient for instance 2HCl be HCl_2 (Nyachwaya et al., 2011). In the context of Indonesia, Sidauruk (2017) found the concept of the law of conservation mass, equalization of chemical reactions, and writing a correct chemical reaction as difficult components of the chemical reaction. When students move to more advanced topics like reduction-oxidation reaction, they were not able to write correct chemical equations (Kusumawati, Enawaty, & Lestari, 2014).

In learning chemical reactions, they apply three-level representations which deteriorate learning because of the complexity of the task (Sana, Adhikary, & Chattopadhyay, 2018; Santos & Arroio, 2016). Students have to translate, interpret and correlate the observable phenomena in their thoughts to the chemical concept (Orgill & Sutherland, 2008; Sana et al., 2018; Treagust, Chittleborough, & Mamiala, 2003). The problem tends to be worse when teachers do not highlight the connectedness of multiple representations in the teaching process (Chandrasegaran, Treagust, & Mocerino, 2008; Gabel, 1999).

As an instance, when hydrochloric acid and grey iron powder are combined, the observable phenomenon is the occurrence of vigorous effervescence, the disappearance of iron, and light green solution (macroscopic) (Chandrasegaran et al., 2011). When students are required to explain the reason for the new colour of the solution, the scientific reason is the forming of iron(II) chloride and the presence of Fe^{2+} from its salts (sub-microscopic) (Chandrasegaran et al., 2008). Unfortunately, some students have misconception when they give their reasons saying that atoms of iron and chlorine turn green when they combine (Chandrasegaran et al., 2011).

To communicate the phenomenon, it can be presented by using chemical reaction: $HCl_{(aq)} + Fe_{(s)} \rightarrow FeCl_{2(aq)} + H_{2(g)}$ in symbolic representation (Chandrasegaran et al., 2009). In writing chemical reactions, students need to consider the symbol of compound or element, for instance, symbol of magnesium which is different at magnesium ribbon (Mg) and burnt magnesium (Mg^{2+}). Chemical reactions differ from ionic reactions because of the deliberation of spectator ions. Therefore, students face difficulties related to the multi-faceted significance of chemical symbols, chemical formulae, either chemical or ionic compounds (Chandrasegaran et al., 2011). These problems also stated by other previous researchers (Agung & Schwartz, 2007; Chiu, 2001) who pointed out that students tend to memorize rather than understanding deeply the use of symbols and writing chemical reactions.

Misconceptions among students are considered problematic because once they acquire it is extremely persistent and difficult to change (Canpolat, 2006; Pabuçcu & Geban, 2006; Stojanovska, Soptrajanov, & Petrusevski, 2012). Moreover, misconceptions meddle with pupils‘ learning process when students utilize them to decipher scientific phenomenon; it engages strongly either emotion or intellectual of students because of the active role of pupils in the construction of their knowledge interpretation (Vosniadou, 2012); and lastly, prevent students from having meaningful learning experiences because of the shortage of ability to link between new knowledge and existing one (Cetin-Dindar & Geban, 2011; Köse, 2008).

Misconception can last for a long time to interfere with the cognitive development of students (Masson, Potvin, Riopel, & Foisy, 2014; Shtulman & Valcarcel, 2012). By comparing brain activation between novice and expert utilize functional magnetic resonance imaging (fMRI), Masson et al. (2014) found that even when misconception has been converted to the scientific view, it remained encoded in a neural network to inhibit scientific knowledge. It is also supported by a study from (Shtulman & Valcarcel, 2012) that employed MediaLab v1.21 software to record the speed and accuracy of respondent‘s thinking. It was found that the process of conceptual change is only suppressing misconceptions and it can not be supplanted.

Thus, to analyse misconceptions that could arise due to the three-level of representations in chemical reactions, a Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI) has been introduced by (Chandrasegaran et al., 2007). It was used as an experimental research intervention in Singapore (Chandrasegaran et al., 2007, 2008, 2009, 2011). As a two-tier multiple-choice test, RSCRDI has the phenomenon and reasoning tier. Phenomenon tier measures factual knowledge or chemical concepts such as the reasons of colour changes, the correct chemical and ionic reactions, the result of reactions if a reactant are replaced with new compounds among others (Chandrasegaran et al., 2011; Fulmer, Chu, Treagust, & Neumann, 2015). After providing answer for the phenomenon tier, test-takers would answer the reasoning tier that measure the scientific reasoning and chemical concepts (Taslidere, 2016).

Phenomenon tier
Dilute sulfuric acid is added to some black copper(II) oxide powder and warmed. The copper(II) oxide disappears producing a blue solution. Why is a blue solution produced?
A The copper(II) oxide dissolves in the acid producing a blue solution.
B Copper(II) oxide reacts with dilute sulfuric acid, producing a soluble salt, copper(II) sulfate.
C Copper(II) oxide is anhydrous. When the acid is added the copper(II) oxide becomes hydrated and turns blue.

Reasoning tier
The reason for my answer is:
1 The ions in copper(II) sulfate are soluble in water.
2 Cu^{2+} ions have been produced in the chemical reaction.
3 Hydrated salts contain molecules of water of crystallisation.
4 Cu^{2+} ions originally present in insoluble copper(II) oxide are now present in soluble copper(II) sulfate.

Figure 1.3 Two-tier multiple-choice questions
Reference: (Chandrasegaran et al., 2011)

However, two-tier instruments such as RSCRDI has been criticized because of the uncertainty of source of errors whether from alternative conception or lack of knowledge, and the possibility of having guessing ability (Chang et al., 2007; Gurel et al., 2015; Yang & Lin, 2015). To improve the quality of diagnostic, the three-tier multiple-choice questions (3TMC) which add CRI (Certainty Response Index) to assess the extent of student‘s belief about the degree of their correctness in responding to a certain question (Caleon & Subramaniam, 2010; Hasan, Bagayoko, & Kelley, 1999). However, if student are asked to state their CRI it is unsure if the uncertainty arises from the phenomenon or reason tier (Caleon & Subramaniam, 2010; Gurel et al., 2015; Kaltakci-gurel, Eryilmaz, & Mcdermott, 2017). To compensate for this problem, the researches in the field utilize the 4TMC by differentiating CRI for the phenomenon and reason tier respectively (Hoe & Subramaniam, 2016; Yan & Subramaniam, 2018; Yang & Lin, 2015). Even it is considered as the best version of the instrument to detect conception, its usefulness is limited to the pure diagnostics instrument rather than diagnostics practices in the classroom (Gurel et al., 2015). To visualize the difference of 2TMC, 3TMC and 4TMC, it is drawn Figure 2.1 in chapter 2 (page 28).

Even though, many studies on the field of misconception exist, there is still relatively lack attention on how to analyse misconceptions. Before discussing further, it is vital to consider two conflicting aims in data analysis. The first objective is to find the most decent model to fit data which often provides a complicated model, while another one is to keep the model as simple as possible that often gives a lower level of representation. Thus, both aims are conflicting and are not focus on one type of analysis (Judd, McClelland, & Ryan, 2009). The development of 4TMC correlates to the first aim. However, some studies try to improve the diagnostics power of original multiple choice question by using distractor to simplify the method of analysis to meet the second aim (Aretz, Borowski, & Schmeling, 2012; Briggs et al., 2006; Herrmann-Abell & DeBoer, 2011; King, Gardner, Zucker, & Jorgensen, 2004; Lin, Chu, & Meng, 2010; Wind & Gale, 2015).

The early study on using distractor to identify misconception started from Hestenes, Wells, and Swackhamer (1992) by working for the project entitled Force Concept Inventory by revising the mechanic's diagnostics instrument from Halloun and Hestenes (1985). In that study, each distractor is assigned to represent a certain form of alternative conceptions and the number of test-takers choosing certain alternative answers was regarded as proof of the ability of the instrument to distinguish the capabilities of students. In chemistry, there are also many studies using this principle to create concept inventory to reveal misconceptions (e.g. Anderson, Fisher, & Norman, 2002; Baser & Geban, 2007; Luxford & Bretz, 2014; Mulford & Robinson, 2002; Schmidt, 1997; Tyson, Treagust, & Bucat, 1999; Uzuntiryaki & Geban, 2005). From the various studies, many misconceptions have been revealed in chemistry which indicated its abstract nature, too many conceptions and followed by algorithm calculations (Chiu, 2001).

The other proponent is the use of distractors is the belief that cogent distractors should be chosen by students with alternative conception or knowledge deficiency because of the nature of its development (Rodriguez, Kettler, & Elliott, 2014). Some studies tried to elicit different conceptual levels by modifying ordinary multiple-choice questions such as ordered multiple choice question (OMC) (Aretz et al., 2012; Briggs et al., 2006) and distractor rationale taxonomy (King, Gardner, Zucker, & Jorgensen, 2004; Lin, Chu, & Meng, 2010).

Therefore, this study will analyse distractors of 2TMC by using the principle of analysis from misconception-driven distractor multiple-choice question to reveal misconception (e.g., Herrmann-abell & Deboer, 2016; Herrmann-Abell & DeBoer, 2011; Wind & Gale, 2015; Wren & Barbera, 2014) as an effort of increasing its diagnostics power (Schultz et al., 2017). The principle of analysis utilizing the Rasch model is still rarely used in diagnostics instruments on science education (Romine, Schaffer, & Barrow, 2015). Its limited use stems from science educators are fewer skills and trained to use Rasch Modeling (Liu, 2010). The principle of this analysis is similar to the analysis of ordinary multiple-choice questions, which is based on selected distractor as visualized by Figure 1.6. The main differences of traditional analysis and using IOCC located at their considerations of another tier. While traditional analysis consider another tier, IOCC analysis based only on its distractors without considering the answer of another tier.

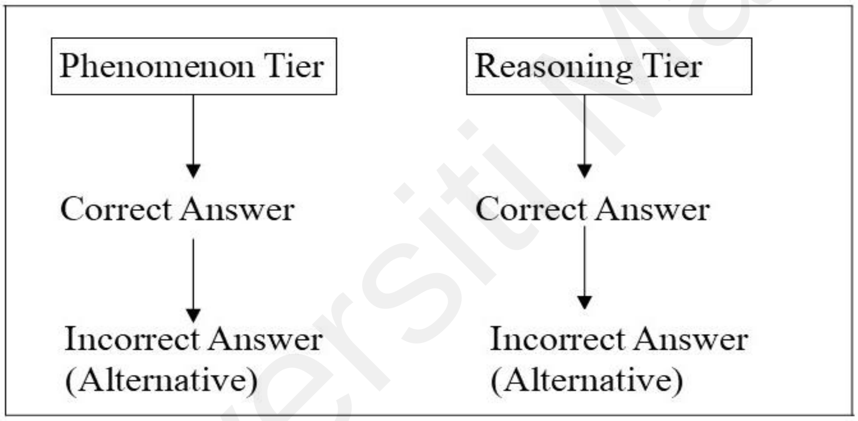


Figure 1.4 Analysis of 2TMC using IOCC

1.3 Objective of the Study

This study proposed the use of the item option characteristics curve (IOCC) as the method of distractor analysis of two-tier multiple-choice questions (2TMC) to identify misconceptions. The objectives of this study were:

1. To reveal selected pre-service chemistry teachers‘ in West Nusa Tenggara, Indonesia's misconceptions on chemical reactions by analyzing two-tier multiple-choice questions (2TMC) using traditional methods.
2. To compare the result of distractor analysis between using the traditional method and item option characteristics curve (IOCC).

1.4 Research Questions

The research questions formulated for this study were:

1. Which misconceptions do selected pre-service chemistry teachers‘ in West Nusa Tenggara, Indonesia showed on chemical reactions when analyzing two-tier multiple-choice questions (2TMC) using the traditional method.
2. Are there any different results of distractor analysis between using the traditional method and item option characteristics curve (IOCC)?

1.5 Significance of the study

A study in the area of misconception detection is always beneficial as a precursor of improvement. By conducting this study, it can give suggestions to the curriculum of pre-service chemistry teachers program to improve lecture teaching styles, laboratory activities, and assessment practice. Some prior researchers in Indonesia had applied such approach (e.g., Agustin, Supardi, & Sunarto, 2018; Farida & Liliasari, 2011; Lastri, Kusumo, & Susilaningsih, 2018; Rahmawati, 2015) to improve the ability of pre-service chemistry teachers in mastering multiple representations in chemistry. Feedbacks from diagnostics instruments could be used as tools by students their to monitor personally (performing self-assessment) their development in the process of knowledge acquisition (Hamid & Mahmood, 2010; Lee, 2007; Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2006)

From the practice of assessment, it suggests moving forward from the assessment which relies on conceptual understanding as only correct (scientific conception) and incorrect (lack of knowledge) as the common practices at schools to consider analysis of misconceptions (Dann, 2014; Flórez & Sammons, 2013). It also refers to the importance of changing teacher methods of finding misconceptions in which teachers seldom use any diagnostics instruments and seem to prefer to use informal methods such as either questions or conversations in the classroom rather than utilize multi-tiered instruments (Morrison & Lederman, 2003).

Misconception analysis is vital because if teachers understand common misconception in certain topics, lead to better learning gains compared to teachers who only can detect correct answers and incorrect answers from a certain instrument (Sadler & Sonnert, 2016). These ideas were reinforced by Ardiansah, Masykuri, and Rahardjo (2018) who analyzed the perspective of four chemistry teachers and noted that 3TMC can help them to map their students' scientific conception and misconception to improve their evaluation process and improve teaching. Not only teachers' view, but as many as 96 students as respondents in schools in West Bengkayang, Kalimantan, Indonesia also recommend the 3TMC because they want to know deeply their conceptual understanding such as truly scientific conception, misconception, lucky guess etc. Therefore, this study introduces the use of diagnostics instrument to pre-service teachers and hope that they would use them when they become teachers.

This study explained in detail to educators how to adopt an instrument to show misconceptions. As noted by referring to study of Soeharto et al. (2019), Gurel et al., (2015) and Wandersee, Mintzes, and Novak (1994), there are a plethora of studies about the diagnostics instrument (Gurel et al., 2015; Oberoi, 2017). They have convincing validity and reliability, but the effort of applying these studies into practice is still meager to enhance the quality of teaching and learning (Lee, Feldman, & Beatty, 2012; Maier, Wolf, & Randler, 2016). The major reasons for the low uptake of the instrument were the difficulty of developing such an instrument, time to adopt the multi-tiered instrument, time to register and analyze its data (Ardiansah, Masykuri, & Rahardjo, 2018; Treagust, 2006).

1.6 Definition of Terms

Diagnostics Instrument

Diagnostics instrument is defined as the tool to explore the students' mind and detect their level of understanding especially scientific conception and misconception (Gurel et al., 2015). In this study, the diagnostics test - Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI) were adapted from Chandrasegaran et al. (2007) and translated into the Indonesian language. The instrument had 15 items based on the three levels of representation for chemical reactions

Phenomenon tier

Phenomenon tier provides some chemical phenomena in the laboratory such as colour changes, the appearance of deposit, the production of heat and gases (Fulmer, Chu, Treagust, & Neumann, 2015; Lin, 2016). All of the phenomena are the consequence of mixing some chemical compounds (Chandrasegaran et al., 2009, 2011). From the phenomena, test developers give some questions and choices of responses (Xiao, Han, Koenig, Xiong, & Bao, 2018).

Reasoning tier

In 2TMC, after responding to phenomenon tier, test-takers have to respond to additional options of answer related to their reasons. The reasons were created in relation to each option in the phenomenon tier (Xiao et al., 2018). A number of choices in reasoning tier are usually more in number because every choice in phenomenon tier tends to have more than reason to be selected (Yang, Li, & Lin, 2008).

Distractors

When test-takers are required to respond to the multiple-choice questions (MCQs), they need to choose some options which can be classified into answer key and distractors. For RSCRDI, the extraction of distractors is from vigorous processes namely literature reviews, open-ended testing, multiple-choice with open justifications and semi-structured interviews (Chandrasegaran et al., 2007). All the processes is the way to make sure that the distractors are from common mistakes of students (Gurel et al., 2015; Lin, 2016). To analyze the distractors that work as alternative answers at phenomenon and reasoning tier, depending on the analysis method used. When using the traditional method, it based on the percentages, while IOCC relies on the probability of each option being selected along the curve.

Misconception

Misconceptions are considered when students view of the concept differently from the experts (Kirbulut & Geban, 2014; Köse, 2008; Maskiewicz & Lineback, 2013). In this study, misconceptions are in the topic of chemical reactions. To quantify misconception, it is decided based on the patterns of answer in which respondents had answered correctly either at phenomenon or reasoning tier (Gurel et al., 2015; Kaltacki, 2012). The data is analysed quantitatively but qualitatively the selection of distractor reveal misconceptions. It is assumed that selected distractor represents the conception of the pre-service chemistry teachers (Herrmann-abell & Deboer, 2016; Wind & Gale, 2015; Yan & Subramaniam, 2018). For the current study, alternative answer (distractors) can be divided into problems in reasoning tier (misconception type-1 (correct at phenomenon only) and problems in phenomenon tier (misconception type-2 (correct at reasoning only)) (Abraham, Grzybowski, Renner, & Marek, 1992; Salirawati, 2011).

Item option characteristics curve (IOCC)

The basic idea of IOCC analysis is to examine trace lines for alternative choices (Ding & Beichner, 2009). The distractor analysis plots were created by plotting the proportion of students selecting answer choices A, B, C, and D for phenomenon tier and 1, 2, 3 and 4 for reasoning tier (*y*-axis) across the range of student achievement measures at each time point (*x*-axis). Accordingly, the *y*-axis values indicate the relative popularity of each answer choice for students with different levels of achievement (*x*-axis). In details, after Rasch estimates of student achievement on the logit scale were obtained from the Winsteps computer program (Linacre, 2014), student achievement estimates on the logit scale were rounded to the nearest integer value (–3 to 4). Then, the frequency of students selecting each answer choice was obtained for each value. At each point on the scale, the proportion of students selecting each answer choice was calculated by dividing the frequency of students who selected a given answer choice by the total number of students observed at each point on the scale (Wind & Gale, 2015). From this figure, we can analyze information towards student conceptual understanding on chemical reactions.

Pre-service chemistry teachers

In this study, pre-service chemistry teachers refer to university students who major in chemistry education. In Indonesia, they have to complete for more than 140 credits for broad of materials such as chemistry, pedagogy, thesis, etc. (Erman, 2016).

1.7 Limitation of the Study

The limitation of this study is due to the sample size, the findings cannot be generalized. In the context of Indonesia, this study only takes pre-service chemistry teachers from one province in the middle part of Indonesia namely West Nusa Tenggara, while there are 34 provinces with more population. For a quantitative study which aims to generalize its result, having more samples in a larger population will increase the value of this study (Etikan, Musa, & Alkassim, 2016). Therefore, the issue stated by Nehm et al., (2010) about the issue of the lack of studies to generalize its result in instrument field cannot be addressed properly in this study.

Limitations from instrumentation and method of data collections is that this study does not proportionally represent factors influencing test instruments in education such as the origin of students, socio-economical background, gender, religion, intelligence quotient (IQ) etc. (McMillan & Schumacher, 2010). From method of data collection, this study is voluntary basis. As a result, it allows respondents to answer questions without putting extra efforts because the result does not influence their grade. Therefore, misconception data possibly does not closely represent the true degree of pre-service chemistry teacher‘s ability.

1.8 Scope of the Study

In studying chemistry in Indonesia, according to the document of chemistry syllabus for senior high school, there are many topics to cover which includes chemical reaction, but this study only takes some reactions: 1) metal combustion, acid-base reaction, precipitation reaction and, metal-ion displacement reactions owing to the availability of instruments, resources, and time constraint. According to Kozma and Russell (2005), there are many representations in learning chemistry but the study only considers three level of representations comprising macroscopic, submicroscopic and symbolic representation.

1.9 Chapter Summary

This chapter has described the explication of some vital prerequisite information concerning the study. The foundation of conducting this study includes the background of the study, the statement of the problem, the objectives and research questions, the rationale and significance of this study, the limitation as well as the definition of terms. In the following chapter, more detailed information from extensive literature reviews to shape the understanding of this study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The gap between what the teacher had taught and how differently the students perceived the concept can be understand by the use of diagnostics instruments such as concept inventory or multiple-tiered multiple choice (William, 2013). In the current study, a diagnostics instrument was to elicit misconceptions. This chapter begins with the explanation of misconception and instruments to detect it. Since this study applies test theory as its foundation, a brief explanation of two fundamental theories in measurement namely classical test theory (CTT) and item response theory (IRT) is provided to enhance the understanding of this study. Type of IRT used is Rasch model and this chapter emphasizes the related studies on applying Rasch model. This chapter ends with the conceptual and theoretical framework.

2.2 Misconception

Various distinct terms of misconception include alternative conceptions, naïve beliefs, children‘s ideas, conceptual difficulties and phenomenological primitives (Gurel et al., 2015; Mutlu & Sesen, 2016). Misconception is the form of understanding a concept which has some error or differences from what scientists view the concept (Arslan et al., 2012; Köse, 2008; Maskiewicz & Lineback, 2013).

In chemistry, various studies documented preservice teachers‘ misconception on topics such as chemical equilibrium (Azizoğlu, Alkan, & Geban, 2006; Bilgin, 2006; Mutlu & Sesen, 2016), electrochemistry and colligative properties (Karsli & Çalik, 2012; Mutlu & Sesen, 2016; Yalcin, 2012), environmental chemistry: greenhouse, acid rain, global warming (Arslan et al., 2012; Çelikler & Aksan, 2014; Sesli & Kara, 2012), chemical kinetics (Mutlu & Sesen, 2016; Tastan, Yalcinkaya, &

Boz, 2010), chemical and physical changes (Bak Kibar, Yaman, & Ayas, 2013; Çalik, Ayas, & Coll, 2007), chemical laws and stoichiometry (Bak Kibar et al., 2013; Haidar, 1997; Kalin & Arikil, 2010), thermochemistry (Gültepe, 2016; Mutlu & Sesen, 2016), periodic table and acid-base (Mutlu & Sesen, 2016; Şenol, 2018). In the local context, Indonesia, misconception of preservice teachers on chemistry concepts such as stoichiometry (Desi, 2013), ionic bond (Muchson & Su, 2015; Yasthophi & Ritonga, 2017), microscopic representation in a chemical reaction (Winarni & Syahrial, 2011) have been documented.

There are some characteristics of the misconception that influence further learning including its persistence and difficulties to change (Canpolat, 2006; Pabuçcu & Geban, 2006; Stojanovska et al., 2012). The robust persistence of misconception is determined by its consistency against age, gender, culture, nationality, and proficiency (Wang, 2004). It is also well-embedded in the cognitive ecology, moreover prepared instructional design to extinguish it tend to fail (Sungur, Tekkaya, & Geban, 2001).

Misconception in a learning process can be diminished through conceptual change which is the concern of science education researchers for more than a fourth century (Arslan et al., 2012; Duit & Treagust, 2003; Treagust, 2006). Understanding this term is advantageous for modeling instructional design because it can help the teacher to help student‘s moving forward from a certain level of understanding (Hasan et al., 1999). What teacher can do by incorporating conceptual understanding is to select which materials relevant to student‘s level and which sub-materials should be emphasized in teaching and learning process (Furtak, Morrison, & Kroog, 2014).

Conceptual change can take place under some prerequisite conditions such as 1) the dissatisfaction of the learner to their current conception 2) the new conception has to be intelligible and plausible 3) the new concept offers the hope of a fruitful program (Posner, Strike, Hewson, & Gertzog, 1982). The view is based on students' epistemologies, where the dissatisfaction incurs dramatic or revolutionary conceptual change and was embedded in radical constructivist epistemological views with an emphasis on the individual's conceptions and his/her conceptual development. If the learner was dissatisfied with his/her prior conception *and* an available replacement conception was intelligible, plausible and/or fruitful, accommodation of the new conception may follow. An intelligible conception is sensible if it is non-contradictory and the student understands its meaning; plausible means that in addition to the student knowing what the concept means, he/she finds the conception believable; and, the conception is fruitful if it helps the learner solve other problems or suggests new research directions.

It is essential to note that conceptual change does not only affect cognitive factor, but also personal factors such as motivation, epistemic beliefs, and learning strategies (Leonard, Kalinowski, & Andrews, 2014). An aspect of motivation that strongly influences conceptual change comprising the enthusiasm to learn something for its own sake, achieve an excellent grade, prepare for a specific future career, circumvent a poor academic performance and appearing less competent than competitors (Elliot & McGregor, 2001; Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011). Epistemic belief explores the student's perspective toward knowledge and learning, for instance, a student perceives knowledge as the body of discrete facts tends to memorize information (Schommer, 1990). Concerning this, Debacker, Crowson, Beesley, Thoma, and Hestevold (2008) highlights four beliefs:

1) the speed of learning take place 2) the complexity of materials to master 3) the certainty of the known information 4) the nature of abilities i.e., innate or result of learning. For learning strategy, it relates to cognitive (regulating one's cognition through by incorporating thinking skills) and metacognitive (monitoring one's comprehension of new material) strategy of students (Leonard et al., 2014). As documented by some researchers (e.g. Sackes 2013; Vilppu Mikkilä-erdmann, & Ahopelto 2011), some learning styles affect conceptual changes namely rehearsal, elaboration, organization, self-regulation, critical thinking and cognitive reflection (Frederick, 2005; Pintrich, Marx, & Boyle, 1993).

2.2.1 Sources of Misconception

There is no overt evidence where alternative conception comes from in science including chemistry because research findings are still speculative and difficult to document (Taylor & Kowalski, 2004; Treagust & Duit, 2009). The first source is from daily life experiences, which explicate the proximity of this subject to the pupil's neighborhood. Every day students experience new things in all aspects of their lives, building conception can be through practical activities, talking with other people around them, and through media they access (Eilks et al., 2012; Talanquer, 2006). Since misconception takes years of observations, trial and error, and consistent practices, the process of accepting a new concept are influenced by the existing concept (Taylor & Kowalski, 2004).

Misconception can arise when teachers use instructional language to explain concepts when teachers use analogies, paraphrasing information, provide inappropriate examples, and the use of everyday languages trigger student acquisition knowledge differently (Doige & Day, 2012; Gurcay & Gulbas, 2015; Jaffar & Dindyal, 2011). It also can be influenced by language use to teach the concept especially if students have different first language (mother tongue) and have low proficiency in instructional language (Mutlu & Sesen, 2016).

Textbooks are also the source of alternative conception since they shape method and strategies of teaching, and act as the guidelines of learning (Kajander & Lovric, 2009). The failure of the textbook to provide clear explanation causes the appearance of misconceptions (Hrast & Savec, 2017). It is also found the definitions of concepts given in textbooks either lack precision or invoke ideas that beginners are not familiar with (Nelson, 2003). A study found the relationships between textbook and alternative conception of both chemical and phase change, dissolution, conservation of atoms, periodicity (Abraham et al., 1992). The other study found the simplification of languages and misleading interpretation in textbooks of electrochemistry is the possible source of misconception (Sanger & Greenbowe, 1992).

One additional reason of students having misconceptions is their lack of ability to apply chemistry “triplets” consisting of macroscopic, submicroscopic and symbolic (Bucat, 2004; Chandrasegaran, Treagust, & Mocerino, 2007; Meijer, 2011; Stojanovska et al., 2012). There is no superiority among all of the representations in learning chemistry because they complement each other to master chemistry (Stojanovska et al., 2012; Talanquer, 2011). These three level of representations in chemistry should be connected, related and applied simultaneously and students find that when they have to do so it chemistry becomes complicated (Kozma & Russell, 1997). As a result, students tend to have a false understanding concerning chemistry concepts.

2.3 Diagnostics Test

Diagnostics test refers to the method of finding students' conceptual understanding in the form of misconceptions. Basically, the method can be classified into subjective methods (interviews and open-ended test) and objective methods (ordinary multiple choice and its modification). In misconception detection, the reported misconception qualitatively is based on distractor analysis. From the distractors, it can be reported the conception of students and followed by quantitative number of students having certain misconception.

2.3.1 Concept Inventory (CI)

In the field of physics and biology education, Gurel et al., (2015) found five instruments for Biology and 13 instruments for Physics to check for students' conceptual understanding. The materials are force, motion, light, energy-momentum, graphs in kinematics, general chemistry, general biology, solution, heat and temperature, meiosis, molecular biology, and life sciences. In chemistry, some studies on concept inventories include Chemistry Concept Inventory (CCI) (Mulford & Robinson, 2002), Solution Concept Test (Uzuntiryaki & Geban, 2005), Conceptual Inventory of Natural Selection (CINS) (Anderson et al., 2002), Heat and Temperature Concepts Test (Baser & Geban, 2007), isomerism, redox and acid-base (Schmidt, 1997), equilibrium (Tyson et al., 1999), Bonding Representation Inventory (Luxford & Bretz, 2014), Thermodynamics Concept Inventory (Wren & Barbera, 2014), particulate nature of matter (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013), structure and motion of matter (Stains, Escriu-Sune, Santizo, & Sevian, 2011), Kinetics Particle Inventory (Treagust et al., 2010), covalent bonding and structure (Raymond F Peterson & Treagust, 1989), covalent bonding and molecular structure (Pentecost & Langdon, 2008), particulate nature of matter and bonding (Othman, Treagust, & Chandrasegaran, 2008), particulate nature of matter (Yezierski & Birk, 2006). In using concept inventory, misconceptions are obtained qualitatively is based on distractors selection and by the percentage of students having that particular misconception.

The main critics of TMC as evaluation and diagnostics instrument is guessing, referring to the condition in which students do not have any ideas of correct answers and select an option without having reasons or justification (Rodriguez et al., 2014). Generally, there are two forms of guessing: “blind guessing” and “informed guessing”. Blind guessing occurs when the respondents have no idea of the correct answer and respond randomly, so it is sometimes called random guessing. The answer of students in multiple choice questions with this issue has not represented any ability or skills (Andrich, Marais, & Humphry, 2012). While informed guessing occurs when an examinee has partial knowledge and the ability to remove some choices and then randomly guess remaining choices (Downing, 1992). The probability of responding correctly with guessing is 20%, 25%, 33.33%, 50% for 5, 4, 3, 2 choices respectively. The more number of distractor, the lower probability of students practice guessing. Guessing is a problem because it can interfere with the student's scores on a test which adversely influences the validity and reliability (Şenel, Pehlivan, & Alatlı, 2015).

2.3.2 Two-Tier Multiple-Choice Question (2TMC)

2TMC is the extension of TMC by adding more sources of identifying student misconceptions (Taslidere, 2016). 2TMC has two parts namely phenomenon tier and reasoning tier.

2.3.2.1 Phenomenon tier

The first-tier is usually set-up the same way as noted with the conceptual inventory (CI), but the second-tier involves students selecting a reason as to why they selected the answer in the first-tier (Adadan & Savasci, 2012; Schaffer, 2012). Phenomenon tier measures factual knowledge or core concepts in a tested domain (Taber & Tan, 2011). For instance, dilute sulfuric acid is added to some green copper (II) carbonate powder. Vigorous effervescence occurs and the copper (II) carbonate disappears producing a blue solution. From the phenomenon, some questions are raised such as the reasons of the changes, chemical reaction, ionic reaction, how if reactants are changed etc (Chandrasegaran et al., 2007, 2011).

2.3.2.2 Reasoning tier

The reasoning tier is the justification of responses at phenomenon tier (Taslidere, 2016). In this tier, conceptual knowledge is asked in responses to phenomenon in the first tier. The rationale of answering phenomenon tier goes beyond knowing (Taber and Tan 2011). Adding reasons can provide deep information about student‘s conceptual understanding as the way of assessing learning experience (Fulmer et al., 2015). The reasoning tier could be developed as open ended or multiple choice questions. Multiple choice questions are used in the reasoning tier because students may provide short reasons or insufficient information which tend to be useless and time-consuming if asked as open-ended questions (Chu, Treagust, & Lim, Chandrasegaran, 2015).

In order to be able to display understanding when engaged in reasoning about chemical reactions and other chemical phenomena, students should be able to constantly navigate between the levels of representation, utilizing each representation at the appropriate stage of their reasoning. The acquisition of knowledge by students without clear understanding may be attributed to the confusion caused in having to deal simultaneously with the macroscopic, submicroscopic and symbolic levels of representation in chemistry. From observations of changes that occur at the macroscopic level, students have to explain these changes at the molecular (particulate) level. The molecular level in turn is represented by symbols and formulas. As a result of having to deal with three levels of representation simultaneously, learners generally experience difficulty in explaining chemical reactions (Gabel, 1998).

2.3.3 2TMC in chemistry

In the field of chemistry education, there are many studies about the development and the use of 2TMC as listed below in Table 2.2. The extensive use of this assessment format is due to the amalgamation of advantages from either subjective test (interviews, essays, open-ended test) or objective test (multiple choice question) (Lin, 2016; Tsui & Treagust, 2010). Its vivid strengths comprise efficient time for administrating and grading, the number of needed workers, the ability of generalization and the breadth of covered topics and subtopics (Adadan & Savasci, 2012; Rollnick & Mahoona, 1999; Saat et al., 2016).

Table 2.1

The study on the 2TMC instrument in chemistry

Materials	References
Chemical equilibrium	(Akkus, Kadayifci, & Atasoy, 2011; Voska & Heikkien, 2000)
Qualitative Analysis inorganic chemistry	(Tan, Goh, Chia, & Treagust, 2002)
Covalent Bonding and Structure	(Peterson, Treagust, & Garnett, 1986; Treagust, 1986)
Nature of Solutions and Solubility	(Adadan & Savasci, 2012)
Boiling Concept	(Coştu, Ayas, Niaz, Çalık, 2007)
Separation of Matter	(Tüysüz, 2009)
Chemical Concept	(Chiu, 2007)
Acid-Base	(Artdej, Ratanaroutai, Coll, & Thongpanchang, 2010)
Representational Systems and Chemical Reactions Diagnostic	(Chandrasegaran et al., 2007; Chandrasegaran et al., 2008b, 2011; Treagust & Duit, 2009)
Electrolyte	(Lu & Bi, 2016)

2.3.3.1 Strengths of 2TMC

There are some strengths of 2TMC: 1) It is more convenient for the teacher to administer rather than an interview because of the limitation of time and resources (Adadan & Savasci, 2012). Even 2TMC can work better in analyzing conceptual understanding compared to TMC, there are still some critics of this instrument. Marek, Maier, and McCann (2008) noted that with two-tier assessments some students have other reasons than the ones found in the second tier for their selection in the first tier, and would rather have the chance to write out another reason for their selection. 2TMC also gives clue to reason tier answer since it is related to the first tier which cannot be found in an interview and open-ended test (Tamir, 1989). In addition, referring to the work of Caleon and Subramaniam (2010) and Hasan et al., (1999), the other problem of 2TMC is its inability to differentiate mistakes because of lack of knowledge or alternative conception; or to differentiate correct answer because of the true understanding or guessing. The probability of having to guess in 2TMC is around 6% if there are 4 choices, lessening from 25% in ordinary MCQs (Milenković, Hrin, Segedinac, & Horvat, 2016a).

2.3.3.2 2TMC and CRI

There are various types of diagnostics instruments introduced to help detect students' conceptual understanding. These are 1) two-tier (2TMC), traditional multiple choice with confidence rating index (TMC-CRI), three-tier (3TMC) and four-tier (4TMC) (Gurel et al., 2015; Xiao et al., 2018). Their similarities are the ability to detect scientific conception, misconception and lack of knowledge but the differences in the sources of information in which 4TMC (phenomenon tier, reasoning tier, 2 CRI), 3 tier (phenomenon tier, reasoning tier, CRI), TMC-CRI (phenomenon and CRI), and 2TMC (phenomenon tier and reasoning tier) (Cetin-Dindar & Geban, 2011; Iqbal, 2016; Kılıç & Sağlam, 2009; Yan & Subramaniam, 2018). Their difference are the introduction of CRI which has been used in psychology, test of intelligence, and science education (Hasan et al., 1999; Taslidere, 2016). Some studies also found a significant correlation of score accuracy and confidence to lie between .40 and .60 (Kleitman & Stankov, 2007; Koriat, A., Lichtenstein, & Fischhof, 1980; Shaughnessy, 1979).

To have a succinct visualization of those mentioned instruments, Figure 2.1 shows the differences between these 4 instruments which is indicated by the different form of an arrow. Phenomenon tier is a set of questions about the chemical reaction which contains a phenomenon of reaction and some responses. Reasoning tier contains some reasons for selecting an option in a phenomenon tier. For CRI, it is to state the confidence level of choosing an answer in phenomenon or reasoning tier.

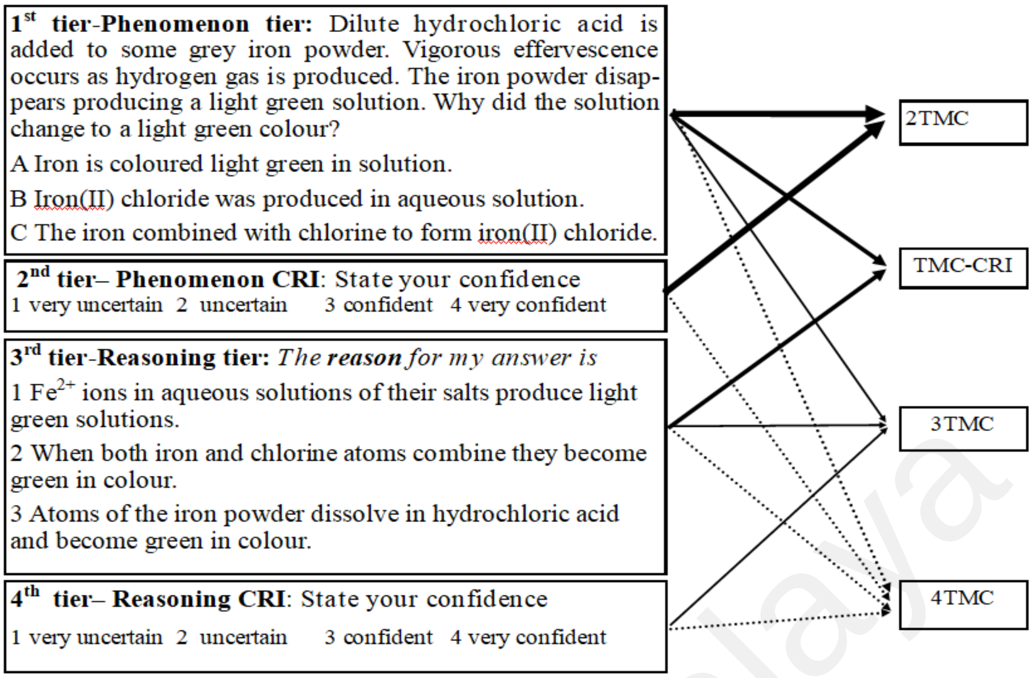


Figure 2.1 The appearance of TMC and multiple-tiered MCQ

Source : (Gurel et al., 2015; Hasan et al., 1999; Iqbal, 2016)

Zakay and Glicksohn (1992) the relationship between confidence and achievement was conducted by which tested fifty-two second-year psychology students'relationships between confidence levels and achievement at Tel Aviv University and reported from t-test result that students with higher confidence tend to achieve lower score as compared to less confident students. The rationale of the finding was the utilization of higher confidence as the defense mechanism of their unpreparedness. From the discussion of all forms of instruments, the concise overview of strengths and weaknesses are shown in Table 2.2.

Table 2.2

Comparison of TMC (concept inventory), 2TMC, 3TMC and 4TMC

Form of Instrument	Strength	Weakness
TMC	Efficient time in administration Easy grading and reporting Objective scoring	Do not provide deep enough investigation into the student's ideas. There is no reason for students responses, therefore, overestimation of students answer can easily take place Potentially interpreting students responses incorrectly depend on how well the items are constructed Guessing if students do not have any ideas or partial knowledge
2TMC	Having all the advantages of TMC	Overestimates the proportions of the misconceptions Even it is smaller than TMC, there is still a chance to guess the correct answer
3TMC	Having all the advantages of TMC and 2TMC Differing the nature of answer at the first and second tier due to misconception or lack of knowledge	Overestimates students scores owing to the difficulties to measure CRI if there is a difference between CRI of the first and second tier
4TMC	Having all the advantages of TMC, 2TMC, and 3TMC Providing the truly assessed misconception which is free of error and distinguishable from other conceptual understanding	Requires longer administrating and grading time. Advantages are limited to diagnostic purpose, and difficult to apply

(Source: Gurel et al., 2015; Kaltacki, 2012)

2.4 Test Theory

2.4.1 Classical Test Theory (CTT)

CTT contains a set of concepts and techniques in the construction and analysis of various instruments. It also provides an underlying concept to develop other approaches of the instrument. The ubiquitous use of this theory is mainly because of its popularity and easy to use (Fan, 1998). CTT is also well-known as True Score Theory which states that each respondent has a true score (T) in a test if an error in measurement is eliminated. The definition of a person's true score is the expected score of a test taker in any numbers of registration. However, in a measurement, it is a succinct possibility to obtain a true score because what to find is an observed score which is equal to the true score (T) plus some error. It is concluded that true score can be possibly higher or lower compared to the observed score (Cappelleri, Lundy, & Hays, 2014; DeVellis, 2006).

CTT is the main theory to analyze the psychometric properties of the diagnostics instrument (Gurel et al., 2015). It is vital to note that there is a difference of correlation point biserial between TMC and 3TMC/4TMC. In TMC, it was applied to check item by item, while in 3TMC/4TMC it is used to check the correlation between phenomenon and reasoning tier toward confidence rating index (CRI).

Table 2.3

The analysis of psychometric properties of diagnostics test

References	Psychometric Properties
Concept Inventory (Chemistry)	
(Mulford & Robinson, 2002)	3
(Uzuntiryaki & Geban, 2005)	3
(Anderson et al., 2002)	1, 8
(Baser & Geban, 2007)	3
(Taskin et al., 2015)	1, 2, Rasch model
(Sadhu et al., 2017)	2, 11
(Milenković, Hrin, Segedinac, & Horvat, 2016b)	1, 2, 3
(Enawaty & Putra Sartika, 2015)	Not Given
(Önder, 2018)	3
(Ikenna, 2015)	3
(Linenberger & Bretz, 2015)	1, 2, IOCC
(Vrabec & Prokša, 2016)	3
Concept Inventory (Physics)	
(Hestenes, Wells, & Swackhamer, 1992)	Not Given
(Berek, Sutopo, & Munzil, 2016)	2, 3
(Milner-Bolotin, 2015)	Not Given
(Soeharto, 2016)	Not Given
(Kusairi, Alfad, & Zulaikah, 2017)	Not Given
(Eshach, Lin, & Tsai, 2018)	5
(Thornton & Sokoloff, 1998)	Not Given
(David Hestenes & Wells, 1992)	Not Given
(Eshach, 2014)	1, 5
(Handhika, Cari, Suparmi, Sunarno, & Purwandari, 2018)	3
(Haryono, 2018)	Not Given
(Wartono, Batlolona, & Putirulan, 2018)	Not Given
(Singh & Rosengrant, 2003)	1, 2, 8, 10
(Beichner, 1994)	1, 5, 8
(Engelhardt & Beichner, 2004)	1, 2, 5, 8, 10
(Wijayanti, Raharjo, Saputro, & Mulyani, 2018)	Not Given
(Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001)	1, 2, 5
(Ding, Chabay, Sherwood, & Beichner, 2006)	1, 2, 8, 10
(Bardar, Prather, Brecher, & Slater, 2006)	1, 2, 3
(Cataloglu & Robinett, 2002)	Not Given
(Tongchai, Sharma, Johnston, Arayathanitkul, & Soankwan, 2009)	1, 2, 5, 8
(Madu & Orji, 2015)	Not Given
(Asri, Rusdiana, & Feranie, 2017)	Not Given
(Ergin, 2016)	3
(Nwafor, 2015)	5
(Kartiko, 2018)	Not Given
(Sadler & Sonnert, 2016)	Not Given
(Wind & Gale, 2015)	Rasch model
(Çetin, Kaya, & Geban, 2009)	3
(Malik, Angstmann, & Wilson, 2019)	Not Given

Table 2.3, (continue)

References	**Psychometric Properties**
(Samsudin, Liliawati, Sutrisno, Suhendi, & Kaniawati, 2014)	Not Given
Concept Inventory (Biology)	
(Klymkowsky & Garvin-Doxas, 2008)	Not Given
(Knight & Wood, 2005)	Not Given
(Desstya, Prasetyo, Suyanta, Susila, & Irwanto, 2019)	
(Shi et al., 2010)	1, 2, 3
(Howitt, Anderson, Hamilton, & Wright, 2008)	Not Given
(Kalas, O'Neill, Pollock, & Birol, 2013)	1, 2, 8
(Butler, Mooney Simmie, & O'Grady, 2015)	Not Given
(Helmi, Rustaman, Tapilow, & Hidayat, 2019)	Not Given
(Murti, Aminah, & Harjana, 2019)	Not Given
(Putri, Rahman, & Priyandoko, 2016)	Not Given
(Orbanić, Dimec, & Cencič, 2016)	3
(Subayani, 2016)	Not Given
2TMC Chemistry	
(Chandrasegaran, Treagust, & Mocerino, 2007)	1, 2, 3
(Chandrasegaran et al., 2008, 2009, 2011)	1, 2, 3
(Siswaningsih, Firman, Zackiyah, & Khoirunnisa, 2016)	2, 3
(Potvin, Skelling-desmeules, & Sy, 2015)	3, 10
(Artdej, Ratanaroutai, Coll, & Thongpanchang, 2010)	1, 2, 3
(Mutlu & Sesen, 2016)	3
(Mutlu & Sesen, 2015)	1, 2, 3
(Chiu, 2007)	Not Given
(Tüysüz, 2009)	3
(Akkus, Kadayifci, & Atasoy, 2011)	1, 2, 3
(Tan, Goh, Chia, & Treagust, 2002)	2, 3, 4
(H. R. Widarti, Permanasari, & Mulyani, 2016)	Not Given
(Hayuni Retno Widarti, Permanasari, & Mulyani, 2017)	Not Given
(Coştu, Ayas, Niaz, Ünal, & Çalik, 2007)	3
(Voska & Heikkie, 2000)	Not Given
(Peterson, Treagust, & Garnett, 1986)	Not Given
(Treagust, 1986)	Not Given
2TMC Physics	
(Chu, Treagust, & Chandrasegaran, 2009)	3
(Fetherstonhaugh & Treagust, 1992)	Not Given
(Chang et al., 2007)	Not Given
(Chen, Lin, & Lin, 2003)	1, 2, 3
(Yumusak, Maras, & Sahin, 2015)	2, 3, 5
(Kanli, 2015)	5
(Saifullah, Sutopo, & Wisodo, 2017)	Not Given
2TMC Biology	
(Mann & Treagust,1998)	Not Given
(Stevens et al., 2017)	Not Given
(Alharbi et al., 2015)	3
(Vitharana, 2015)	Not Given
(Monteiro, Nóbrega, Abrantes, & Gomes, 2012)	3

Table 2.3, (continue)

References	**Psychometric Properties**
(Odom & Barrow, 1995)	1, 2, 6
(Sesli & Kara, 2012)	1, 2, 3
(Kılıç & Sağlam, 2009)	1, 2, 5
(Griffard & Wandersee, 2001)	Not Given
(Wang, 2004)	1, 2, 3, 9
(Lin, 2004)	1, 2, 7
(Cheong, Treagust, Kyeleve, & Oh, 2010)	1, 2, 3, Rasch
(Tsui & Treagust, 2010)	3
(Haslam & Treagust, 1987)	1, 2, 3
3TMC	
(Korur, 2015)	1, 2, 3
(Suliyanah, Putri, & Rohmawati, 2018)	2, 3
(Ainiyah, Ibrahim, & Hidayat, 2018)	
(Taufiq, Sriyati, & Priyandonko, 2018)	Not Given
(Cahya & Sanjaya, 2015)	
(Osman, 2017)	3
(Aydeniz, Bilican, & Kirbulut, 2017)	Not Given
(Irsyad & Linuwih, 2018)	Not Given
(Prastiwi, Kholiq, & Setyarsih, 2018)	Not Given
(Wijaya, Supriyono Koes, & Muhardjito, 2016)	Not Given
(Lin, 2016)	3
(Ardiansah et al., 2018)	3
(Oberoi, 2017)	Not Given
(Uygar Kanli, 2014)	5, 9
(Liampa, Malandrakis, Papadopoulou, & Pnevmatikos, 2017)	2, 4, 9
(Gurcay & Gulbas, 2015)	1, 3, 4
(Taslidere, 2016)	2, 3, 8, 10
(Saat *et al.*, 2016)	1, 2, 3
(Arslan, Cigdemoglu and Moseley, 2012)	1, 2, 3, 8
(Cetin-Dindar and Geban, 2011)	1, 2, 3, 9
(Peşman & Eryilmaz, 2010)	1, 2, 3, 8
(Kirbulut & Geban, 2014)	1, 2, 3, 9

Table 2.3, (continue)

References	**Psychometric Properties**
(Kaltakçi & Didiç, 2007)	1, 2, 3
(Milenković et al., 2016)	1, 2, 3
(Eryilmaz, 2010)	1, 2, 3, 10
(Cheung & Yang, 2018)	Not Given
(Sen & Yilmaz, 2017)	Not Given
(Prodjosantoso, Hertina, & Irwanto, 2019)	Not Given
4TMC	
(Caleon & Subramaniam, 2010)	1, 2, 3
(Kaltakci-gurel et al., 2017)	1, 2, 3, 9, 10
(Anggrayni & Ermawati, 2019)	3
(Maier et al., 2016)	3

Table 2.3, (continue)

References	Psychometric Properties
(Ammase, Siahaan, & Fitriani, 2019)	Not Given
(Fariyani, Rusilowati, & Sugianto, 2017)	Not Given
(Afif, Nugraha, & Samsudin, 2017)	Not Given
(Hermita et al., 2017)	Not Given
(Sari, Sopandi, Koesbandiyah, & Arviana, 2018)	Not Given
(Hoe & Subramaniam, 2016)	1, 2, 3, 9
(Yan & Subramaniam, 2018)	1, 2, 3
(Sreenivasulu & Subramaniam, 2014)	1, 2, 3, 6
(Sreenivasulu & Subramaniam, 2013)	1, 2, 3, 6
(Yang & Lin, 2015) adopted 2TMC from (Yang, Li, & Lin, 2008)	Not Given

Note: 1) Discrimination Index, 2) Difficulty Index 3) Cronbach's Alpha (Reliability) 4) Facility Index 5) kr-20 6) split half 7) test-retest 8) Correlation point biserial 9) Pearson correlation 10) Factor Analysis (EFA) 11) Product moment correlation

2.4.2 Rasch Model

Another theory as framework of data analysis is Rasch model. The basic principle of this theory are two statements comprising: (a) a test-taker having better result than other test-takers should have higher probability of correctly responding to any item of the questions in a test and (b) one item in a test being more demanding than another means that the probability of correctly respond to that item for any test-taker is smaller compared to the second item (Bond & Fox, 2015). Data of learning outcomes cannot be treated as interval data because the scoring method by calculating and adding a number of correct answers can only assume data as ordinal data. Either interval or ordinal data can rank students, but the interval of ordinal data is not equal among data. Therefore the transformation of data is needed to meet the nature of running statistical analysis for comparative study such as t-test, ANOVA and ANCOVA (Saidfudin et al., 2010). To do so, one way is to transform data by employing the Rasch model. This model can work to address measurement problems by telling the condition when someone responds an item, defining excuses of the responses, directing how to estimate the responses and determining the relation of responses to the estimated situation (Wright, 1977).

In analyzing students learning the outcome, the Rasch model can give a better representation and explanation even in a small number of students. This offers the high precision of comparison and the true degree of the level of achievement (Osman, Badaruzzaman, & Hamid, 2011). From the Rasch Model, one of interesting feature is also the ability to visualize data using wright map (item-person map) which is a graphical and empirical representation of a progress variable (Boone, Staver, & Yale, 2014; Wilson, 2008).

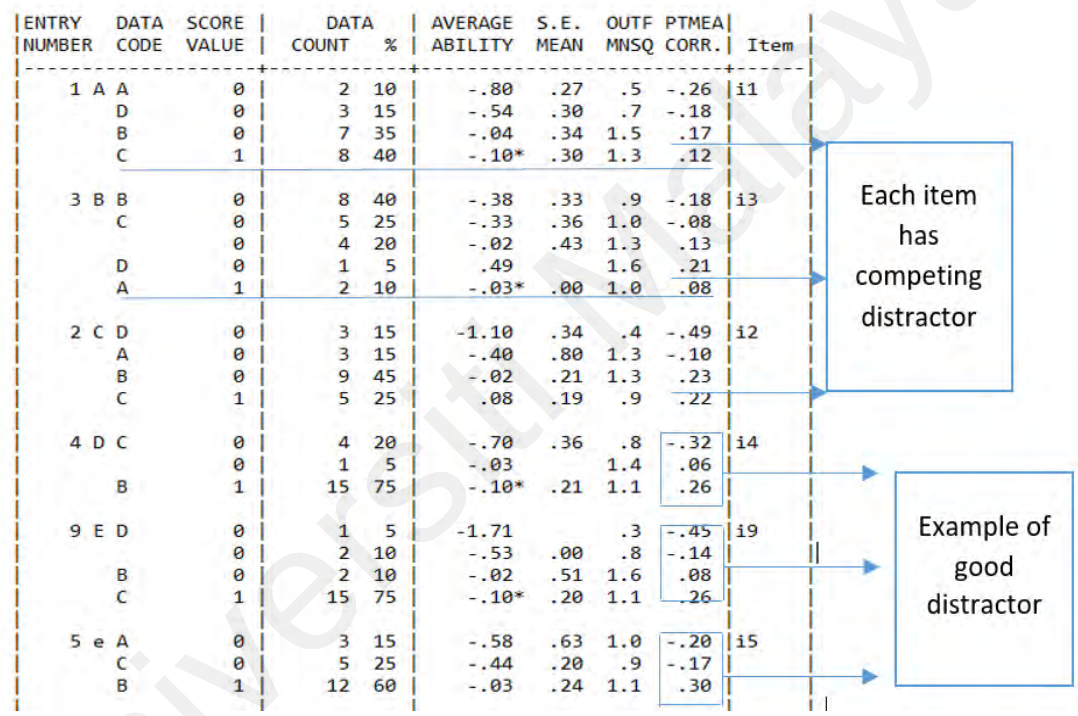
To estimate respondent measure by considering a person‘s ability and item difficulties, Rasch model calls the term as logit. Logit = Log (P/(N-P), where P= number of correct item from given items , N= number of given items. Logit is classified into person logit and item logit. Person Logit : Ψ [p] = ln (p/(1-p), item logit : Ψ [p-value] = ln (p-value/(1-p-value), where Ψ symbolize logit transformation.

In nature, logit score delineates natural log odds of each person to succeed in an item for the determination of the zero point scale (Ludlow & Haley, 1995). Item difficulty is the attribute that affects the person‘s response while the person‘s ability shapes the item difficulty estimates (Abdullah, Noranee, & Khamis, 2017). The proponent of Rasch model measurement are two theorems: 1) A person who is more capable has a higher probability of correctly responding to all the items provided. 2). An easier item is more likely to be answered correctly by all respondents or test-takers (Linacre, 1999; Sumintono & Widhiarso, 2015).

One vital information for the study is distractor analysis. The quality of distractor can also be seen by referring to item option & distractor frequencies in misfit order from Winstep. The vital data to reveal is the point correlation between the data code (scored 1, or non-occurrence, scored 0, of this category or distractor and the person raw scores or measure). The criteria of good distractor is it has a lower correlation compared to correct answer. It must be selected by minimum 10% of respondents as the criteria of misconception distractors (Yan & Subramaniam, 2018).

Generally, there are some characteristics of a good distractor. They are (a) the distractors should be theoretically plausible; (b) are common errors or misconceptions from literatures, or other empirical data; (c) avoid technically phrased distractors; (d) use of familiar yet incorrect phrases; (e) true statements that do not correctly answer the stem; (f) avoid the use of humour; (g) develop as many effective options as possible; (h) place distractors in logical or numerical order; (i) distractors are independent and may not overlap; (j) distractors are kept homogeneous (content and structure); (k) the length of distractors about equal; (l) ―none of the above” and ―all of the above” should be used carefully; (m) avoid giving clues to the correct answer and (n) phrase distractors positively, avoid negatives (Gierl, Bulut, Guo, & Zhang, 2017; Haladyna, 2016; Haladyna & Rodriguez, 2013; Haladyna, Downing, & Rodriguez, 2002).

Distractor‘s position also influences the strength to distract test-takers to choose the correct answer (Gierl et al., 2017). According to Schroeder, Murphy, and Holme (2012), it is argued that if a distractor place earlier, it tends to distract stronger than the latter position. They argued that test-takers choose the earlier option if it is plausible without reading latter options, moreover, students tend to select earlier positioned distractor if it is the most plausible distractor and correct answer have equal confidence. This finding is also supported by Tellinghuisen and Sulikowski (2008) which argued that the performance of students and the quality of multiple-choice items are influenced by the position of distractors. In the current study, distractors on phenomenon tier in the earlier position were item P1, P2, P4, P7, P14, P15, while the rest is the latter position. In reasoning tier, distractors in the earlier position were item R2, R4, R5, R7, R8, R9, R10, R11, R14 (9 items in total).



ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE ABILITY	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	Item
1 A	A	0	2	10	-.80	.27	.5	-.26	i1
	D	0	3	15	-.54	.30	.7	-.18	
	B	0	7	35	-.04	.34	1.5	.17	
	C	1	8	40	-.10*	.30	1.3	.12	
3 B	B	0	8	40	-.38	.33	.9	-.18	i3
	C	0	5	25	-.33	.36	1.0	-.08	
		0	4	20	-.02	.43	1.3	.13	
	D	0	1	5	.49		1.6	.21	
	A	1	2	10	-.03*	.00	1.0	.08	
2 C	D	0	3	15	-1.10	.34	.4	-.49	i2
	A	0	3	15	-.40	.80	1.3	-.10	
	B	0	9	45	-.02	.21	1.3	.23	
	C	1	5	25	.08	.19	.9	.22	
4 D	C	0	4	20	-.70	.36	.8	-.32	i4
		0	1	5	-.03		1.4	.06	
	B	1	15	75	-.10*	.21	1.1	.26	
9 E	D	0	1	5	-1.71		.3	-.45	i9
		0	2	10	-.53	.00	.8	-.14	
	B	0	2	10	-.02	.51	1.6	.08	
	C	1	15	75	-.10*	.20	1.1	.26	
5 e	A	0	3	15	-.58	.63	1.0	-.20	i5
	C	0	5	25	-.44	.20	.9	-.17	
	B	1	12	60	-.03	.24	1.1	.30	

Figure 2.2 Example of competing distractor analysis

The use of the Rasch model in the research of multi-tiered multiple-choice questions is still limited to some cases. Some studies (e.g., Chong & Goolamally, 2019; Romine et al., 2015; Sadhu & Laksono, 2018; Lu & Bi, 2016) utilized Rasch model to comprehensively evaluate item analysis by measuring item difficulty measure (SE), MNSQ infit measure, outfit MNSQ measure. According to Boone, Staver, and Yale (2014), the criteria used for the suitability of the outliers or misfits items following: (a) the value of accepted outfit mean square (MNSQ): 0.5 <MNSQ <1,5 (b) the value of tolerated Z-Standard Outfit (ZSTD): -2.0 <ZSTD <+2,0, (c) the value of accepted Correlation Points (Pt Mean Corr): 0.4 <Pt Measure Right <0.85.

Rasch model also has been used to analyze which is more demanding for students between phenomenon tier or reasoning tier. The study combines logit, depicting difficulties of the item, correlation data, and model fit statistics. Data to analyze are two sets from other studies (i.e. data from Singapore and Korea on the Light Propagation Diagnostic Instrument (LPDI), data from the United States on the Classroom Test of Scientific Reasoning (CTSR). Even the result is not consistently delineate which one is more challenging, it can be said that it is more convenient to state propositional knowledge or to make a choice that reflects their understanding of the context (tier 1) than it is to reason through their choice (tier 2) (Fulmer et al., 2015).

Currently, the distractor analysis by utilizing Rasch model software has been used to deepen the analysis of multiple-choice questions (concept inventory). IRT methods have also been proposed that are appropriate for misconceptions distractor-driven multiple-choice (MDDMC) items whose answer choices cannot be ordered in terms of “correctness” or progression along a latent variable. One approach within this category was illustrated by Herrmann-Abell and DeBoer (2011, 2014) based on Rasch measurement theory. The major benefit of Rasch measurement theory is that it is based on the principles of invariant measurement. In other words, Rasch models facilitate the interpretation of student achievement and item difficulty within a single frame of reference, such that inferences about student achievement do not depend on item characteristics and inferences about items do not depend on student characteristics (Wright & Masters, 1982).

Hermann-Abell and DeBoer applied the dichotomous Rasch model (Rasch, 1960/1980) to estimate student and item locations on a single linear continuum that represents a construct, or latent variable. For each item of interest, they used graphical displays to illustrate the proportion of students selecting each answer choice along the range of student achievement estimates. The resulting displays provide diagnostic information that describes the relationship between student achievement levels and the popularity of misconceptions that are included in misconceptions distractor-driven multiple-choice (MDDMC) items. This approach goes beyond what would be obtained using proportions of students selecting each distractor at the pre- and post-test time points because it provides information about the degree to which each answer choice (misconception) is attractive to students at different levels of achievement. Furthermore, the use of the Rasch model is desirable in that estimates of student achievement can be described separately from item difficulty estimates, and that these estimates are on an interval-level scale.

The next study applies and extends the Rasch-based distractor analysis methodology illustrated by Herrmann-Abell and DeBoer (2011,2014) using misconceptions distractor-driven multiple-choice (MDDMC) items that address physical science concepts for eighth-grade students using a pre- and post-test design.

Some studies on this instrument are Herrmann-Abell and DeBoer (2011), Wind and Gale (2015) in which both studies apply similar data analysis i.e. Rasch analysis to reveal students misconception. Both studies have different materials i.e. chemistry and physics for Herrmann-Abell and DeBoer (2011), Wind and Gale (2015) respectively. From both studies, it is concluded that distractor analysis tends to reveal a misconception by showing a high probability of respondents to select the distractors as the correct answer.

As an illustration, in an item which has 5 options that present different categories, a conquest was applied to analyze each option. In coding the answer, answer A was coded 1, answer B till E were coded 2, 3, 4 and 5 respectively. The different code in this analysis was used to show the difference between each option. Figure 2.6 shows the result of plotting each option. From the Figure, it can be seen clearly that the correct answer (option B) has a tendency to be selected higher as the level of students‘ competency rise. It also depicts information in the average ability ($\theta = 0$), the probability of picking the correct answer is 50%, while other options are lower in which option accounted for 30% chance, D option around 10% each, and the remaining 10% for C and E. These result fitted construct map which ordered options: B > D > E=C > A (Briggs et al., 2006).

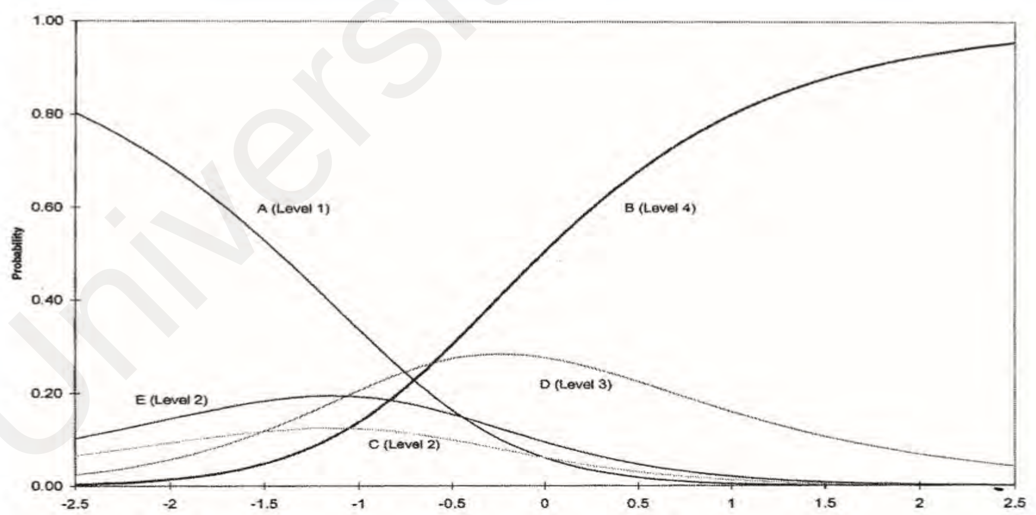


Figure 2.3 Example of IOCC result
Source : (Briggs et al., 2006)

2.5 Theoretical Framework

Based on the analysis of literature review, the importance of this study accentuates on the effort to find the best way of presenting data of two-tier multiple-choice question (2TMC). Therefore, test theory to use is classical test theory which stipulates on the method of findings distractors as the measure to show misconceptions. It is considerable as the most well-applied methods to assess misconceptions. To enhance the analysis, this study used item option characteristics curve (IOCC) to show the probability of each option to be selected. The idea of using IOCC are based on the conclusions of previous studies which stated that IOCC can enhance diagnostics power. Diagnostics power refers to the ability of instrument to explicate more information regarding conceptual understanding.

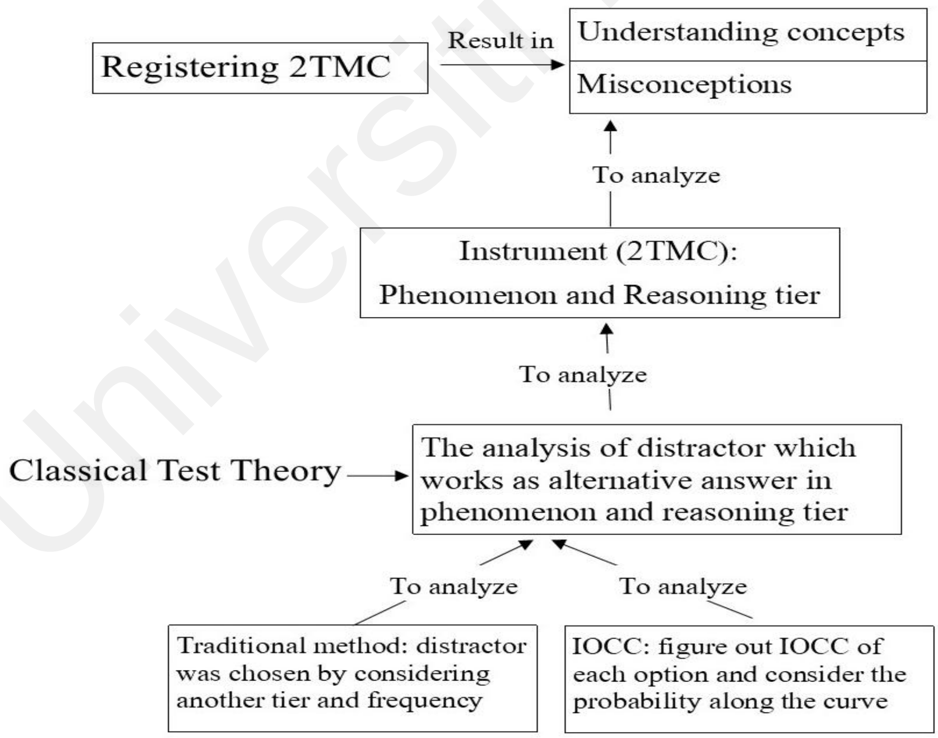


Figure 2.4 Theoretical framework

2.6 Conceptual Framework

Conceptual framework shows variables and concepts of interest in the study. In Figure 2.5, there are two main components of developing multiple choice namely correct answer and distractors. Distractors are the result of vigorous processes to obtain respondents common mistakes (Gurel et al., 2015; Lin, 2016). These common mistakes are elicited from literature reviews, open ended testing, multiple choice with open justifications and semi-structured interviews (Chandrasegaran et al., 2007). In analysing misconceptions using multiple-choice questions, distractors play a very crucial role because it is the only tool to measure respondents' alternative answers. What they choose as an answer, it is the proof of their conceptions.

Multiple-choice questions have two tier: phenomenon and reasoning. Phenomenon tier provides some chemical phenomena in the laboratory such as 1) the appearance of vigorous effervescence and colour changes when dilute hydrochloric acid is added to some grey iron powder, 2) the appearance of deposit and colour changes when powdered zinc is added to blue aqueous copper(II) sulfate and the mixture shaken. From the phenomena, test-takers are asked such as the reasons of colour changes, the appearance of effervescence, the appearance of deposits, choosing correct ionic reactions etc. (Chandrasegaran et al., 2011; Prodjosantoso et al., 2019). It is provided answer choices for every question, and followed by reasoning tier which arrange some reasons to select choice on phenomenon tier. Reasoning tier usually have more choices compared to phenomenon tier (Lin, 2016; Prodjosantoso et al., 2019).

Both tiers produce an instrument namely two-tier multiple choice questions (2TMC) which have been widely used to detect misconceptions (Fulmer et al., 2015; Xiao et al., 2018). For the purpose of this study, an arrow showed that the instrument can produce some types of findings to consider such as achievement on phenomenon tier, achievement on reasoning tier, misconception type-1 (correct phenomenon only) and misconception type-2 (correct reasoning only). To visualize details of which part of 2TMC produce certain findings, there were some arrows directly from both phenomenon and reasoning tier into the four types of findings. There was also the arrow to visualize that there were two methods of analysing type of misconception namely traditional method and IOCC. Traditional method relies on the answer of other tier to determine alternative answer, for instance, to determine alternative answer of reasoning tier we need to consider the correctness of phenomenon tier, while the analysis using IOCC depend only on its tier, meaning that the analysis of distractor at reasoning tier is independent without considering phenomenon tier.

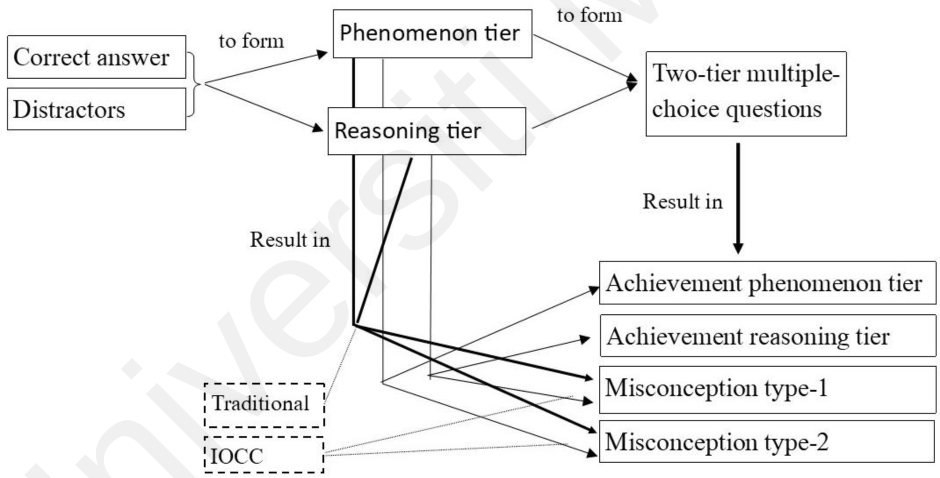


Figure 2.5 Conceptual Framework

2.7 Summary

This chapter has provided discussions about misconception, a term referred to different concept with Chemist‘s view into certain phenomena. To detect misconception, there has been many studies carried out which mostly depend on classical test theory in establishing their psychometric properties. Therefore, it also provided the use of Rasch model as other theory to support the study on the area of diagnostics test. Lastly, after considering many PhD dissertations on misconception analysis which mostly consider constructivism as theoretical framework (e.g. Bain, 2017; Drogemuller, 1994; Greenwood, 2017; Jang, 2003; Kutluay, 2005; Naah, 2012; Schaffer, 2013; Sopapun, 2002), it is decided theory of this study which were supplemented by test theory either classical or Rasch model. The theoretical framework was discussed, followed by the presentation of the conceptual framework.

CHAPTER 3

METHODOLOGY

3.1 Introduction

Assessing students‘ level of understanding is the way to extract the extent of students mastering materials on a subject. The result of the assessment process must represent as close as possible their true level of understanding (Lesage, Valcke, & Sabbe, 2013). Using diagnostics assessment, this study revealed misconceptions of pre-service chemistry teachers. This chapter begins with the explanation of the research design followed by the population of interest, instrumentation, and data analysis. To enhance the consideration of each part, it is provided theoretical information and justification for each part of the methodology.

3.2 Research Design

This study employed a quantitative method and the research design could be considered as a survey design. The data in this study was collected as numerical data in the form of a number of misconception, descriptive statistics of achievement in phenomenon and reasoning tiers, and types of misconceptions based on distractor analysis (Gay, Mills, & Airasian, 2009; Given, 2008). The main reason for the design was the nature of research objectives and research gaps in the literature. The study was also considered as a survey design because it directly explicates the phenomena of pre-service teachers‘ conceptual understanding and compared its result depending on an instrument form without any manipulation of the sample characteristics in a point of time (Creswell, 2012).

3.3 Population and Sample

3.3.1 Population

The population of this study was pre-service chemistry teachers from three universities in a province in the middle part of Indonesia to represent the three varieties of universities in Indonesia: public, Islamic, and private university. The number of population of the study were 323 students in which 130 students were in the first year, 103 students were in the second year and 90 students in the third year. For the population, there were179 students from university A, 96 students from the university B and 48 students from university C. Looking to the quality of the majors based on accreditation from the Ministry of Higher Education, university A and C has grade B while university B is not certified yet due to the new study programme in the university. Based on the rank from www.4icu.org and Ministry of Education, University A and B were ranked first and second respectively. The three institutions were located in almost the same area. It is vital to note that these three universities were the only ones offering this bachelor programme in the province.

3.3.2 Sample

In the current study, the selection of the sample was based on stratified random sampling and the procedure were:

1. The population were divided into three groups according to their year of education namely first year, second year and third year.
2. 65% of the population from each group was randomly selected using SPSS 25. The consideration of choosing the number of samples was the result of calculation employing G* Power 3.0.10.0, where the minimum sample for data analysis utilizing one-way ANOVA, effect size 0.40 and alpha value 0.05 was 102 respondents as shown in Figure 3.2. From this analysis, it was

chosen 72 students as representative of first year (40.19%), 61 students for second year (32.06%) and 52 students for the third year (27.75%).

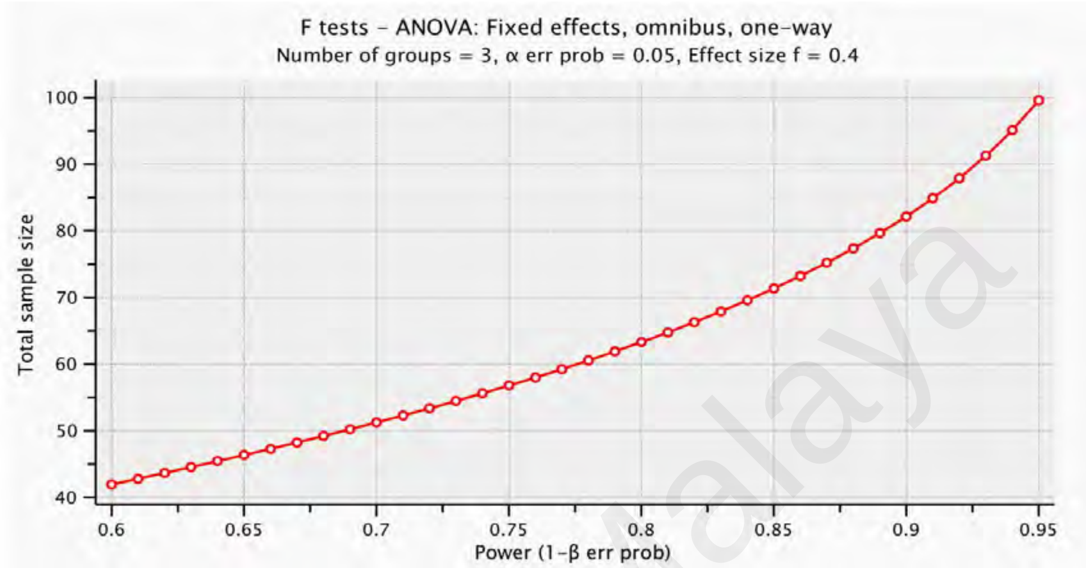


Figure 3.1 Sample Size

The descriptive statistics of the selected sample is shown in Table 3.1. The sample of the study was 185 pre-service teachers (19 males and166 females) between the ages of 18-21 years old. Different number of sample was the result of sampling method which accentuating on the difference of academic years rather than gender information. These pre-service teachers were from the three universities: university A (61.62%), university B (27.56%) and university C (10.81%). The pre-service teachers were at different academic year - first year (40.19%), second year (32.06%) and third year (27.75%). The sample who from three different academic years affects ability of respondents to answer questions or understanding the concept of chemical reactions. The more years student spend in the university, they will engage to more subjects and result in deeper understanding towards chemical concepts. Majority of them are the inhabitants of the province, while only a small number of students are from other provinces such as East Nusa Tenggara, provinces in Java island etc. These pre-service teachers had different academic background such as senior high schools, vocational schools and Islamic boarding schools. School background tend to affect the prior ability of students specifically freshmen since they have different curriculum in senior high school level. Vocational school does not emphasize on chemistry learning, while Islamic boarding school reduce the amount of time to study chemistry which tend to influence student ability compared to general senior high school.

Table 3.1

Demography of the sample of the study

Characteristics	Number of samples	Percentages (%)
Year of education		
1st year	72	40.19%
2nd year	61	32.06%
3rd year	52	27.75%
Gender		
Male	19	10.27%
Female	166	89.73%
University		
A	114	61.62%
B	51	27.56%
C	20	10.81%
Total	185	100%

3.4 Instrument

3.4.1 Two-Tier Multiple-Choice Question (2TMC)

This instrument was adopted from Chandrasegaran et al. (2007) namely Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI) (shown in Appendix B) which published in one of the best journals in the field of chemistry education namely *Chemistry Education Research and Practice* (Teo, Goh, & Yeo, 2014). Proof of permission to use the instrument is shown in Appendix A. In its application in Singapore, the reliability of the 15-item 2TMC was established by a Cronbach alpha coefficient of 0.65. The other studies applying this instrument to the practice of teaching and learning for some purposes comprise the effort of facilitating students to reduce misconception and as a medium of assessing the effectiveness of certain teaching models (Chandrasegaran et al., 2008b; Chandrasegaran, Treagust, & Mocerino, 2009; Chandrasegaran et al., 2011).

Selection of this instrument was based on some considerations: 1) form of the instrument was two-tier multiple-choice questions which has been considered as good instrument to detect misconceptions because it can assess student‘s conception deeper compared to ordinary multiple-choice questions. The extensive use is due to the amalgamation of advantages from either subjective test (interviews, essays, open-ended test) or objective test (multiple choice question) (Lin, 2016; Tsui & Treagust, 2010). Its vivid strengths comprise efficient time for administrating and grading, the number of needed workers, the ability of generalization and the breadth of covered topics and subtopics (Adadan & Savasci, 2012; Rollnick & Mahoona, 1999; Saat et al., 2016); 2) questions emphasize on multiple representations comprising macroscopic, sub-microscopic and symbolic; 3) the opportunity of analyzing the aspect of distractor driven misconception in the options of phenomenon tier and reasoning tier. As a justification, development of the instrument employed interviews, open-ended test, literature reviews and some-stage pilot test to ensure the distractor is from the common misconception of such item (Chandrasegaran et al., 2007). From 15 items, the number of choices is not the same i.e. 2 options (3 items), 3 options (9 items) and 4 options (3 items) for phenomenon tier and 3 options (7 items) and 4 options (8 items), depicting a very attentive selection of distractors empirically and theoretically to represent common misconception.

3.4.2 Translation of the Instrument

Since the mentioned populations are native Indonesia language speakers and the instrument is in English, the translation is necessary for adopting this instrument to meet the local characteristics (Rode, 2005; Wild et al., 2005). The fact that the delivery language for the majority of the sample is in the Indonesian language, except half of the students in University A that taught in bilingual (English-Indonesian language) since the introduction of International Standard Teacher Education (ISTE) for math and science in the university in 2011. The advantages of translation are 1) avoid error in testing because of misinterpretation of questions from students 2) efficient time of administration because of the reduction of required time for students to interpret the question (Wild et al., 2009, 2005).

Method of the translation of 2TMC was back-translation (Callegaro Borsa, Figueiredo, Denise, & Bandeira, 2012; Gjersing, Caplehorn, & Clausen, 2010; Gudmundsson, 2009). The procedures were: 1) The instrument was first translated from English as its original language to Indonesian language by its researchers and reviewed by three graduate chemistry students who currently study at a University in Thailand, Japan, and Australia 2) The raw translation document wass sent to two experts in Chemistry who is good at Indonesian language and English. For the first lecture, he was graduated his Ph.D study from Germany with ten years of teaching experience. Currently, he taught basic chemistry and biochemistry in bilingual (English and Indonesian language) in a public university in West Nusa Tenggara. The second expert is a Professor on inorganic chemistry who was graduated his master from the United States and currently teaching undergraduate and master chemistry education for more than 15 years. From their validations, some parts are revised comprising:

a. Revision of word translation: 1) Bunsen flame (item P1), from ―lampu Bunsen" into ―pemanas Bunsen", 2). Powder (item P1), from ―bubuk" into ―serbuk", 3). Light green colour, from berwarna hijau terang menjadi berwarna hijau pucat, 4). Dissappears, from ―menghilang" into ―habis bereaksi" 5). Become warmer (item P8), from ―menjadi lebih panas" into ―menjadi agak hangat" 6). reddish-brown (item P11), from ―merah kecoklatan" into ―coklat kemerahan" 7) produce, from ―membentuk" into ―menghasilkan".

b. Incomplete words such as writing ―larutan" as the name of many solutions, also ―encer" as the characteristics of the used solutions.

c. Revision of sentence translation such as: 1) item P6 from ―Larutan asam sulfat encer ditambahkan ke dalam bubuk tembaga(II) karbonat yang berwarna hijau. Buih yang banyak dihasilkan dan tembaga(II) karbonat memhilang membentuk warna biru terang" into from ―Larutan asam sulfat encer ditambahkan ke dalam bubuk tembaga(II) karbonat yang berwarna hijau sehingga dihasilkan buih yang banyak dan larutan berwarna biru sampai semua tembaga(II) karbonat habis". 2). (Item P9) from ―Reaksi antara ion H^+ dengan ion OH^- akan menghasilkan air" becomes ―yang bereaksi adalah ion H^+ dengan ion OH^- menghasilkan air". 3) (item P10) from ion Cu^{2+} dihasilkan dari reaksi kimia" into ―Telah dihasilkan ion Cu^{2+} dari reaksi kimia", from ―ion Cu^{2+} yang semula berada pada tembaga(II) oksida yang bersifat tidak larut dalam air sekarang berada pada temabaga(II) sulfat yang bersifat larut dalam air" becomes" ion Cu^{2+} yang semula sebagai tembaga(II) oksida yang bersifat tidak larut sekarang berubah menjadi tembaga(II) sulfat yang larut dalam air. 4) (item P4) from ―larutan ion Cu^{2+} di dalam larutan membentuk warna biru, sedangkan ion Zn^{2+} membentuk larutan tidak berwarna" becomes ―Dalam larutan ion Cu^{2+} berwarna biru, sedangkan ion Zn^{2+} tidak berwarna".

Lastly, the third step is back translation of Indonesian language version of the instrument by a chemistry lecture. He has a Ph.D. in computational chemistry and he was graduated with his master degree from a university in Australia. Currently, he teaches basic chemistry and physical chemistry in bilingual (English and Indonesian language) in Science Faculty and Education Faculty at university in West Nusa Tenggara for almost ten years. 4) The result of back translation and the first draft is reviewed to measure the correctness of the translation. Since their result has the same meaning, there is no revision made and the draft is used for pilot study and its instrument was available at appendix C. Using a chemist to do the translation has been done in other studies where Kumpha, Suwannoi and Treagust (2014) translated 2TMC chemical bonding instrument from Tan and Treagust (1999). Similarly, Vrabec and Prokša (2016) translated Bonding Representations Inventory (BRI) from Luxford and Bretz (2014) from English to Slovak. The use of chemists in translation process is beneficial to evade translation problems such as idiomatic expression, chemical symbol and languages, no equivalent vocabularies, specific phrases and cultural relevance (Su & Parham, 2002).

3.4.3 Validity of the Instrument

Validity is the extent of the instrument can be used to measure the expected construct in a study. It refers to the usefulness, appropriateness, correctness, and meaningfulness of instrument in a study (Fraenkel & Wallen, 2009; Shultz, Whitney, & Zickar, 2014). Since 2TMC is a test instrument, the face and content validity are estimated. Content validity is based on the expert judgment, while face validity requires the respondent to make such judgment about the validity of the instrument (Delgado-Rico, Carrctero-Dios, & Rueh, 2012; Shultz et al., 2014). To measure the content validity of the instrument, the draft of the translated version is reviewed by 3 chemistry lecturers in Chemistry (regarded as subject matters expert (SME)) (Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003; Shultz et al., 2014). The first and second lecture is the same as the experts who reviewed the translation from English to Bahasa. The third lecture represented a private university who has 5 years of teaching experience and has an educational background (bachelor, master, and Ph.D. in chemistry education). They evaluated the appropriateness of the instrument to measure misconceptions concerning chemical reaction involving multiple representations (Raykov & Marcoulides, 2011). Generally, the panel agreed all questions were appropriate to measure chemical reactions using multiple representations. Distractors look well functioned and tend to distract students.

3.5 Pilot Study

A pilot study is defined as a small-scale trial of the proposed procedures which can be used for several purposes (Fraenkel et al., 2012). The purpose of this pilot study was to estimate administration time, the reliability, and goodness of model fit of translated instrument. The sample of the pilot test was 69 students (10 males and 59 females) between the age of 18 to 20 years old. The sample were students from university A as in the real study in 3 different academic years comprising first year (39.13%), second year (30.43%), and third year (30.43%). The administration of the test lasted 45 minutes The descriptive statistics of the sample for the pilot study shown in Table 3.2.

Table 3.2

Pilot test sample

Characteristics	Number (n)	Percentages (%)
Gender		
Male	10	14.49%
Female	59	85.81%
Year of education		
1st year	27	39.13%
2nd year	21	30.43%
3rd year	21	30.43%
Total	69	100%

Data analysis of the pilot study was to estimate construct validation using the Rasch model to find reliability, separation and psychometric properties. As stated in chapter 2 after conducting a literature review, the application of the Rasch model is still limited in the area of diagnostics instruments. It is also supported by the statement of Liu (2010) in the book entitled ―Using and developing measurement instruments in science education: A Rasch modeling approach". The cause of the stagnant replacement of CTT into Rasch model was the lack of training and skills of science educators to apply the theory (Liu, 2010; Romine et al., 2015). The ubiquitous use of CTT could be seen from the application of Cronbach‗s alpha. It is employed to refers to reliability in 2015 by 69 distinct articles at four best tier science education journals in a sole year (Taber, 2018). The four leading papers are International Journal of Science Education (IJSE), Journal of Research in Science Teaching (JRST), Research in Science Education (RISE) and Science Education (SE) (Keith S Taber, 2018; Teo et al., 2014). High Cronbach‗s alpha does not always mean a high degree of internal consistency because it also affected by test length and it also insufficient to estimate unidimensionality or homogeneity (Cortina, 1993; Green, Lissitz, & Mulaik, 1977; Tavakol & Dennick, 2011). The rule of scoring system as follows: 1) students get score 0 if they were incorrect at both phenomenon or reasoning tier, 2) students get score 1 if they were correct at phenomenon tier only or reasoning tier only, 3) students get score 2 if they were correct at both phenomenon and reasoning tier.

Table 3.3

Rubrics of scoring

	Phenomenon	Reasoning	Score
Pattern of answer	Incorrect	Incorrect	0
Pattern of answer	Correct	Incorrect	1
Pattern of answer	Incorrect	Correct	1
Pattern of answer	Correct	Correct	2

(References: Fulmer et al., 2015; Liu, Lee, Linn, & Liu, 2011; Park & Liu, 2019; Sadhu & Laksono, 2018; Xiao et al., 2018).

3.5.1 Reliability

Reliability is the degree to which an instrument consistently give a similar result among numerous administration (Fraenkel & Wallen, 2009; Shultz et al., 2014). To measure reliability, this study applied Cronbach‘s alpha internal consistency to elicit the correlation between a score of an individual item in the test and the total gained score for all items (Chua, 2013). According to Sumintono and Widhiarso (2015), person reliability elicits the stability of student responses in each instrument, while item reliability elicits the stability of item score. The lowest value for Cronbach‘s alpha based on Nunnally (1978) was 0.5 for multiple choice questions, while a reliability of 0.65–0.70 is ―minimally acceptable” and a reliability between 0.70 and 0.85 is ―respectable” for person and item reliability (DeVellis, 2012).

Table 3.3

Reliability and separation of the instrument

	Value
Cronbach's Alpha	.65
Person Reliability	.60
Item Reliability	.76
Person separation	1.23
Item Separation	1.80

In the current study, person reliability was 0.60, below expected score. In the area of diagnostics instrument, some studies (e.g., Caleon & Subramaniam, 2010; Hoe & Subramaniam, 2016; Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018) which published their works on some good articles also find unsatisfactory result for reliability with reliability lower than 0.5 (minimum value is 0.15). Based on Teo, Goh, and Yeo (2014), Chemistry Education Research and Practice is one of two top-tiered chemistry education journals, and International journal of science education is one of four top-tiered science education journal. For Research in Science Education, it is Q1 article since 2009 (cited from schimagojr.com). The next value to consider is separation. Based on Sumintono and Widhiarso (2015), one equation to estimate from item separation is H (separation) = {(4 x separation) + 1}/3 = 2.733, or 3. It means that the items can differentiate the ability of respondents into high, moderate and low. Similarly, the separation higher than one indicate good spread of item and person (Chan, Ismail, & Sumintono, 2015; Gracia, 2005).

3.5.2 Construct Validation

As the proof of construct validation, analysis of items in the current study utilized Rasch model to renew item analysis based on CTT as conducted by other researchers such as He, Liu, Zheng, and Jia (2016) and Romine et al. (2015) by renew analysis of Zheng, Fu, & He (2014) and Schaffer (2013) respectively. Renew analysis refers to the changes of analysis based on classical test theory into Rasch model. There are some fit statistics to measure such as mean square (MNSQ), tolerated Z-Standard (ZSTD) and Correlation Points (Pt Mea Corr). According to Boone, Staver, and Yale (2014), the criteria: (a) the value of accepted infit and outfit mean square (MNSQ): 0.5 <MNSQ <1,5 (b) the value of tolerated infit and outfit Z-Standard (ZSTD): -2.0 <ZSTD <+2,0 (c) the value of accepted Correlation Points (Pt Mean Corr) must be positive value. The result is in Table 3.4.

Table 4.4

The result of model fit

	Infit		Outfit		
Item	MNSQ	ZSTD	MNSQ	ZSTD	Pt Mea Corr
1	0.8	-1.3	0.88	-0.6	0.38
2	1.19	1.5	1.21	1.5	0.35
3	0.95	-0.3	0.89	-0.6	0.58
4	0.76	-2	0.77	-1.8	0.38
5	0.93	-0.5	0.91	-0.6	0.55
6	0.75	-2.1*	0.76	-1.8	0.43
7	1.17	1.3	1.16	1.2	0.48
8	1.08	0.7	1.07	0.5	0.32
9	1.21	1.7	1.18	1.4	0.34
10	0.98	-0.1	0.96	-0.2	0.41
11	0.94	-0.4	0.93	-0.4	0.56
12	1.05	0.5	1.06	0.5	0.41
13	1.15	1.2	1.22	1.6	0.14
14	0.95	-0.3	0.91	-0.6	0.49
15	1.02	0.2	1.05	0.4	0.32
Mean	.99	.01	.99	.03	.41
SD	.15	1.19	.15	1.10	.11
Min	.75	-2.1	.76	-1.8	.14
Max	1.21	1.7	1.22	1.6	.58

Note: * is the sign for value outside acceptable range

The result of analysis using Winstep version 3.73 found that each item was concluded to fit the Rasch measurement model. Only Item 6 did not fit the criteria of infit tolerated Z-Standard (ZSTD). High values ZSTD at iem 6 suggests that the item failed to differentiate among students in terms of the targeted ability and thus may measure a different construct from the rest of the items. Obviously a large outfit statistic is more problematic than a small outfit statistic (Liu, Lee, Linn, & Liu, 2011). Even there were some values outside the acceptable range especially at both infit and outside Z-tolerated score (ZSTD), they did not appear together. As a result, all items are considered good items to measure students' conceptual understanding (Sumintono & Widhiarso, 2015).

3.5.3 The Analysis of competing distractor

Distractor analysis is important factor to see the quality of an item. There are some vital information to consider the quality of an item such as the number of test-takers selecting a distractors and its point correlation. The other standard for good distractor for diagnostics test is it should be selected by more than 10% test takers (Yan & Subramaniam, 2018). Based on the above criteria, there are 10 distractors that have percentage below 10% such as P3 (A), P4 (A), P5 (7) P12 (B), R2 (3), R4 (1), R6 (1), R13 (4) and R14 (1). P/R belongs to phenomenon or reasoning tier and the number belongs to item order (1-15) and word or number in the bracket represented the distractor.

By using the Rasch model, it is revealed the point correlation between the data code (scored 1), or non-occurrence (scored 0), of this category or distractor and the person raw scores or measure. It was expected that a higher (positive) score showed a better option from the most correct answer to the most incorrect answer. In this study, it was found that 15 items in their phenomenon and reasoning tiers did not entail competing for the distractor. Thus, it was proven by correlation value that a correct answer always has a higher correlation. For example, item P1 (see Table 3.5 which showed distractor analysis) has correlation 0.31 for correct answer, while distractors were negative for option A (-0.26) and option B (-0.03). Detailed result of the analysis of competing disctractor is attached in Appendix C.

Table 3.5

Example of Distractor Analysis

Item	Data Code	Score Value	Data Count	Data Percent	Average Ability	SE Mean	Pt mea Corr
1	A	0	36	52	-0.58	1	-0.26
	B	0	13	19	-0.4	0.18	-0.03
	C	1	20	29	0.11	0.23	0.31
2	A	0	32	46	-0.66	1	-0.31
	B	1	37	54	-0.08	0.16	0.31

Note: Coloured table is the main focus for distractor analysis

3.6 Data collection

Prior to data collection, the preservice teachers were informed that the test was a diagnostic test and the results of the test would not affect their grades. To collect data, the researcher employed paper-and-pencil-based test that provided the chance for the researcher to observe the process of data collection, go gain a better response rate and the affordability of respondents (Zuidgeest, Hendriks, Koopman, Spreeuwenberg, & Rademakers, 2011). Since the instrument was considered as a test, data collection was conducted by administering the test class-by-class to the selected sample (University A six classes, University B three classes, and University C three classes). To do this, the researchers contact personally to lectures of each university and asking their permission to give the test to their students in their teaching schedule. The lectures also accompany the researcher to ensure the smoothness of the process and to control the active participation of the students.

3.7 Data Analysis

3.7.1 Research Questions 1: Misconceptions analysis by detecting alternative answer

Analysis of misconceptions was based on distractor analysis on phenomenon and reasoning tier that showed how preservice teachers provided alternative answers. For reasoning tier, the answer on phenomenon tier was firstly determined, and if it was correct, then the analysis would continue to see if the student had chosen the correct or alternative answer at reasoning tier. If the preservice teacher had chosen the alternative answer at the reasoning tier, n simple manner, this analysis is said to be ―correct at phenomenon tier only, while reasoning tier was incorrect". For instance, Item 1, there were 99 participants (53.51%) who answer correctly at the phenomenon tier but only 71.71% of them chose the incorrect answer at the reasoning tier. This could be categorized as misconception type-1. The illustration was available in the Figure 3.3.

When a piece of silvery magnesium ribbon is lighted using a Bunsen flame, the magnesium burns with a dazzling white flame and a white powdery ash is produced.

What change in mass would you expect to find?

A The mass of the white powdery ash is less than the mass of the magnesium burned.

B The mass of the white powdery ash is the same as the mass of the magnesium burned.

C The mass of the white powdery ash is greater than the mass of the magnesium burned. (99 participants or 53.51%)

The reason for my answer is:

1 Atoms of magnesium have combined with oxygen molecules.

2 The mass of the ash is greater as there are larger molecules in magnesium oxide. (28 participants)

3 The atoms of the magnesium ribbon get separated and form the white powder which is less dense than the magnesium burnt. (36 participants)

4 The white powdery ash will have the same mass because the atoms in the magnesium are only being heated up and are still in the solid form. (7 participants)

Figure 3.2 Example of analysis to find alternative answer using traditional method

From the findings, it was clear that the alternative answer for reasoning tier was option 3 (selected by 36 participants). From the data, we can explain the misconceptions of pre-service chemistry teachers. The same analysis could be conducted for reasoning tier.

3.7.2 Research Questions 2: 2TMC analysis using IOCC

Before running analysis using item option characteristics curve (IOCC), the analysis should begin with the analysis of ability differences among the year of study. The purpose of this analysis was to see the possibility of having good spread of student ability. To see the differences among a group of the sample, one-way Analysis of Variance (ANOVA) was employed. Student answers were coded as a dummy variable, 1 for the correct answer, 0 for an incorrect response. Before conducting this test, raw data were converted to logits by Winstep version 3.7.3 in order to change the data from an ordinal variable into an interval variable. The new set of data were measured its normality and homogeneity to meet all assumptions of the ANOVA test.

Test assumption

Before deciding the use of the parametric test in data analysis, there are some assumptions need to meet which is its basic requirements. Fulfillment or violation of the assumption determines the type of statistical test whether a parametric test or nonparametric test and the interpretation of its result. The assumptions of ANOVA are data scale, independence, normality, homogenous variance (Field, 2009; Huberty & Olejnik, 2006; Tabachnick & Fidell, 2007).

Data scale and Independence

Data of the dependent variable must be categorized interval or ratio scale for scientific conception, misconception, and lack of knowledge. In this study, each data ranges from 0 (lowest score) to 15 (highest score) and categorized as ordinal scale. Independence also refers to the subjects of this study has independence in terms of their responses. Performance of a student is not influenced by his or her colleagues (Stevens, 2009).

Normality

To estimate normality, this study used the skewness and kurtosis. Skewness is the skew of distribution indicates how much a distribution ―leans” toward low scores or high scores, relative to the mean. It also assesses the extent to which a variable‘s distribution is symmetrical or skewed (towards the left or right tail) of the distribution (Hair, Hult, Ringle, & Sarstedt, 2014; Pagano, 2009). The kurtosis measures how much the distribution is peaked or flattened as compared to the normal distribution (Hair et al., 2014). The criteria for data has normal distribution based on skewness and kurtosis test i.e. the value must be the ranges of ±1.96 (Chua, 2013; Raykov & Marcoulides, 2011).

Homogeneity

Homogeneity refers to the equal variance of the population from samples, meaning that the variability of scores for each of the group is similar. The method to perform this test is the Levene test from SPSS. In interpreting the result, the population is regarded to be homogeneous if its significant score is higher than 0.05 (Chua, 2013; Stevens, 2009).

ANOVA Test

One-way Analysis of Variance (ANOVA) was conducted in order to test the hypothesis of the study. If there is a significant result at ANOVA test, it is added with Scheffe post hoc test to determine which groups have a significant differences of result.

Item Option Characteristics Curve

This analysis aimed to empirically reveal the misconception of pre-service chemistry teachers. Data for this analysis were the student's responses in 2TMC instrument with the demography of the sample can be seen in Table 3.1. The principle of this analysis was the changes in student's level of understanding over time, and it is expected that pre-service teacher scientific conception increase by the rise of the year of education.

The procedure of data analysis in the study are:

1. To reveal the misconceptions of pre-service chemistry teachers using distractor analysis, students' answers were coded A, B, C, and D (phenomenon tier) and 1, 2, 3 and 4 (reasoning tier) for each item. Winstep version 3.7.3 was employed for the analysis to figure out the probability of each distractor to be chosen over time. The figure of the item options characteristic curve (IOCC) was analyzed for each item to determine which students are holding misconception.
2. To analyze alternative answers based on the figure of the item options characteristic curve (IOOC), the distractor which was consistently had high probability for being selected along the curve was considered as alternative answer. For instance, the Figure 3.4 below is the IOCC of item P1. From the figure, option C is the correct answer, while the distractors are option A and B. Looking the line of both distractors, the probability of selecting option A was higher compared to option B, and it is considerable as alternative answer. Detailed result for this analysis was available in chapter 4 table

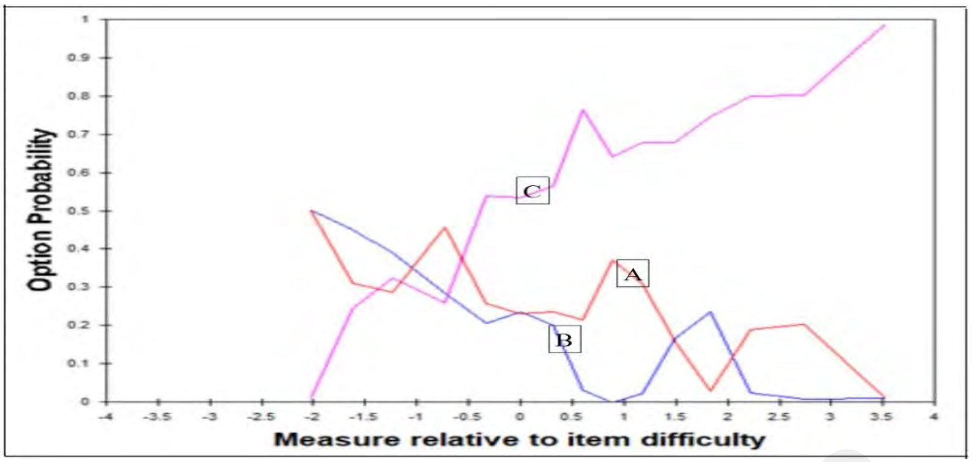


Figure 3.3 IOCC item P1

Table 3.6

Summary of Data Analysis

Research Question	Instrument	Data Analysis
To reveal selected pre-service chemistry teachers' in West Nusa Tenggara misconceptions on chemical reactions by analyzing two-tier multiple-choice questions (2TMC) using traditional method.	Phenomenon and reasoning tier 2TMC	Finding percentage of each distractor
To compare the findings of alternative answers between analysis using traditional method and item option characteristics curve (IOCC) as proposed by this current study.	Phenomenon and reasoning tier 2TMC	Drawing IOCC, and compare each item alternative answer

3.8 Summary of the chapter

This chapter has highlighted the design, sample, instrumentation, data collection and data analysis of the study. The study was survey design, which directly test pre-service chemistry teachers' understanding on chemical reaction using a diagnostic test. After establishing validity and reliability of the instrument, data were collected and analysed using traditional method and a proposed method. In the next chapter, the result of the two analyses would be discussed to see the possibility of using item option characteristics curve (IOCC) as an alternative method for detecting misconception using two-tier instruments.

CHAPTER 4

RESULTS

4.1 Introduction

This study analyzed the misconceptions of pre-service chemistry teachers in West Nusa Tenggara on chemical reactions. The analysis was based on alternative answer (distractors) to gauge their alternative answer in responding two-tier multiple-choice questions. The analysis was merely similar to many studies utilizing 2TMC in chemistry (e.g., Artdej, Ratanaroutai, Coll, & Thongpanchang (2010), Chiu (2007), Tüysüz (2009), Akkus, Kadayifci, & Atasoy (2011), Tan, Goh, Chia, & Treagust (2002), Voska & Heikkie (2000), Peterson, Treagust, & Garnett (1986), Treagust (1986)) which is mostly based on Classical Test Theory. To strengthen the diagnostics power, this study compared and added on the analysis using item option characteristics curve (IOCC) as many previous studies (e.g., Herrmann-abell & Deboer, 2016; Herrmann-Abell & DeBoer, 2011; Wind & Gale, 2015). In using IOCC, there was comparison of achievements on phenomenon and reasoning tier which is the requisite analysis before conducting IOCC to confirm the availability of data spread. The spread was useful to allow the curve along the graph to show the changes of probability of each option to be selected along the student measure from low achievers to high achievers. From the analysis, strong distractors interfere students to select correct options are identified and followed by analysis of unusual curve in IOCC graph of its item. Figures and explanations of the IOCC graphs are provided as part of delineating the strengths of the analysis.

4.2 Research Objective 1: Analysis of misconceptions using traditional method

The first analysis to describe was pre-service chemistry teachers‘ misconceptions based on distractor analysis using traditional analysis. Before explaining further findings, it was necessary to provide the information of item coding. There are seven chemical reactions, each of them has two or three items which is used for classifications in part of description of pre-service chemistry teachers‘ misconceptions based on the result of registering two-tier multiple- choice questions. There were fifteen items which have phenomenon and reasoning tier each, for instance item 1 consist of item P1 and item R1, where P = phenomenon, R = reasoning.

The method of analysis was illustrated in chapter 3 section 3.7.1 which visualized by Figure 3.3. In this analysis, there were some findings to analyze as presented in Table 4.1 such as 1) how many pre-service teachers (percentages) can answer correctly phenomenon tier, 2) Percentage of pre-service teachers correctly answer both phenomenon and reasoning tier 3) When pre-service teachers answer reasoning tier incorrectly, phenomenon answer was correct, which alternative answer they choose.

For instance, from 185 respondents of item 1 ―when a piece of silvery magnesium ribbon is lighted using a Bunsen flame, the magnesium burns with a dazzling white flame and a white powdery ash is produced. What change in mass would you expect to find?" there were 53.51% (99 respondents) who correctly answer phenomenon tier (answering option C ―the mass of the white powdery ash is greater than the mass of the magnesium burned"), and from the number, only 28.28% (28 respondents) who correctly respond reasoning tier (option 1 ―Atoms of magnesium have combined with oxygen molecules”. They chose other options as alternative answer such as option 2 “the mass of the ash is greater as there are larger molecules in magnesium oxide” (28.28%), option 3 “the atoms of the magnesium ribbon get separated and form the white powder which is less dense than the magnesium burnt” (36.36%) and option 4 “the white powdery ash will have the same mass because the atoms in the magnesium are only being heated up and are still in the solid form” (7.07%), and concluded that the alternative answer for reasoning tier (misconception type-1) was option 3. Detailed result for other items was in Table 4.1 with the alternative answer was given sign by thicken their words.

Table 4.1

Distractor analysis for phenomenon tier

Item	Percentage (n) of pre-service teachers answer correctly	Percentage (n) of pre-service teachers have a scientific conception	Alternative Reason	Percentage
1	53.51% (99)	28.28% (28)	(Option 2) n = 28	(28.28%)
			(Option 3) n = 36	**(36.36%)**
			(Option 4) n=7	(7.07%)
2	44.86% (83)	37.35% (31)	(Option 1) n=11	(13.25%)
			(Option 2) n = 19	(22.89%)
			(Option 3) n = 22	**(26.51%)**
3	27.56% (51)	31.37% (16)	(Option 2) n = 11	(21.57%)
			(Option 3) n = 14	**(27.45%)**
4	27.03% (50)	36.00% (18)	(Option 1) n=4	(8.00%)
			(Option 2) n=21	**(42.00%)**
			(Option 4) n=7	(14.00%)
5	21.08% (39)	53.85% (21)	**(Option 1) n= 10**	**(25.64%)**
			(Option 2) n=8	(20.51%)
6	52.43% (97)	56.70% (55)	(Option 1) n=19	(19.59%)
			(Option 3) n=23	**(23.71%)**
7	25.41% (47)	68.09% (32)	**(Option 1,3) n=7**	**(14.89%)**
			(Option 2) n=1	(2.13%)
8	49.78% (92)	51.09% (47)	(Option 1) n=15	(16.30%)
			(Option 2) n= 22	(23.91%)
			(Option 3) n=8	(8.69%)
9	45.41% (84)	71.43% (60)	**(Option 1) n=13**	**(15.47%)**
			(Option 2) n=11	(13.09%)

Table 4.1, (continue)

Item	Percentage (n) of pre-service teachers answer correctly	Percentage (n) of pre-service teachers have a scientific conception	Alternative Reason	Percentage
10	31.89% (59)	37.29% (22)	(Option 1) n=8	(13.56%)
			(Option 2) n=16	**(27.12%)**
			(Option 3) n= 13	(22.03%)
11	48.65% (90)	44.44% (40)	(Option 1) n=18	(20.00%)
			(Option 2) n=32	**(35.55%)**
12	34.59% (64)	40.63% (26)	(Option 1) n=9	(14.06%)
			(Option 3) n= 29	**(45.31%)**
13	52.97% (98)	29.59% (29)	(Option 1) n=13	(13.26%)
			(Option 3) n=42	**(42.86%)**
			(Option 4) n=14	(14.28%)
14	39.50% (73)	43.84% (32)	(Option 1) n=6	(8.22%)
			(Option 2) n=7	(9.59%)
			(Option 3) n= 28	**(38.36%)**
15	32.43% (60)	61.67% (37)	(Option 1) n=5	(8.33%)
			(Option 2) n=18	(30%)

From this analysis, it was found that the percentage of pre-service chemistry teachers who answered correctly on phenomenon tier were ranging from 21.08% to 53.51% with average 39.14 (SD=11.22). The percentage of pre-service teachers who can answer correctly reasoning tier when they are correct in phenomenon tier are ranging from 28.28% to 71.43% with average 46.11% (SD=13.80). From the table, it can be seen that some distractors are selected by more pre-service teachers compared to correct answer. For instance, questions number 1 in the reaction between magnesium metal and oxygen, only 28.28% pre-service teachers chose correct reasoning, while distractors are chosen by 36.36% (option 3). Similar trends can also be witnessed in some items such as item P12, P13 etc.

This analysis was also conducted for reasoning tier with the detailed result was in Table 4.2. In the table, alternative answer when respondents correctly answer reasoning tier while they were incorrect at phenomenon tier are presented. As an instance, from 185 respondents of item 1 —when a piece of silvery magnesium ribbon is lighted using a Bunsen flame, the magnesium burns with a dazzling white flame and a white powdery ash is produced. What change in mass would you expect to find?”, there were 30.81% (57 respondents) who correctly answer reasoning tier “(option 1 “Atoms of magnesium have combined with oxygen molecules”, and from the number, only 49.12% (28 respondents) who correctly respond phenomenon tier (option C “the mass of the white powdery ash is greater than the mass of the magnesium burned”). They chose other options as alternative answer such as option A “the mass of the white powdery ash is less than the mass of the magnesium burned” (40.75%), option B “the mass of the white powdery ash is the same as the mass of the magnesium burned” (10.53%) and concluded that the alternative answer for phenomenon tier (misconception type-2) was option A. The alternative answer was given a sign by thicken their words, and detailed result for the whole items were in Table 4.2.

Table 4.2

Distractor analysis for reasoning tier

Item	Percentage (n) of pre-service teachers reason correctly	Percentage (n) of \ pre-service teachers have a scientific conception	Alternative Answer	Percentage
1	30.81% (57)	49.12% (28)	**(Option A) n=23**	**(40.75%)**
			(Option B) n=6	(10.53%)
2	26.49% (49)	63.27% (31)	**(Option A) n=18**	**(36.73%)**
3	24.86% (46)	34.78% (16)	(Option A) n=6	(13.04%)
			(Option C) n=24	**(52.17%)**
4	36.22% (67)	26.87% (18)	(Option A) n=10	(14.93%)
			(Option B) n=24	**(35.82%)**
			(Option C) n=15	(22.39%)
5	37.30% (69)	30.43% (21)	(Option B) n=13	(18.84%)
			(Option C) n=35	**(50.72%)**
6	52.43% (97)	56.70% (55)	(Option B) n=26	(26.80%)
			(Option C) n=16	(16.49%)
7	29.19% (54)	59.26% (32)	**(Option B) n=11**	**(20.37%)**
			(Option A) n=2	(3.70%)
			(Option D) n=8	(14.81%)

Table 4.22, (continue)

Item	Percentage (n) of pre-service teachers reason correctly	Percentage (n) of \ pre-service teachers have a scientific conception	Alternative Answer	Percentage
8	44.32% (82)	51.09% (47)	(Option B) n=6	(6.52%)
			(Option C) n=29	**(31.52%)**
9	53.51% (99)	60.61% (60)	**(Option B) n=39**	**(39.39%)**
10	32.97% (61)	36.07% (22)	(Option A) n=8	(13.11%)
			(Option C) n=31	**(50.82%)**
11	36.76% (68)	58.82% (40)	**(Option B) n=14**	**(20.59%)**
			(Option C) n=14	**(20.59%)**
12	45.95% (85)	30.59% (26)	**(Option B) n=30**	**(35.29%)**
			(Option C) n=29	(34.12%)
13	29.19% (54)	53.70% (29)	**(Option B) n=25**	**(46.29%)**
14	32.97% (61)	52.46% (32)	**(Option A) n=13**	**(21.31%)**
			(Option B) n=13	**(21.31%)**
			(Option D) n=3	(4.91%)
15	43.78% (81)	45.78% (37)	**(Option A) n=22**	**(27.16%)**
			(Option B) n=22	**(27.16%)**

From this analysis, it was found that the percentage of pre-service chemistry teachers who answer correctly on reasoning tier are ranging from 24.86% to 53.51% with average 37.12 (SD=9.03). Compared to phenomenon tier (M=39.14, SD=11.22), the average was lower which indicating it was slightly more difficult with 2.02 mean difference. This findings were consistent with findings of some previous studies (Fulmer et al., 2015; Liu et al., 2011; Tan et al., 2002) which concluded that reasoning tier tend to be more difficult compared to phenomenon tier.

The percentage of pre-service teachers who can answer correctly reasoning tier when they are correct in phenomenon tier are ranging from 26.87% to 63.27% with average 47.29% (SD=12.39). This average only has small difference with the percentage of pre-service teachers who can answer correctly reasoning tier when they are correct in phenomenon tier. Their difference was only 1.18 higher, while standard deviation difference was only 1.41 lower. It implies that respondents can answer both tiers slightly better when they know the reasoning tier.

Description of pre-service teachers's misconception

This description is based on the analysis using traditional method as presented in Table 4.1 and 4.2. The difference of this analysis and prior studies (Chandrasegaran et al., 2009, 2011) is the analysis of reasoning tier as in Table 4.2 which were not presented in the previous articles. The classification of explanations are based on phenomena which represented different reaction. They are the burning of magnesium (item 1-2), reactions of dilute hydrochloric acid and grey iron powder (item 3-4), dilute sulfuric acid and green copper(II) carbonate (item 5-6), strong acids and strong alkalis (item 7-8), lead(II) nitrate and potassium iodides (item 12-13), displacement reaction (item 14-15).

The burning of magnesium ribbon. In the explanation of mass changes of burnt Magnesium ribbon, rather than explaining the case because of a chemical reaction, some respondents think that it is caused by the changes in molecule size (28.28% of respondents) and density (36.36%). When the respondents know the reason, as many as 40.75% of prospective teachers answer the mass of white powdery ash is lower than magnesium burning. This condition has shown there is problem in chemical reaction and conservation mass law as some prior studies (Agung & Schwartz, 2007; Ozmen & Ayas, 2003).

In item 2, the request to select the symbol of magnesium ribbon, some prospective teachers stated incorrect reasons by selecting the cause as magnesium has high reactivity (22.89% of respondents) and has positive nuclei (26.51% of respondents). When the reason is correct, there are still 36.73% of test-takers think that the magnesium symbol is Mg^{2+}. There is an unexpected relation between distractor in phenomenon and reasoning tier in which the popularity of magnesium has a charge of 2+ is lower than other distractors.

Chemical reactions of dilute hydrochloric acid and grey iron powder. In the reaction, the respondents are asked to identify the changes of solution color to a light green solution. In this question, less than a fourth of pre-service teachers can respond to it due to the production of iron(II) chloride and the presence of Fe^{2+} ions. However, as many as 21.57% and 27.45% of respondents state incorrect reasoning by answering due to chemical reaction and dissolving of iron atoms respectively. The use of a macroscopic approach is evident since 52.72% thought the reason of color changes is the dissolve of iron atoms. Concerning the reaction, the pre-service teachers are also asked the process of the formation of hydrogen gas (item 4) and the selection of ionic reaction (item 5). Hydrogen gas is produced as the reaction of reactive metals with dilute acid and the correction reaction is $Fe_{(s)} + 2H^{+} \rightarrow Fe^{2+}_{(aq)} + H_{2(aq)}$ which were correctly selected by less than 25% of respondents. However, when they know the answer, 42% of respondents stated incorrect reason by thinking that hydrogen is highly reactive because of having one valence electron. When they have a correct reason, some think it is a general truth that the reaction between an acid and all metals will always form hydrogen. In item 5, pre-service teachers who know general guidelines of writing ionic reaction, they choose incorrect reaction i.e. $Fe^{2+}_{(aq)} + SO_4^{2-}{}_{(aq)} \rightarrow FeSO_{4(aq)}$

The chemical reaction between dilute sulfuric acid and green copper(II) carbonate. In the explanation of the reason of the appearance of a blue solution as the result of the reaction between dilute sulfuric acid and green copper(II) carbonate, some respondents think the cause is displacement reaction rather than the presence of Cu^{2+} ions. When they know its reason, some respondents think the reason is the dissolve of copper(II) carbonate in acid solution. The following question is the type of produced gas which can be identified by 25.41% of respondents. Mostly they know the reason for the carbon dioxide which is due to the reaction of carbonate ions and hydrogen ions.

Chemical Reaction between strong acids and strong alkalis. In the reaction between dilute nitric acid and aqueous sodium hydroxide, around a half of pre-service teachers know the result of the reaction between equal number H^+ ions and OH^- ions is a neutral solution. In contrast, there were 23.91% of respondents state a wrong reason i.e., the reaction between Na^+ ions and NO_3^- form $NaNO_3$. This findings are similar to Kind (2004) which found students could not identify neutralization reactions and students perceived this as removing acid properties. The alkali may stop the action of an acid, or alternatively the acid may break down. Students' problems also may arise because acids and alkalis both look like water (Kind, 2004). When they know the reason, accounting for 31.52% of respondents think sodium hydroxide becomes more dilute. When the reactants are displaced by HCl and KOH, pre-service teachers mostly provide a correct reason. Unfortunately, many students (39.39%) who answer reasoning tier correctly think the overall reaction would be different

The reaction of dilute sulfuric acid and black copper(II) oxide. In the reaction, around a third of respondents know the color changes as the effect of production of a soluble salt, copper(II) sulfate and the presence of Cu^{2+} ions. However, many pre-service teachers (27.12%) think the cause is the production of Cu^{2+} ions in the solution. When respondents know the reason, pre-service teachers select incorrect answer i.e., the changes of properties from anhydrous to hydrate. When a dilute hydrochloric acid is replaced by sulfuric acid in the reaction with reddish-brown iron(III) oxide, many pre-service teachers think that the same result because iron(III) oxide molecules are insoluble in dilute H_2SO_4.

The chemical reaction between lead(II) nitrate and potassium iodides. In the reaction, pre-service teachers are asked to identify the correct ionic reaction. As many as 45.31% of respondents think that all ions have to be written in the chemical reaction. However, when they know the correct concepts, around a third pre-service teacher, unfortunately, choose incorrect options. They choose options which writes all involved ions or write spectators ion only. If potassium iodide, KI is replaced by sodium iodide, NaI, respondents who know that the result of the reaction would be the same, unfortunately the reason is incorrect i.e., both compounds produce the same number of ions.

Displacement reaction. In the reaction between powdered zinc and aqueous copper(II) sulfate, the pre-service teachers have been requested to provide the reasons of solution turning colorless with around a third of respondents answer correctly. The reason is the completion of the reaction of copper(II) sulfate, but 38.36% of pre-service teachers answer incorrectly by thinking soluble, blue Cu^{2+} ions have formed insoluble, reddish-brown copper atoms. When the process of the production of reddish-brown deposit as the main attention, students think it was influenced by the impact of Cu^{2+} lost its electron rather than it is removed to form copper atoms. However, when they know the reason, many students think the correct answer is the precipitation of copper and oxidation of copper.

4.3 Research Objective 2: Misconception analysis using IOCC

Misconception analysis in this part are from the aspect of quantitative only i.e. by identifying strongest distractor based on the result of analysis of figures namely Item Option Characteristics Curve. Strongest distractor tells specific misconception that hold by the pre-service chemistry teachers. The distractor refers to alternative answer that chosen by students rather than choosing correct answer. The results were further compared to the findings from traditional analysis at Table 4.6 and Table 4.7. The comparison was based on the similarity of option between two methods of analysis.

4.3.1 Comparison of achievement on phenomenon and reasoning tier

Before running the analysis of the item options characteristic curve (IOCC), it was conducted one-way ANOVA analysis to see the difference of score in phenomenon and reasoning among the year of education. This analysis was vital to do because it can be considered as the determinant of successful analysis using IOCC which requires any spread of ability along its curves. In IOCC analysis, there must be any differences in scores between each group of respondents. This analysis seems like an assumption test in the parametric test which is compulsory to do before IOCC analysis. Descriptive statistics of pre-service teachers‘ abilities as in Table 4.3. It can be seen from the table that pre-service teachers' achievement generally increases by the year of education, even it is a quite surprising result in which the reasoning ability of the second year pre-service teachers is lower than the reasoning ability of the first-year pre-service teachers.

Table 4.3

Descriptive statistics of phenomenon and reasoning tier

Variable	**Mean**	**SD**	**Std Error**	**95% Confidence Interval**	
				Lower Bound	**Upper Bound**
Phenomenon 1st year	-1.0944	.61221	.08490	-1.2649	-.9240
Phenomenon 2nd year	-.4746	1.0606	.14708	-.7699	-.1793
Phenomenon 3rd year	.1458	.92316	.12802	-.1112	.4028
Reasoning 1st year	-.7533	.61148	.08480	-.9235	-.5830
Reasoning 2nd year	-1.1835	.87176	.12089	-1.1835	-1.4262
Reasoning 3rd year	.3229	.64279	.08914	.1439	.5018

Before running one-way ANOVA, the data were firstly reported their normality test and homogeneity as in Table 4.4. It is found that the data are normally distributed for each group.

Table 4.4

Normality test

Variable	Skewness		Kurtosis	
	Statistics	Std. Error	Statistics	Std. Error
Phenomenon 1st year	1.135	.330	.676	.650
Phenomenon 2nd year	.425	.330	.764	.650
Phenomenon 3rd year	-.665	.330	1.031	.650
Reasoning 1st year	.227	.330	-.736	.650
Reasoning 2nd year	-1.021	.330	.998	.650
Reasoning 3rd year	.635	.330	.524	.650

Skewness data for phenomenon and reasoning are ranging from -1.021 to 1.135, while kurtosis data are ranging from -.736 to 1.031. All data are in the range of acceptable value ±1.96 which means that the data are normally distributed. The data are also checked for homogeneity and it is found that either phenomenon or reasoning group have the significance below 0.000, indicating that the homogeneity variance is not assumed. However, the parametric test still can be conducted because this study has a large sample size, meaning that every group has more than 50 samples (Pallant, 2007).

Since all data sets are normally distributed, it is conducted one-way ANOVA test and their results in the following table for phenomenon tier and table for reasoning tier. An analysis of variance showed that there is a difference of pre-service teachers achievement in phenomenon tier among the first year, second year and third-year pre-service teachers, $F(2,182)=70.624$, $p = 0.000$. An analysis of variance showed that there is a difference of student achievement in reasoning tier among the first year, second year and third-year pre-service teachers, $F(2,182)=27.897$, $p = 0.000$.

Table 4.5

Result of ANOVA for phenomenon and reasoning tier

		Sum of Squares	Df	Mean Square	F	Sig.
Phenomenon	Between Groups	71.765	2	35.882	70.624	.000
	Within Groups	92.471	182	.508		
	Total	164.235	184			
Reasoning	Between Groups	42.177	2	21.089	27.897	.000
	Within Groups	137.582	182	.756		
	Total	179.759	184			

Since both analyses suggest that there is a significant difference, the following analysis is post hoc analysis to determine which group have a significant difference with others.

Table 4.6

Scheffe post hoc analysis result

Dependent Variabe	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Phenomenon	third year	second year	.64118*	.16410	.000	.2534	1.0290
		first year	1.18077*	.15823	.000	.8069	1.5547
	second year	third year	-.64118*	.16410	.000	-1.0290	-.2534
		first year	.53959*	.15130	.001	.1821	.8971
	first year	third year	-1.18077*	.15823	.000	-1.5547	-.8069
		second year	-.53959*	.15130	.001	-.8971	-.1821
Reason	third year	second year	1.56256*	.13454	.000	1.2446	1.8805
		first year	1.11441*	.12972	.000	.8079	1.4209
	second year	third year	-1.56256*	.13454	.000	-1.8805	-1.2446
		first year	-.44814*	.12404	.001	-.7413	-.1550
	first year	third year	-1.11441*	.12972	.000	-1.4209	-.8079
		second year	.44814*	.12404	.001	.1550	.7413

From the post hoc analysis, it is found that:

1. There is a significant difference of achievement between first year and second-year of pre-service chemistry teachers on phenomenon tier, mean difference is 1.18077 ($p < 0.05$)
2. There is a significant difference of achievement between second year and third year of pre-service chemistry teachers on phenomenon tier, mean difference is .64118 ($p < 0.05$)
3. There is a significant difference of achievement between first year and third year of pre-service chemistry teachers on phenomenon tier, mean difference is -1.18077 ($p < 0.05$)
4. There is a significant difference of achievement between first year and second year of pre-service chemistry teachers on reasoning tier, mean difference is .44814, ($p < 0.05$)
5. There is a significant difference of achievement between second year and third year of pre-service chemistry teachers on reasoning tier, mean difference is 1.56256 ($p < 0.05$)
6. There is a significant difference of achievement between first year and third-year of pre-service chemistry teachers on reasoning tier, mean difference is 1.11441 ($p < 0.05$)

4.3.2 Comparison of alternative answer using traditional method and IOCC

After running one-way ANOVA analysis, it was conducted the distractor analysis by initially drawing IOCC for each item using Winstep version 3.73. All IOCC figures of each item were available in appendix F. From the analysis of figures with the method as illustrated in section 3.7.2 by reviewing the probability of each distractor to be selected, Findings of alternative answers were reported at Table 4.7 below, while the percentage of test-takers chose any choices were presented at Appendix D.

Table 4.7

Alternative answers in phenomenon and reasoning tier

Item	Phenomenon tier		Reasoning tier	
	Correct Answer	Alternative answer	Correct Answer	Alternative answer
1	C	A	1	3
2	B	A	4	3
3	B	C	1	3
4	D	B	3	2
5	A	C	3	2
6	A	B	2	3
7	C	B	4	3
8	A	C	4	2
9	A	B	3	1
10	B	C	4	2
11	A	B	3	2
12	A	C	2	3
13	A	B	2	3
14	C	A	4	3
15	C	A	3	2

Based on analysis using traditional analysis and IOCC, mostly all items have the same conclusion in the comparison between Table 4.1 and Table 4.2 (traditional method) and Table 4.7 (IOCC). Therefore, detailed misconception of students is the same as the part of 4.2 (description of pre-service chemistry teacher misconception) which generally can be categorized as: 1) The burning of magnesium ribbon, 2) reactions of dilute hydrochloric acid and grey iron powder, 3) dilute sulfuric acid and green copper(II) carbonate, 4) strong acids and strong alkalis, 5) lead(II) nitrate and potassium iodides, 6) displacement reaction.

4.3.3. Analysis of striking IOCC

Generally, past studies (e.g., Wind & Gale, 2015; Herrmann-Abell & DeBoer, 2011) reported their result by explaining exemplary graphs and movement of the probability of each distractor. However, this study only reported some striking points from the IOCC analysis comprising the curve for 2-options (question P2, P9, and P13), unexpected curve after 0 logits (P4, P6, P15, R1), the inconsistency of strongest distractor (P8) and unworking distractor (P3 and R4). Before looking to them, the example of good IOCC is item P12 (Figure 4.1) in the question ionic equations of reactions between colorless aqueous solution of lead(II) nitrate and potassium iodide, KI. Before logit around -0.75, the probability of two distractors (option B and C) has higher probability compared to the key answer (option A). However, after the point, both distractors plunged, while the key answer uplifts till the highest ability of pre-service chemistry teachers.

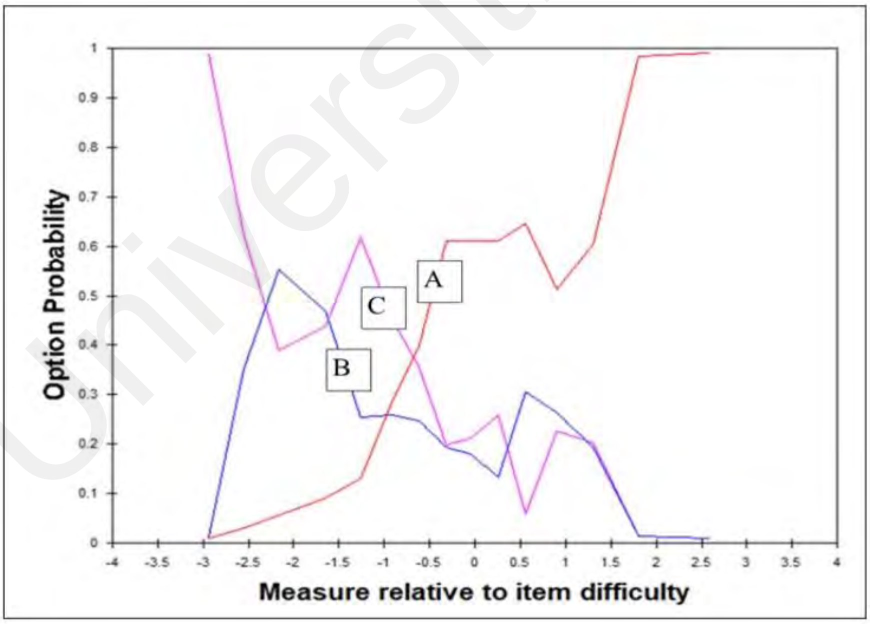


Figure 4.1 IOCC item P12 (Example of good item)

The curve for two-option questions

From the study, some items with two options namely item P2, P9 and P13 in phenomenon tier found the high probability of the distractor to be selected at around zero logits (see figures at appendix F). As seen in Figure 4.2 for item 2, the correct answer (option B) raise its popularity to be selected after more than 0 logits. Basically, two-option multiple-choice is not a good way of constructing multiple choice questions and contradicts the suggested number of options by Epstein (2007); Vyas and Supe (2008) which is optimal for four options or five options. However, three-option is also enough when all distractors are plausible and well-functioned (Rodriguez, 2005). Upon talking about well-functioned distractor, two choices are sufficient for those items. In item P2, it is conceivable for the choice only Mg, and Mg^{2+} since both are the only reasonable symbol for magnesium that can distract respondents. Theoretically, it is difficult and less reasonable to add some choices by adding other choices, as instance Mg^{+}, Mg^{3+}, or Mg^{2-} which were not popular and accepted.

For the case of item P9 and P13, it is feasible to add additional option “undetermined”, which was often used as additional options for “the same” and “different”. However, looking to the chemistry aspect, it is not well-functioned since the properties of the involved compounds tend to be similar and possibly the participants can detect their similar properties by looking the group they belong. In item P9, the compounds involved Kalium (K) and Sodium (Na) which belongs to the IA group and also there are dilute HCl and HNO_3 (considered as strong acid). So, it is possible for respondents to identify are their reaction whether the same or different. In item P13, the compounds consist of Kalium (K) and Sodium (Na) as item P9 which is mixed with the same reactant, lead(II) nitrate. As a result, the condition for both items tend to be similar.

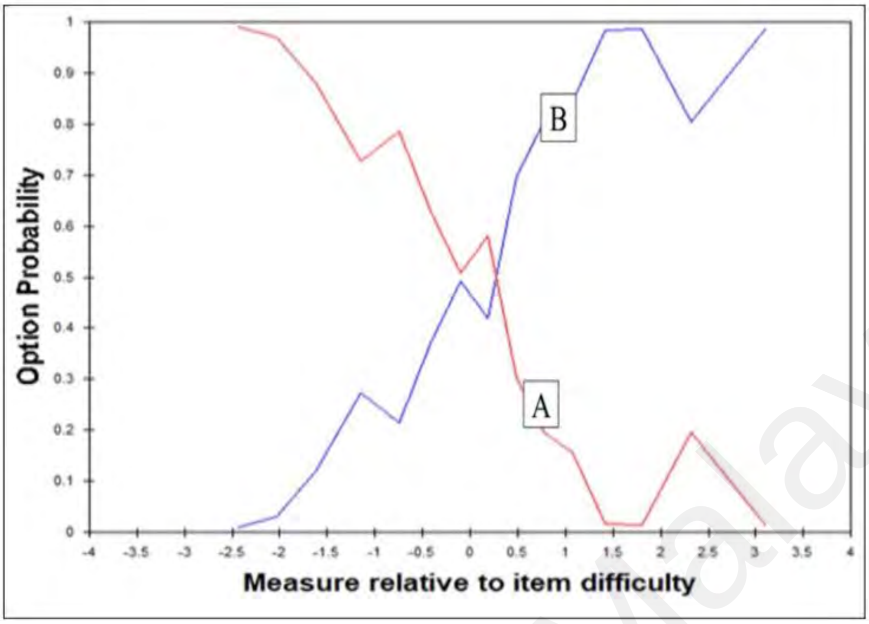


Figure 4.2 IOCC 2-choice question (item P2)

Unexpected curve after 0 logits.

From this study, some unexpected curves were observed in item P4, item P6, item P15 for phenomenon tier and item R1 for reasoning tier. The similarity of the unexpected curve was the direct drop of correct answer probability in high logit measure (high achievers) of pre-service teachers and accompanied by the high rise of a distractor to be selected. This unexpected curve was considered as problem because it does not fulfil the criteria of good answer choices as shown in the example of Figure 4.3.

In item P4, the question of the produced gas in the reaction between dilute hydrochloric acid and grey iron powder, the probability of selecting correct answer and distractors are comparable before logit 0 and respondents with logit around -0.3 to 1.5 can differentiate them precisely as seen in Figure 4.3. Carefully looking to the answer choice, the only difference for the reason of produced gas is kind of metals all metals (distractor) and reactive metals (correct answer)”. Possibly due to the less carefully of respondents, unexpected curve after logit 1.5 was witnessed.

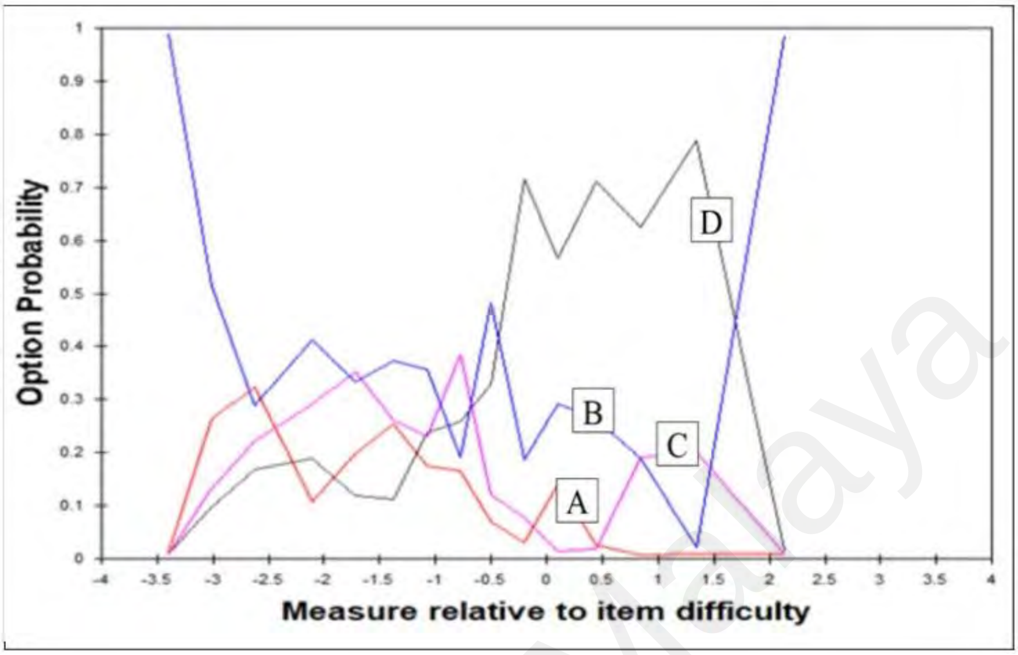


Figure 4.3 IOCC item P4

In item P6, a question about the color changes to a blue solution when adding dilute sulfuric acid into some green copper(II) carbonate powder is also apparent. In the logit measure from around 0 to 1, there is a sharp drop of probability to select a correct answer and followed by the rise of probability choosing other choices. Basically, both distractors have strong influence to be opted for whether the color changes as the result of dissolve (option B) or chemical reaction (C). The problem is possibly because they do not fully understand the color change is the consequence of the presence of Cu^{2+} which is the result of a reaction between copper and dilutes sulfuric acid. As a result, without considering the chemical concept and only contemplate the macroscopic, it is possible to suggest that copper(II) carbonate dissolve in the solution.

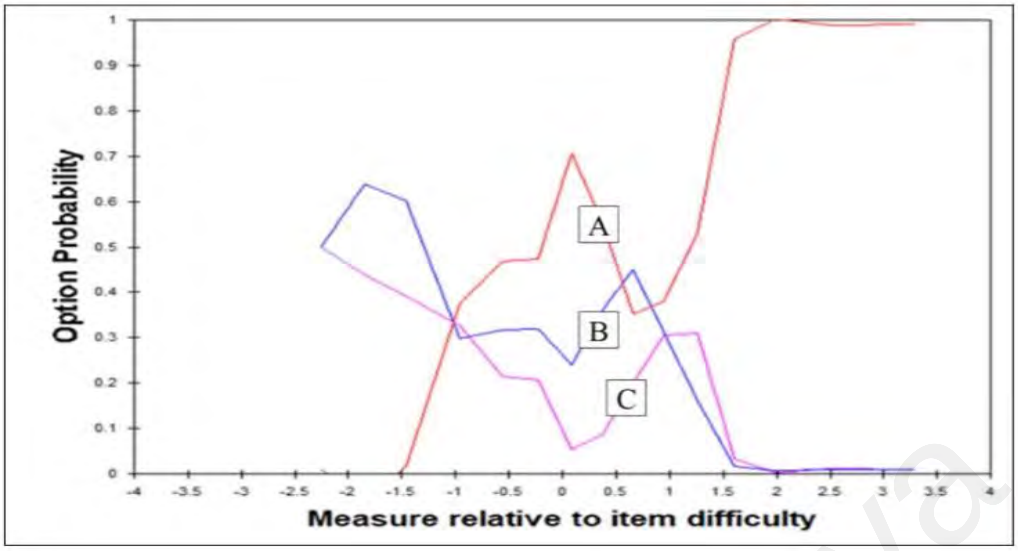


Figure 4.4 IOCC item P6

The last item in phenomenon tier was P15, inquiring the formation of the reddish-brown deposit in the mixing between powdered zinc and blue aqueous copper(II) sulfate. From the Figure 4.5, a sharp plummet of right answer (option C) in the logit measure around 1.6 was observed, followed by the dramatic increase of distractors (option A and option B). Possibly, option A looks more scientific because they consider copper oxidation which refers to the loss of electrons.

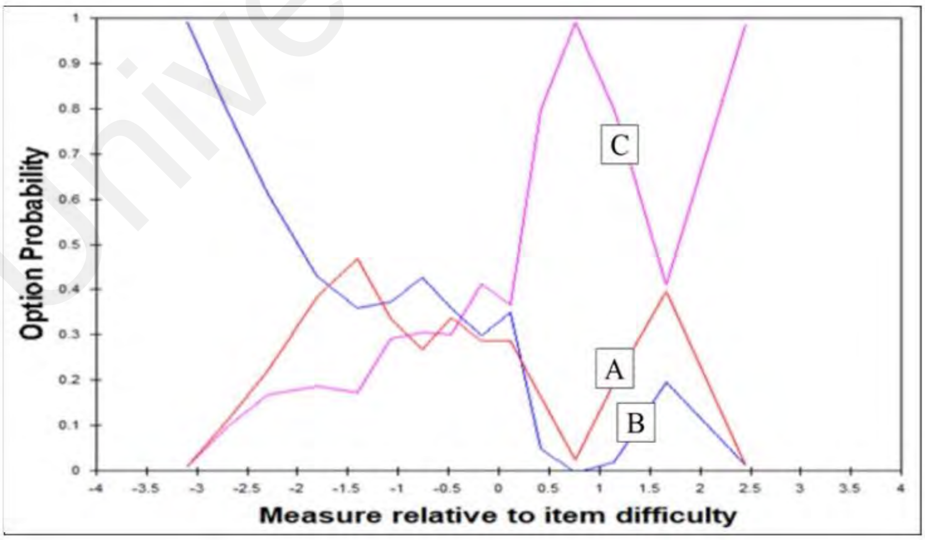


Figure 4.5 IOCC item P15

The only item in reasoning tier with an unexpected curve is item R1 “the reasoning for the mass change of burnt magnesium”. In the logit around 0.75, there is a dramatic decline of true answer probability, while distractor witnessed a surge. It is plausible because option A (P-tier) and option 3 (R-tier) indicated a plunge, suggesting a clue for pre-service teachers to choose (Griffard & Wandersee, 2001; Gurel et al., 2015). At option three, magnesium ribbon are considered to be separated which indicated the possibility of pre-service teachers‘ to use their macroscopic observations “magnesium ribbon changes to be ash”.

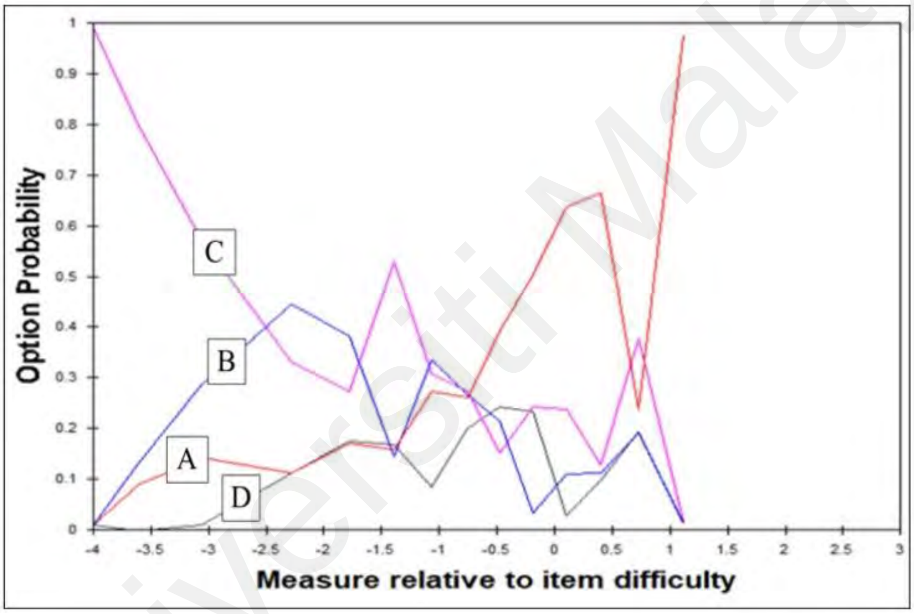


Figure 4.6 IOCC item R1

The inconsistency of the strongest distractor

In the current study, the popularity of distractor to be selected in relation to the correct answer sometimes inconsistent along the graph. For instance, item P15 as in the figure, option A has higher probability before logit 0.2, after the point option B started to precede option A. It implies that option B is more popular for high score pre-service teachers compared to option A. Before logit 0.25, the stronger distractor was the ionic precipitation which possibly due to macroscopic observation. After logit 0.25, the stronger distractor is oxidation of copper, meaning that the compound lost its electron in the reaction process.

The other item having the same characteristics is item P8, in the reaction between dilute nitric acid and aqueous sodium hydroxide. Before logit 0.7, mostly the probability of stating there is no reaction taking place is more popular, but after the point, there is a stronger probability of arguing that aqueous sodium hydroxide turns more dilute.

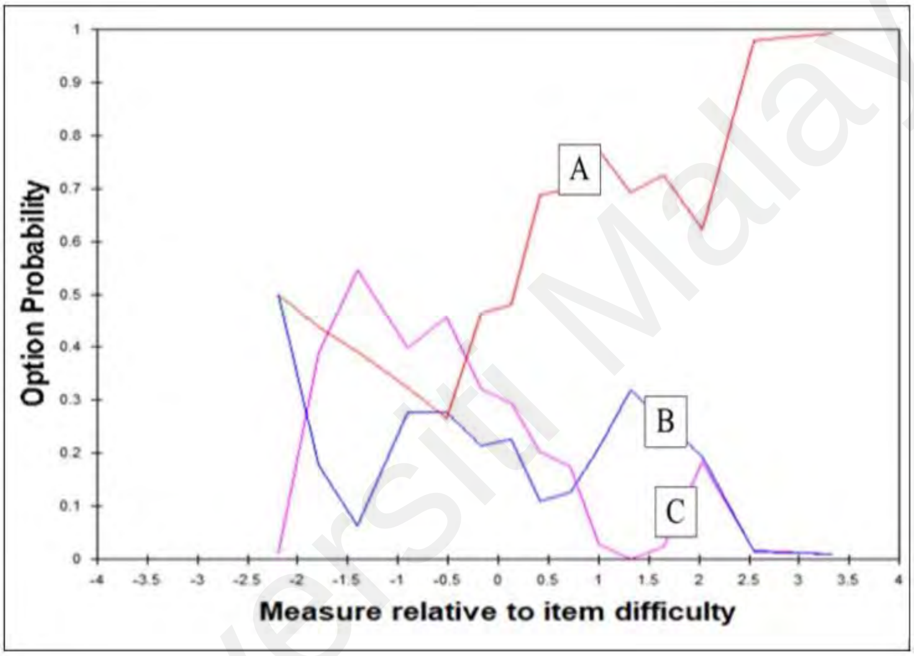


Figure 4.7 IOCC item P8

Less-functioned distractor

According to the IOCC analysis, there were two distractors can be identified as less-functioned distractor namely option A item P3 and option 1 item R4. In the figure, it was evidence that the distractor did not popular to be selected along the curve. It could only distract pre-service teachers with logit in the range of -2.70 to -2.00. The question asks about the colour changes in the reaction between dilute hydrochloric acid and some grey iron powder. A small number of pre-service teachers argue the answer because the colour of iron is green in solution.

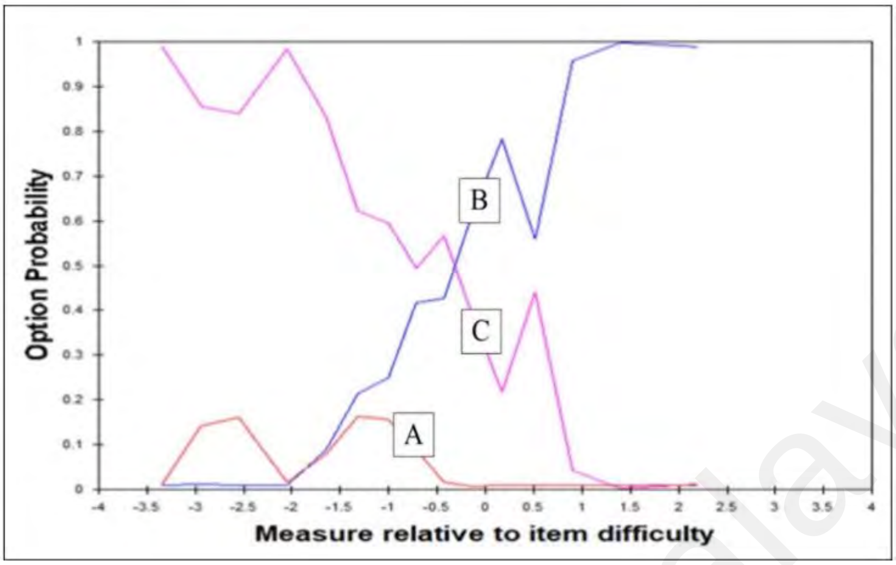


Figure 4.8 IOCC item P3

The other distractor is option 1 in item R4, the question about the reason for the forming of hydrogen gas in the reaction between dilute hydrochloric acid and some grey iron powder. Option 1 ―iron ions are more reactive than hydrogen ions" was not popular with probability below 0.1 along the graph and it did not being chosen by any pre-service chemistry teachers with logit measure more than 0. The possible reason is the similarity of the distractor with correct answer. If the correct answer mentions iron is more reactive, the distractor mention iron ions. However, the correct answer also having more explanations which possibly strengthen respondents to choose. Compared to other option, the option is the shortest one, without any reasoning or additional information which tend to less distract respondents.

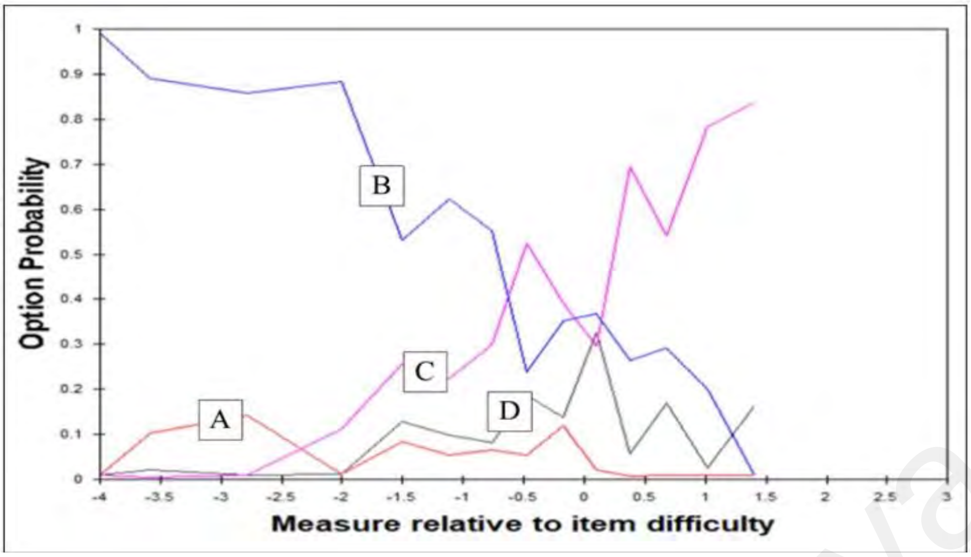


Figure 4.9 IOCC Item R4

4.4 Summary of the chapter

This chapter has provided findings of the study. The finding was any difficulties of pre-service chemistry teachers in each item in two-tier multiple-choice questions. For instance, in the phenomena of burning magnesium ribbon, there were 36.36% of respondents thought the increase of mass of products is the result of magnesium ribbon get separated and less dense without considering chemical reaction takes place. However, when some the respondents can identify the correct reasoning, they chose incorrect option by assuming that the mass of magnesium ribbon decrease. After analyzing using IOCC and compare both analyses findings, the alternative answer of using the traditional method and IOCC found the same conclusions.

Even both analyses found the same alternative answers, there are some points to highlights from IOCC analysis including the unexpected curve for 2 choices (item P2 and P13), unexpected curve after 0 logit of student measure (item P4, P6, P15, R1), the inconsistency of alternative answer or distractor with highest probability (item P8) and unworking distractor for item P3 (option A) and item R4 (option 1). It implies that distractor analysis using item option characteristics curve (IOCC) reveals not only misconception distractor, but it also found some other vital information which tends to expand the power of diagnostics of 2TMC.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1 Introduction to the chapter

In the context of this study, the view of a diagnostics test to elicit misconception is from the aspect of how to determine misconceptions which do not rely on one method, but it also proposes the use of item option characteristics curve. In the previous chapter, the findings were presented and this chapter represents the key findings to link back the similarities and differences with related prior studies in the field. It is also provided the possible reasons for the key findings such as misconception of pre-service teachers on multiple representations on chemical reactions and the differences of misconception level of pre-service chemistry teachers by an additional time of their study in university.

The next discussion is to elicit the use of Item Options Characteristics Curve (IOCC) at two-tier multiple-choice questions as the proposed analysis by this study. This chapter highlights reasons of this analysis are more powerful compared to the common practices of analyzing 2TMC. After discussing all key findings, it is stated the implications of the study, conclusions, and suggestions to future studies.

5.2 Discussion

Discussion of findings was divided into two parts namely misconception of pre-service chemistry teachers and the use of IOCC to reveal misconceptions. In the first part, we discuss the findings of these studies related to misconceptions using both methods. We also added explanations on the changes of ability imperatively understanding and misconception level of pre-service chemistry teachers. The second part explicates the use of IOCC as proposed methods of analyzing two-tier multiple-choice questions (2TMC) to reveal misconceptions. These organizations were based on research questions order even it is modified to give a coherency of the discussion by referring to the findings.

5.2.1 Misconception of pre-service chemistry teachers

This study found that pre-service chemistry teachers experience misconceptions in submicroscopic representation and symbolic representation of chemical reactions as some studies using the same instruments (Chandrasegaran et al., 2008, 2009, 2011). Kind (2004) mentioned the observable indications of chemical reactions take places such as color changes, the appearance of deposit, heat production, and gas production. The problem is the respondents confuse about the reasoning on macroscopic or sub-microscopic representational systems. As an instance, when iron powder reacts with colorless dilute hydrochloric acid, it is produced a green solution of iron(II) chloride (explication at the macroscopic system). The colour change is attributed to the presence of Fe^{2+} in the solution. In this study, 21.57% of respondents suggested it was because of the reaction between iron and chlorine atoms. It indicates possible confusion between the color change at the macroscopic level with changes to the elements iron and chlorine at the sub-microscopic level.

As the previous study utilizing the same instrument (e.g., Chandrasegaran et al., 2008, 2009, 2011), test-takers also have problems with understanding of the multi-faceted significance of chemical symbols, chemical formulae as well as chemical and ionic equations. Pre-service chemistry teachers get difficulties such as identification of sound ionic equations (at item 5, 12, and 13), wrong identification of reacting compounds (item 8), having problems in writing a balanced molecular equation, balanced complete ionic equation, understanding spectators ions. This study is consistent with finding from Agung and Schwartz (2007) which study the ability on stoichiometry and equations of the reaction of twelve grade students at Jakarta and West Java as two leading provinces in Indonesia. As pointed by Chiu (2001), students tend to memorize rules rather than understanding the principles, so they still get difficulties in facing these problems.

Theoretically, pre-service chemistry teachers have misconceptions possibly due to the effect of problems in knowledge acquisition which is influenced by the knowledge and experiences that they bring with them into the science classroom. As stated by Bodner (1986), ―knowledge is constructed in the mind of the learner". It also assumes that we do not discover knowledge; we actively construct it. We invent concepts and models to make sense of our experiences. We then continually test and modify these constructions in light of new experiences (Bodner, 2004).

The other possible problem is the use of multiple representations as found by some previous studies (Chandrasegaran et al., 2011). These terms are abstract and cannot be experienced especially for sub-microscopic and symbolic systems of representation. In addition, the student's thinking is laboriously affected by tangible sensory information. They also sometimes have limited conceptual understanding and poor visual-spatial ability (Chandrasegaran et al., 2008; Keig & Rubba, 1993). In addition, from the aspect of teaching style, teachers also do not highlight the use and connectedness of multiple representations (Chandrasegaran et al., 2007; Gabel, 1999).

Findings of this study confirm the difficulty of students in Indonesia to use macroscopic and sub-microscopic representation in chemistry as reported by Rahayu and Kita (2010) on the particulate nature of matter. According to the study, from 447 Indonesian students aging between 15-18 years old, the achievement of students tends to be lower in sub-microscopic level rather than macroscopic level.

This study is consistent with the finding from Imaduddin (2018) which studied the ability of 36 pre-service chemistry teachers in a university in Central Java, Indonesia about their sub-microscopic misconception on acid and base concept. In general, he found that university students cannot scientifically connect the interrelation among chemistry representation. For instance, they suggest that CH_3COOH as base compound since its symbolic showing OH. It means that they cannot the name of the compound, acetic acid, and its symbol in spite of the lack of knowledge of its component i.e., CH_3COO^- and H^+ as its sub-microscopic level.

The ability of pre-service chemistry teachers connecting three levels of representations can be categorized as problems. To solve the problems, some scholars conducted studies on its improvements such as Farida and Liliasari (2011) which given interventions by using Moodle 2.0 for teaching the topics of salt hydrolysis. From the study, it is concluded that the ability of pre-service chemistry teachers in a University in West Java, Indonesia to connect multiple representations witnessed upward trends between before and after the learning process. Some studies (e.g. Lastri, Kusumo, & Susilaningsih, 2018; Rahmawati, 2015) developed a learning module based on multiple representations as a response of low ability of students in sub-microscopic and symbolic representations. In addition, scholars introduce the use of multimedia in the teaching process to familiarize the application of multiple representations (Agustin et al., 2018)

The findings to highlight from this study are the increase of ability of pre-service teachers to respond chemical reaction diagnostics test as suggested by finding the significant difference of ability in phenomenon and reasoning tier. The analysis was using one-way ANOVA by firstly transforming data into logit. This transformation is beneficial for some reasons. Firstly, the use of raw score fails to delineate score correctly (Saidfudin et al., 2010). For instance, in a chemistry test, student A and B gets 80 and 90 correct answers respectively, while student C and D get 115 and 125 with the differences of A and B, C and D are 10 correct answers. From the result, we cannot concede that student B is better than A because we do not know the category of questions (easy or difficult items) they answer correctly (Neumann, Neumann, & Nehm, 2011). Secondly, the data analysis which transform data into logit is considered useful based on the study from Osman et al. (2011) and (2012). In the evaluation of concrete design course examination of a university student in Malaysia, Osman et al. (2011) applied the Rasch model to present the data. Student‘s achievement was reported using the Wright map, person measure order, and scalogram to precisely showed student‘s performance. They concluded that the Rasch model is beneficial because it can help to reveal the true degree of student‘s abilities and also analyse items to carefully analyse the difficulties of students. With the same methodology, Osman et al. (2012) also analyse the achievement on the course of engineering design II of students in Malaysia. From both studies, the Rasch model is more useful compared to the classical method to analyse student‘s data even there is a small sample (Saidfudin et al., 2010).

For science educators, the benefit of using the Rasch model seems clear if considering the work of Boone, Townsend, and Staver (2011). The use of raw score (incorrectly assuming the data are linear) concluded a significant difference of self-efficacy, while after altering the data into Rasch data set (linear or interval scale) found an insignificant difference of two groups as a function of time point. The study employed Preservice Science Teacher Self-Efficacy Beliefs Instrument (STEBI-B) with the sample of United States Midwestern University which responded to the questionnaire at their beginning and ending of a lesson. From the study, it seems that the enactment of linear data is clearly important.

In chemistry, a study to compare the use of the Rasch model and classical test theory to analyse the effect of the intervention was Pentecost and Barbera (2013). In a study using the instrument of Chemistry Concept Inventory (Barbera, 2013) which tested a number of 2392 students, they illustrate the use of wright map to show the differences of student‘s achievement in pre-test and post-test. The article adds to the literature the powerful advantages of utilizing the Rasch model.

The next things to consider is the possible reasons for changes ability of pre-service chemistry teachers which can be seen from the subject taught in their study. In their first year, they study fundamental chemistry (eight credits) which reviews high school chemistry. In their second years, subjects related to chemical reactions are inorganic chemistry (eight credits) which allowed them to understand more about physical and chemical properties of many chemicals including metals. Another subject is basics analytical chemistry which required pre-service teachers to understand many chemical reactions.

In their third year of the program, they study chemical bond and physical inorganic chemistry. In the subject of the chemical bond, they need to understand chemical reactions from molecular aspects. In the physical inorganic chemistry, pre-service chemistry teachers study many chemical reactions from characteristics and energy. However, all subjects such as biochemistry, physical chemistry, and organic chemistry can contribute to the rising ability on chemical reactions for answering phenomenon and reasoning tier.

The next possible reason is laboratory activities. In their first year, they have two credits of experiments which is part of fundamental chemistry courses. One experiment in the course is mixing magnesium ribbon with oxygen. Mostly all reactions are the experiments in a wide range of courses along with their pre-service teachers‘ chemistry program. The number of laboratory activities increase substantially in their second years because there are many chemistry core courses which required chemical laboratories. For some selected pre-service chemistry teachers, they also work as laboratory assistants. Their job scope includes preparation of experiments, supervise and guide juniors on experiments and mark junior‘s reports.

There are many studies prove that misconception can be reduced by the application of appropriate interventions. An example of intervention to reduce misconception on a chemical reaction that emphasizes multiple representations employs some strategies such as 1) Additional laboratory activities to familiarize students with the chemical reactions, 2). Explanation of the observed chemical changes at the particulate and symbolic levels using small group and class discussions, 3). Emphasizing the significance of coefficients and subscripts in chemical and ionic equations, 4). Deducing ionic equations from observed chemical changes, not by simply canceling out ―spectator ions” in chemical equations along nine months of the program (Chandrasegaran et al., 2009, 2011). At the end of the program, it is found that the ability of students is better compared to students with traditional teaching method as shown by the significant difference at a t-test (Chandrasegaran et al., 2008).

A study which successfully eliminate misconception in inthe topic of boiling concept was Coştu et al., (2007) which evaluated the ability of 52 pre-service science teachers in the primary science education department in a university. In every teaching activity, the researchers have targeted concepts to eliminate such as 1) The properties of the boiling phenomenon, the nature of bubbles in boiling liquid 2) Low external pressure effects on boiling temperature, boiling and vapor pressure, 3) High external pressure effects on boiling temperature, Boiling and vapor pressure. At the end of teaching activity, the students were tested using the two-tier instrument and found a statistically significant improvement of conceptual understanding imperatively misconception elimination.

These findings which found there is the differences of conceptual understanding among years of education e.g., the increase of understanding level and the decrease of misconception level are similar to some prior studies. In a survey study, Herrmann-Abell & DeBoer (2011) tested 13360 students in the USA about middle schools chemistry ideas which its samples consisted of all grades junior high schools, senior high schools, and university students. Topics tested are atomic structure, properties, reactions and many more consisting 91 multiple-choice items. In the study, with the same analysis by firstly transform data into person logit and carried out one-way ANOVA, it is found that students have a significant difference in ability by year of education. However, some grades did not exhibit significant difference such as seventh grade and eighth students, grade eleven to twelve, and college chemistry to graduate chemistry which is similar to second year and third-year students of this study.

In a survey study in the concept of matter (structure and composition, either chemical or physical properties and its changes, and conservation), Liu (2007) also found a gradual increase of ability of students. In the study, the sample was ranging from grade three until grade twelve. One-way ANOVA analysis found a statistically significant improvement in performance. However, looking too detailed post hoc analysis, some ordering grades did not reveal statistically different ability such as the comparison of sixth and seventh grade, tenth and eleventh grade.

Another survey study to test the same materials as Liu (2007) is the study from Liu and Lesniak (2005) which analyzed the result of Third International Mathematics and Science Study (TIMSS) data sets for the United States which offered the most diverse populations. In the study, the analysis is based on Rasch model analysis and found that there is an upward trend of ability which is indicated by person logit. Elementary schools' students have logit below the average of item difficulties, while high school students were at above of the average of item difficulties of conservation topics.

In a qualitative study by involving 54 students from first grade till tenth grade which interviewed the concept of substances and its combinations of water, vinegar, and baking soda, Liu and Lesniak (2006) found there was a change of conception. A cogent progression could be seen from the different level of education in which a higher degree showing a better understanding. In contrast, there is an overlapping conception between two ordered grades, for instance, ninth grade and tenth grade. The overlapping conception refers to the same conceptions or abilities. The findings were the same with this study which found there is no significant difference of ability between the second year and third-year pre-service chemistry teachers.

In a study tested chemistry conception including oxidation-reduction, chemical equilibrium, nature of particles, organic compounds and acid-base in Taiwan, it also concluded that there is a difference of ability of each grade of students. The sample of the study is students from elementary schools, junior high schools, and senior high schools. It can be said that each level of education tends to have different kinds of concepts which will be more demanding to answer. For instance, 20% of juniors and 8% of seniors thought that the size of the gas particles in a balloon increases when the balloon is expanded by heating.

Kahveci (2009) conducted a study to detect conceptions of pre-service chemistry teachers on concepts of elements, compounds, particulate nature of matter and chemical bonding. In the comparison of conceptions, a Kruskal-Wallis test and subsequent Mann-Whitney U showed a significantly different performance between third, fourth and fifth pre-service teacher education. It is exceeded an upward trend of the ability of pre-service chemistry teachers.

5.2.2 The Use of IOCC to Reveal Misconceptions

In this current study, the analysis of misconceptions utilized item option characteristics curve to gauge pre-service chemistry teacher understandings on chemical reactions. The basic idea of IOCC analysis is to examine trace lines for alternative choices (Ding & Beichner, 2009). The distractor analysis plots were created by plotting the proportion of students selecting answer choices A, B, C, and D for phenomenon tier and 1, 2, 3 and 4 for reasoning tier (*y*-axis) across the range of student achievement measures at each time point (*x*-axis). Accordingly, the *y*-axis values indicate the relative popularity of each answer choice for students with different levels of achievement (*x*-axis). In details, after Rasch estimates of student achievement on the logit scale were obtained from the Winsteps computer program

(Linacre, 2014), student achievement estimates on the logit scale were rounded to the nearest integer value (–3 to 4). Then, the frequency of students selecting each answer choice was obtained for each value. At each point on the scale, the proportion of students selecting each answer choice was calculated by dividing the frequency of students who selected a given answer choice by the total number of students observed at each point on the scale (Wind & Gale, 2015).

From the result of the analysis IOCC, it can be said that the correct option is the most popular option to select by students especially students with more than -1.00 logit as shown in details in appendix F. However, this analysis also implies that the students with logit measure lower than -1.00 tend to choose distractor and if they can answer correctly, there is a possibility that they answer correctly guessing since the questions are multiple-choice questions.

This study can be considered as the extension of using IOCC in the diagnostics test. In the study from (Herrmann-abell & Deboer, 2016; Herrmann-Abell & DeBoer, 2011), IOCC was used only to detect misconception. This application is further utilized by Wind and Gale (2015) to analyze data of difference of ability in pre-test and post-test. To continue, this framework is used to analyze distractors on phenomenon and reasoning tier to show student conception.

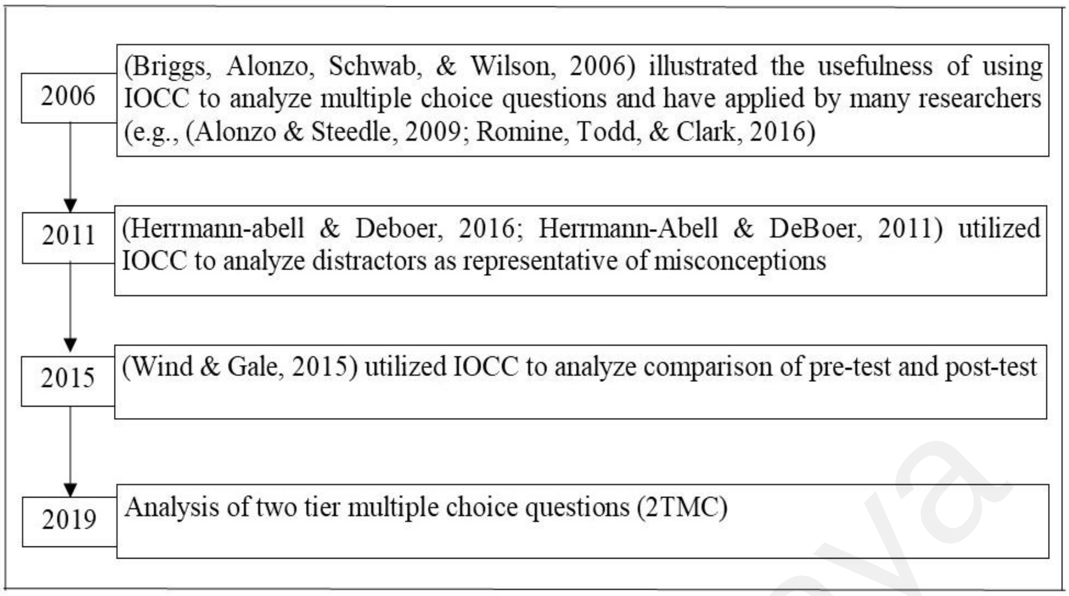


Figure 5.1 The use of IOCC in diagnostics test

Utilizing IOCC to analyse two-tier multiple-choice questions have some advantages compared to traditional method. The first one is traditional analysis can only show the percentages of students having problems, but it does not tell the progression along with the rise of ability. For instance, in item P1, the number of students selecting option A, B, C were 23, 6 and 28 respondents respectively. When we use IOCC, Figure 5.1 illustrated that option C indicated a strong misconception especially for students with ability lower than logit -0.5, while option B only can distract students with ability lower than -1.5. However, above logit of 1.00, option B and C have the same probability to be selected by the students. By having this information, the result of the analysis will be more meaningful because we can detect more details of the student‘s difficulties. These information give teachers a clearer picture of the type of misconceptions that different ability of students would have and at what level they attained these misconceptions. Having a more detailed information about the present of misconceptions in their students‘ minds, will allow teachers to develop instructional materials to combat these misconceptions. The next advantage of IOCC is the appearance of data looks more attractive compared to traditional analysis. As a result, when the data need to be presented, data analysis using IOCC will get more attention because of its better visualization.

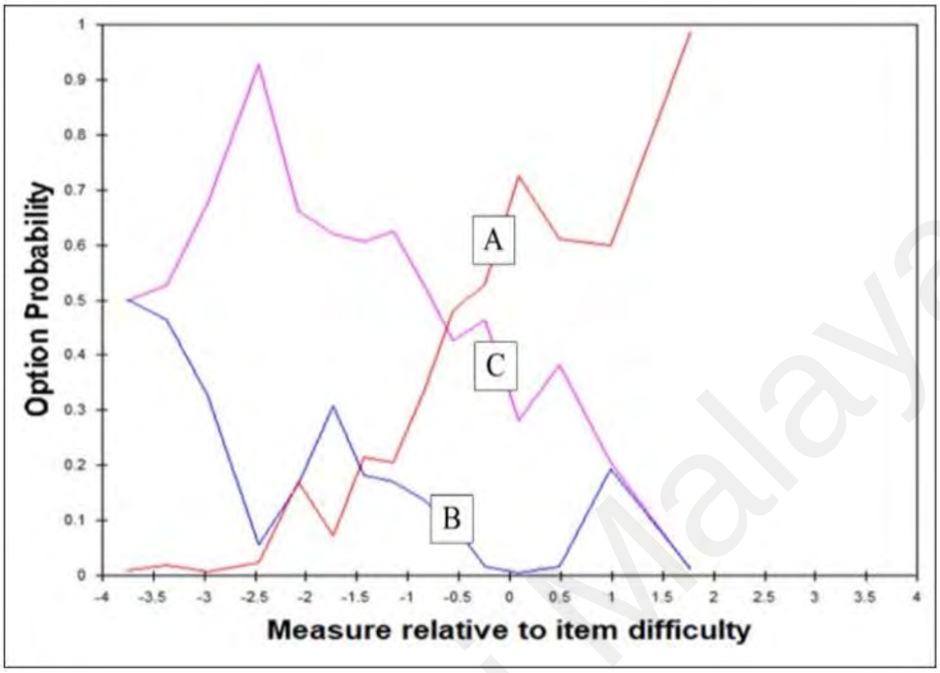


Figure 5.2 IOCC item P1

The next advantage of using IOCC is in the determination of alternative answer, there were some items based on traditional method could not decide one alternative answer. These were alternative answer for item R7, item P11, item P12, item P14, and item P15 because there were a fair number of selected two different options. For this difficulty, IOCC could decide which option could be a stronger alternative answer. For instance, item R7 –the reason for gas produced in the reaction between dilute sulfuric acid and some green copper(II) carbonate powder” with option 1 –oxygen molecules are released from the CO_3^{2-} ions” and option 3 –copper atoms liberate hydrogen from dilute acids” which was selected by 7 participants. Based on IOCC at appendix F, it was clear that the alternative answer was option 3 which has higher probability along the graph compared to option 1. The same case was evident in some items such as item P11, P12, P14 and P15 in which traditional method get difficulties to measure one alternative answer, while IOCC can straight to choose single alternative answer for all the items.

5.3 Implication of Research Findings

The implications of research findings are discussed in these three aspects: a) classroom assessment, b) test developers and c) pre-service chemistry teachers providers.

5.3.1 Classroom Assessment

As stated by the study from Ardiansah, Masykuri, and Rahardjo (2018), the use of the multiple-tiered instrument is very important in the classroom to get a clearer picture of student‘s ability and difficulties. So far, there are many forms of multiple-choice have been introduced to detect misconceptions, and this study offers insights on how the diagnostics power of two-tier multiple-choice questions could be increased. This study reinforces the reasons for utilization 2TMC compared to other forms such as 3TMC and 4TMC. Another vital aspct from this study is the use of the Rasch model to present student‘s data and instrument analysis which can be imitated by other educators. Generally, this study addressed some problems in science education such as 1) student performance is calculated using raw scores 2) test reliability is determined through the calculation of a single statistic (e.g., Cronbach alpha) 3) parametric tests such as t-tests and ANOVAs are conducted using student raw scores (Boone & Scantlebury, 2006; Romine, Schaffer, & Barrow, 2015).

The instrument of this study can be used for future assessment to measure conceptual understanding. The use of Rasch model to estimate validity and reliability of the translated version of the instrument add confidence of its quality to measure (mis)conceptions (Velde, Beaton, Hogg-johnston, Hurwitz, & Tennant, 2009;

Young, Yang, Brazier, & Tsuchiya, 2011). As a result, this move can reduce the scarcity of reliable instrument to measure conceptual understanding in science (Aydeniz et al., 2017; Caleon & Subramaniam, 2010; Hudson & Treagust, 2013; Zubeyde Demet Kirbulut & Geban, 2014). Moreover, as Chemical Concept Inventory (CCI) from Mulford and Robinson (2002) which was widely adopted and validated in other situations (Barbera, 2013; Kruse & Roehrig, 2005; Schwartz & Barbera, 2014), other valid and reliable diagnostics instrument should also be adapted (Schultz et al., 2017).

5.3.2 Test developers

This study has benefitted test developers especially diagnostics test developer because this study informs the importance of using the Rasch model to analyse instrument and the method of data analysis (Boone et al., 2011). It can be said that this study reinforces the use of Rasch model in the development of a multiple-tiered instrument which is initiated by some prior scholars (e.g Romine et al., 2015; Sadhu & Laksono, 2018). Concerning the Rasch model, this study highlights a recommendation of contemplating IOCC to evaluate distractors in MCQs. This findings also highlight the recommendation by Ding and Beichner (2009) to use IOCC as the way of analysing multiple-choice questions.

5.3.3 The pre-service chemistry teachers' program

This study could be useful for universities that have chemistry education program. Using the instrument to detect potential misconceptions, would prompt teacher educators to improve their teaching style and enhance laboratory experiences. Currently, hey may have problems with limited sources, however, they could utilize technology to improve students involvement in any experiments such as: virtual laboratory to enhance chemistry achievement (Tatli & Ayas, 2013), animations for particulate nature of matters (Chang, Quintana, & Krajcik, 2009), multimedia to improve multiple representations (Ardac & Akaygun, 2004), Interactive Multimedia Module with Pedagogical Agents (IMMPA) EC Lab (Tien & Osman, 2017).

This study revealed some difficulties of selected pre-service chemistry teachers to master materials related to multiple representations system and chemical reactions which are very crucial in obtaining good content knowledge (Kahveci, 2009). This is important because content knowledge has strong effects to the confidence and teaching ability in chemistry (Adadan & Yakmaci-guzel, 2013; Kind, 2014; Kirschner et al., 2016; Kruse & Roehrig, 2005). Therefore, the findings and methods designed in this study can be a recommendation for the need for periodic revision in pre-service teacher education (Kahveci, 2009; Kind, 2017).

5.4 Contribution of the study

This study found that the use of item option characteristics curve (IOCC) in misconception studies was a novel analysis of two-tier multiple-choice questions. By using IOCC, it replaced the uses of three-tier multiple-choice questions, and four-tier multiple-choice questions in assessment practices. Moreover, using IOCC, revealed a more detailed analysis of student‘s conception. This in turn would provide chemistry teachers with a method to analyze the misconceptions and better develop instructional interventions to alert students to these misconceptions.

5.5 Conclusion

This study found that pre-service chemistry teachers in Indonesia hold misconceptions about chemical reactions. Their misconception is the influence of the involvement of three level of representation in chemistry. In general, pre-service chemistry teachers, experience misconceptions comprise of burning of magnesium, reactions of dilute hydrochloric acid and grey iron powder, dilute sulfuric acid and green copper(II) carbonate, strong acids and strong alkalis, lead(II) nitrate and potassium iodides, displacement reaction. However, their conception from first year to the third year seems improved as the effects of quantitatively more advanced and deeper subjects they learn in the university.

From this study which used traditional method and IOCC, it was found that each item in either phenomenon tier or reasoning tier had a distractor that could interfere with the selected option of pre-service chemistry teachers and the phenomenon indicated misconceptions. It showed that under ability equal to logit zero, respondents had high potency to be deflected by a distractor which revealed misconceptions. Distractor analysis by item option characteristics curve (IOOC) also revealed some unexpected curves after 0 logits, less-functioned distractors, and the inconsistency of the strongest distractor. The finding of this suggested the use of IOCC to analyse student‘s abilities.

5.6 Suggestion for future research

In the area of misconception analysis based on a quantitative study with survey design, consideration of the sample is also beneficial to find more meaningful data. There are some variables that need further consideration such as gender (male-female) (Hudson & Treagust, 2013), attitude toward science, and motivation (low, moderate, high) in the comparison of misconception number (Çam, Topçu, & Sülün, 2015). Gender in many studies can influence science achievement (Acar, Türkmen, & Bilgin, 2015). Female students perform better in some countries such as Caribbean Islands and India (Kutnick, 2000; Larson, Stephen, Bonitz, & Wu, 2014), while male students exceed female students in Hongkong and Canada (Adamuti-Trache & Sweet, 2013; Sun, Bradley, & Akers, 2012). Another factor influence science achievement is the attitude toward science which positively correlated (Yetisir, 2014). However, attitude toward chemistry is not statistically significant influencing misconceptions (Çam et al., 2015).

The use of IOCC in the literature mostly have a large sample size, therefore it is vital to conduct a study to see the feasibility of using IOCC in a smaller sample size or less than 40 students. So far, the study from Hermann-Abel et al (2011), Wind and Gale (2015) have samples of more than 400 samples, while this study is only 185 students. As a result, if IOCC can be utilized, it can be used by researchers and educators to analyse multiple choice questions data to improve science learning.

REFERENCES

Abdullah, N., Noranee, S., & Khamis, M. R. (2017). The use of rasch wright map in assessing conceptual understanding of electricity. *Social Sciences & Humanities*, *25*, 81–92.

Abraham, M. R., Grzybowski, E. B., Renner, J. W., & Marek, E. A. (1992). Understandings and misunderstandings of eighth graders of five chemistry concepts found in textbooks. *Journal of Research in Science Teaching*, *29*, 105–120.

Acar, Ö., Türkmen, L., & Bilgin, A. (2015). Examination of gender differences on cognitive and motivational factors that influence 8 th graders ' science achievement in Turkey. *Eurasia Journal of Mathematics, Science & Technology Education*, *11*(5), 1027–1040. https://doi.org/10.12973/eurasia.2015.1372

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

Adadan, E., & Yakmaci-guzel, B. (2013). Use of multiple representations in developing preservice chemistry teachers ' understanding of the structure of matter. *International Journal of Enviromental & Science Education (IJESE)*, *8*(1), 109–130.

Adamuti-Trache, M., & Sweet, R. (2013). Academic effort and achievement in science: Beyond a gendered relationship. *Research in Science Education*, *43*, 2367-2385.

Afif, N. F., Nugraha, M. G., & Samsudin, A. (2017). Developing energy and momentum conceptual survey (EMCS) with four-tier diagnostic test items. *AIP Conference Proceedings*, *1848*(May). https://doi.org/10.1063/1.4983966

Agung, S., & Schwartz, M. S. (2007). Students' understanding of conservation of matter , stoichiometry and balancing equations in Indonesia. *International Journal of Science Education ISSN:*, *29*(13), 1679–1702. https://doi.org/10.1080/09500690601089927

Agustin, A., Supardi, K. I., & Sunarto, W. (2018). Pengaruh penggunaan peta konsep berbasis multilevel terhadap pemahaman konsep kimia siswa. *Chemistry in Education*, *7*(2), 8–13.

Ainiyah, M., Ibrahim, M., & Hidayat, M. T. (2018). The profile of student misconceptions on the human and plant transport systems. *Journal of Physics: Conference Series*, *947*(1), 0–6. https://doi.org/10.1088/1742-6596/947/1/012064

Akkus, H., Kadayifci, H., & Atasoy, B. (2011). Development and application of a two-tier diagnostic test to assess secondary students' understanding of chemical equilibrium concepts. *Journal of Baltic Science Education*, *10*(3), 146–155.

Alharbi, N. N. S., Treagust, D. F., Chandrasegaran, A. L., Won, M., Alharbi, N. N.

S., Treagust, D. F., & Chandrasegaran, A. L. (2015). Influence of particle theory conceptions on pre-service science teachers ' understanding of osmosis and diffusion in fl uence of particle theory conceptions on pre-service science teachers ' understanding of osmosis and diffusion. *Journal of Biological Education*, *9266*, 1–14. https://doi.org/10.1080/00219266.2014.923488

Ammase, A., Siahaan, P., & Fitriani, A. (2019). Identification of junior high school students' misconceptions on solid matter and pressure liquid substances with four tier test. *Journal of Physics: Conference Series*, *1157*(2), 0–6. https://doi.org/10.1088/1742-6596/1157/2/022034

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*(10), 952–978. https://doi.org/10.1002/tea.10053

Andrich, D., Marais, I., & Humphry, S. (2012). Using a theorem by andersen and the dichotomous rasch model to assess the presence of random guessing in multiple choice items. *Journal of Educational and Behavioral Statistics*, *37*(3), 417–442. https://doi.org/10.3102/1076998611411914

Anggrayni, S., & Ermawati, F. U. (2019). The validity of Four-Tier's misconception diagnostic test for Work and Energy concepts. *Journal of Physics: Conference Series*, *1171*(1), 0–13. https://doi.org/10.1088/1742-6596/1171/1/012037

Ardac, D., & Akaygun, S. (2004). Effectiveness of multimedia based on instruction that emphasizes on students' understanding of chemical change. *Journal of Research in Science Teaching*, *41*(4), 317–337.

Ardiansah, Masykuri, M., & Rahardjo, S. B. (2018). Senior high school students' need analysis of three-tier multiple choice (3TMC) diagnostic test about acid-base and solubility equilibrium. *Journal of Physics: Conference Series*, 1–8. https://doi.org/10.1088/1742-6596/1022/1/012033

Aretz, S., Borowski, A., & Schmeling, S. (2012). The role of confidence in ordered multiple-choice items about the universe ' s expansion. *ESERA 2017 Conference Dublin City University, Ireland*, (2006), 3–5.

Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, *34*(11), 1667–1686. https://doi.org/10.1080/09500693.2012.680618

Artdej, R., Ratanaroutai, T., Coll, R. K., & Thongpanchang, T. (2010). Thai grade 11 students' alternative conceptions for acid-base chemistry. *Research in Science & Technological Education*, *28*(2), 167–183.

Asri, Y. N., Rusdiana, D., & Feranie, S. (2017). ICARE Model Integrated with Science Magic to Improvement of Students' Cognitive Competence In Heat and Temperature Subject. *Advances in Social Science, Education and Humanities Research (ASSEHR)*, *57*, 137–139. https://doi.org/10.2991/icmsed-16.2017.30

Aydeniz, M., Bilican, K., & Kirbulut, Z. D. (2017). Exploring pre-service chemistry teachers' conceptual understanding of particulate nature of matter through three-tier diagnostics test. *International Journal of Education in Mathematics, Science and Technology*, *5*(3), 221–234.

Azizoğlu, N., Alkan, M., & Geban, Ö. (2006). Undergraduate pre–service teachers ‘ understandings and misconceptions of phase equilibrium. *Journal of Chemical Education*, *83*(6), 947–953.

Bain, K. (2017). Students‘ understanding of energy in chemical reactions and processes. *Unpublished Doctoral Dissertation, West Lafayatte: Purdue University*.

Bain, K., & Towns, M. H. (2018). Investigation of undergraduate and graduate chemistry students‘ understanding of thermodynamic driving forces in chemical reactions and dissolution. *Journal of Chemical Education*, *95*(4), 512–520. https://doi.org/10.1021/acs.jchemed.7b00874

Bak Kibar, Z., Yaman, F., & Ayas, A. (2013). Assessing prospective chemistry teachers‘ understanding of gases through qualitative and quantitative analyses of their concept maps. *Chemistry Education Research and Practice*, *14*(4), 542–554. https://doi.org/10.1039/c3rp00052d

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

Bardar, E. M., Prather, E. E., Brecher, K., & Slater, T. F. (2006). Development and validation of the light and spectroscopy concept inventory. *Astronomy Education Review*, *5*(2), 103–113. https://doi.org/10.3847/AER2006020

Baser, M., & Geban, Ö. (2007). Effectiveness of conceptual change instruction on understanding of heat and temperature concepts. *Research in Science & Technological Education*, *25*(1), 115–133. https://doi.org/10.1080/02635140601053690

Bayraktar, S. (2009). Misconceptions of Turkish pre-service teachers about force and motion. *International Journal of Science and Mathematics Education*, *7*(2), 273–291. https://doi.org/10.1007/s10763-007-9120-9

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, *62*(8), 750–762. https://doi.org/10.1119/1.17449

Berek, F. X., Sutopo, S., & Munzil, M. (2016). Concept enhancement of junior high school students in hydrostatic pressure and archimedes law by predict-observe-explain strategy. *Jurnal Pendidikan IPA Indonesia*, *5*(2), 230–238. https://doi.org/10.15294/jpii.v5i2.6038

Bilgin, İ. (2006). Promoting pre-service elementary students‘ understanding of chemical equilibrium through discussions in small groups. *International Journal of Science and Mathematics Education*, *4*, 467–484. https://doi.org/10.1007/s10763-005-9015-6

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa*, *80*(2), 139 – 148. Retrieved from http://www.pdkintl.org/kappan/kbla9810.htm

Bond, T. ., & Fox, C. M. (2015). *Applying the Rasch model fundamental measurement in the human sciences* (3rd ed.). New York, NY, US: Routledge Taylor & Francis Group.

Boone, W.J, Staver, J. ., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer Netherlands.

Boone, William J., & Scantlebury, K. (2006). The role of rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253–269. https://doi.org/10.1002/sce.20106

Boone, William J, Townsend, J. S., & Staver, J. (2011). Using Rasch Theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts : An Exemplar utilizing STEBI Self-Efficacy Data. *Science Education*, 258–280. https://doi.org/10.1002/sce.20413

Briggs, D., Alonzo, A., Schwab, C., & Wilson, M. (2006). *Diagnostic assessment with ordered multiple-choice items*. *Educational Assessment*. https://doi.org/10.1207/s15326977ea1101_2

Bucat, R. (2004). Pedagogical content knowledge as a way forward: applied research in chemistry education. *Chemistry and Education: Research and Practice*, *5*(3), 215–228. https://doi.org/10.1039/b4rp90025a

Butler, J., Mooney Simmie, G., & O'Grady, A. (2015). An investigation into the prevalence of ecological misconceptions in upper secondary students and implications for pre-service teacher education. *European Journal of Teacher Education*, *38*(3), 300–319. https://doi.org/10.1080/02619768.2014.943394

Cahya, A., & Sanjaya, I. G. M. (2015). The development of three tier diagnostic test to identify student misconception in chemical bonding on 10 th grader. *UNESA Journal of Chemical Education*, *4*(3), 456–465.

Caleon, I. S., & Subramaniam, R. (2010). Do students know What they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4

Çalik, M., Ayas, A., & Coll, R. K. (2007). Enhancing pre-service elementary teachers' conceptual understanding of solution chemistry with conceptual change text. *International Journal of Science and Mathematics Education*, *5*(1), 1–28. https://doi.org/10.1007/s10763-005-9016-5

Callegaro Borsa, J., Figueiredo, B., Denise, D., & Bandeira, R. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia*, *22*(53), 423–432. https://doi.org/10.1590/1982-43272253201314

Çam, A., Topçu, M. S., & Sülün, Y. (2015). Preservice science teachers ' attitudes towards chemistry and misconceptions about chemical kinetics. *Asia-Pacific Forum on Science Learning and Teaching*, *16*(2), 12–14.

Canpolat, N. (2006). Turkish undergraduates' misconceptions of evaporation, evaporation rate, and vapour pressure. *International Journal of Science Education*, *28*(15), 1757–1770. https://doi.org/10.1080/09500690600779957

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures. *Clinical Therapeutics*, *36*(5), 648–662.

Cataloglu, E., & Robinett, R. W. (2002). Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career. *American Journal of Physics*, *70*(3), 238–251. https://doi.org/10.1119/1.1405509

Çelikler, D., & Aksan, Z. (2014). Determination of knowledge and misconceptions of pre-service elementary science teachers about the greenhouse effect by drawing. *Procedia - Social and Behavioral Sciences*, *136*, 452–456. https://doi.org/10.1016/j.sbspro.2014.05.355

Cetin-Dindar, A., & Geban, O. (2011). Development of a three-tier test to assess high school students‘ understanding of acids and bases. *Procedia - Social and Behavioral Sciences*, *15*, 600–604. https://doi.org/10.1016/j.sbspro.2011.03.147

Çetin, P. S., Kaya, E., & Geban, Ö. (2009). Facilitating conceptual change in gases concepts. *Journal of Science Education and Technology*, *18*(2), 130–137. https://doi.org/10.1007/s10956-008-9138-y

Chan, S. W., Ismail, Z., & Sumintono, B. (2015). The impact of statistical reasoning learning environment: A rasch analysis. *Advanced Science Letters*, *21*(5), 1211–1215. https://doi.org/10.1166/asl.2015.6077

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students‘ ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, *8*(3), 293. https://doi.org/10.1039/b7rp90006f

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

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2009). Emphasizing multiple levels of representation to enhance students‘ understanding of the changes occuring during chemical reaction. *Journal of Chemical Education*, *86*(12), 1433–1436.

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2011). Facilitating high school students‘ use of multiple representations to describe and explain simple chemical reactions. *Teaching Science*, *57*(4), 13–19.

Chang, Hsin-yi, Quintana, C., & Krajcik, J. S. (2009). The impact of designing and evaluating molecular animations on how well middle school students understand the particulate nature of matter. *Science Education*, 73–94. https://doi.org/10.1002/sce.20352

Chang, Huey-Por, Chen, J., Guo, C., Chen, C., Chang, C., Lin, S., … Tseng, Y. (2007). Investigating primary and secondary students‘ learning of physics concepts in Taiwan. *International Journal of Science Education*, *29*(4), 465–

482. https://doi.org/10.1080/09500690601073210

Chen, C. H. H. I. H., Lin, H. U. H., & Lin, M. I. N. G. I. (2003). Developing a two-tier diagnostic instrument to assess high school students ' understanding − the formation of images by a plane mirror. *Proceedings of the National Science Council*, *12*(3), 106–121.

Cheong, I. P., Treagust, D., Kyeleve, I. J., & Oh, P. (2010). Evaluation of students' conceptual understanding of malaria. *International Journal of Science Education*, *32*(18), 2497–2519. https://doi.org/10.1080/09500691003718014

Cheung, K. L., & Yang, D. (2018). Examining the differences of Hong Kong and Taiwan students ' performance on the number sense three-tier test. *Eurasia Journal of Mathematics, Science & Technology Education*, *14*(7), 3329–3345.

Chiu, M. (2001). Algorithmic problem solving and conceptual understanding of chemistry by students at a local high school in Taiwan. *Proceedings of the National Science Council, Republic of China*, *11*(1), 20–38.

Chiu, M. H. (2007). A national survey of students' conceptions of chemistry in Taiwan. *International Journal of Science Education*, *29*(4), 421–452.

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

Chong, S. H., & Goolamally, N. (2019). Validation of a two-tier multiple choice (2TMC) diagnostic instrument on the mole concept and solution concentration: A Rasch analysis. *Proceeding International Conference on Education*, 244–257.

Chu, H.-E., Treagust, D. F., & Lim, G. S. E. Chandrasegaran, A. L. (2015). Efficacy of multiple choice items: Do two-tier multiple-choice diagnostic items have the power to measure students' conceptions similar to open-ended items? *Paper Presented at the 11th European Science Education Research Association (ESERA) Conference, Helsinki, Finland.*

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

Chua, Y. P. (2013). *Mastering research statistics*. McGraw-Hill Education.

Cortina, J. (1993). What is coefficient alpha: an examination of theory and applications. *Journal of Applied Psychology*, *78*, 98–104.

Coştu, B, Ayas, A., Niaz, M., Ünal, S., & Çalık, M. (2007). Facilitating conceptual change in students' understanding of boiling concept. *Journal of Science Educational Technology*, *16*, 524–536.

Coştu, Bayram, Ayas, A., Niaz, M., Ünal, S., & Çalik, M. (2007). Facilitating conceptual change in students' understanding of boiling concept. *Journal of Science Education and Technology*, *16*(6), 524–536. https://doi.org/10.1007/s10956-007-9079-x

Cox, M., Steegen, A., & De Cock, M. (2016). How aware are teachers of students ' misconceptions in astronomy? A qualitative analysis in Belgium. *Science Education Internationl*, *27*(2), 277–300.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

Dann, R. (2014). Assessment *as* learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 149–166. https://doi.org/10.1080/0969594X.2014.898128

Debacker, T. K., Crowson, H. M., Beesley, A. D., Thoma, S. J., & Hestevold, N. L. (2008). The challenge of measuring epistemic beliefs : An analysis of three self-report instruments. *The Journal of Experimental Education*, *76*(3), 281–312.

Delgado-Rico, E., Carrctero-Dios, H., & Rueh, W. (2012). Content validity evidences in test development: An applied perspective. *Imernalonal Journal of Clinical and Health Psychology*, *12*(3), 449–460.

Desi, D. (2013). Identifikasi miskonsepsi mahasiswa program studi pendidikan kimia pada pokok bahasan stoikiometri mata kuliah kimia dasar I. *Jurusan Pendidikan MIPA FKIP Unsri. (Unpublished)*.

Desstya, A., Prasetyo, Z. K., Suyanta, S., Susila, I., & Irwanto, I. (2019). Developing an instrument to detect science misconception of an elementary school teacher. *International Journal of Instruction*, *12*(3), 201–218.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Newbury Park, CA: Sage Publications.

DeVellis, Robert F. (2006). Classical test theory. *Medical Care*, *44*(11), 50–59.

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, *5*, 1–17. https://doi.org/10.1103/PhysRevSTPER.5.020103

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research*, *2*(1). https://doi.org/10.1103/PhysRevSTPER.2.010105

Doige, C. A., & Day, T. (2012). A typology of undergraduate textbook definitions of ‗Heat' across science disciplines. *International Journal of Science Education*, *34*(5), 677–700. https://doi.org/10.1080/09500693.2011.644820

Downing, S. (1992). True-false, alternate-choice, and multiple-choice items. *Educational Measurement: Issues and Practice*, *11*(3), 27–30.

Drogemuller, R. (1994). Evaluation of a two-tier multiple choice test assessing prior concepts of Photosynthesis. *Unpublished Master Thesis, University of South Australia*.

Duit, R., & Treagust, D. F. (2003). Conceptual change - A powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*(6), 671–688. https://doi.org/10.1017/CBO9781107415324.004

Eilks, Ingo, Witteck, Torsten, Pietzner, & Verena. (2012). The role and potential dangers of visualisation when learning about sub-microscopic explanations in chemistry education. *Centre Dor Educational Policy Studies*, *2*(1), 125–145. Retrieved from http://www.pef.uni-lj.si

Elliot, A. J., & McGregor, H. a. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology*. https://doi.org/10.1037//0022-3514.80.3.501

Enawaty, E., & Putra Sartika, R. (2015). Description of students' misconception in chemical bonding. *Proceeding of International Conference On Research, Implementation And Education Of Mathematics And Sciences*, (May), 17–19.

Engelhardt, P. V, & Beichner, R. J. (2004). Students' understanding of direct current resistive electric circuits. *American Journal of Physics*, *72*(1), 98–115.

Epstein, R. M. (2007). Assessment in medical education. *The New England Journal of Medicine*, *356*, 387–396.

Ergin, S. (2016). The Effect of Group Work on Misconceptions of 9th Grade Students about Newton's Laws. *Journal of Education and Training Studies*, *4*(6), 127–136. https://doi.org/10.11114/jets.v4i6.1390

Erman, E. (2016). Factors contributing to students ' misconceptions in learning covalent bonds. https://doi.org/10.1002/tea.21375

Eryilmaz, A. (2010). Development and Application of Three-Tier Heat ans Temperature Test_Sample of Bachelor and Graduate Students.pdf. *Egitim Arastirmalari - Eurasian Journal of Educational Research*, (40), 53–76.

Eshach, H. (2014). Development of a student-centered instrument to assess middle school students conceptual understanding of sound. *Physical Review Special Topics - Physics Education Research*, *10*(1), 1–14. https://doi.org/10.1103/PhysRevSTPER.10.010102

Eshach, H., Lin, T. C., & Tsai, C. C. (2018). Misconception of sound and conceptual change: A cross-sectional study on students' materialistic thinking of sound. *Journal of Research in Science Teaching*, *55*(5), 664–684. https://doi.org/10.1002/tea.21435

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, *58*(3), 357–381.

Farida, I., & Liliasari, L. (2011). Pengembangan kemampuan interkoneksi multipel representasi mahasiswa calon guru kimia melalui sistem manajemen belajar berbasi web. *Simposium Puslitjaknov, Denpasar Bali*.

Fariyani, Q., Rusilowati, A., & Sugianto, S. (2017). Four-tier diagnostic test to identify misconceptions in geometrical optics. *Unnes Science Education Journal*, *6*(3), 1724–1729. https://doi.org/10.15294/usej.v6i3.20396

Fetherstonhaugh, T., & Treagust, D. F. (1992). Students‘ understanding of light and its properties: Teaching to engender conceptual change. *Science Education*, *76*(6), 653–672. https://doi.org/10.1002/sce.3730760606

Field, A. P. (2009). *Discovering statistics using SPSS* (3th ed.). Los Angeles, California: Sage Publications.

Flórez, M., & Sammons, P. (2013). *The Assessment for learning strategy*. *CfBT Education Trust*. Retrieved from http://cdn.cfbt.com/~/media/cfbtcorporate/files/research/2013/r-assessment-for-learning-2013.pdf%5Cnhttp://eric.ed.gov/?id=ED546817

Fraenkel, J R, & Wallen, N. E. (2009). *How to design and evaluate research in education. Boston: McGraw-Hill.* (7th ed.). New York, NY: McGraw-Hill.

Fraenkel, Jack R, Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York: McGraw-Hill.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Fulmer, G. W., Chu, H.-E., Treagust, D. F., & Neumann, K. (2015). Is it harder to know or to reason? Analyzing two-tier science assessment items using the Rasch measurement model. *Asia-Pacific Science Education*, *1*(1), 1. https://doi.org/10.1186/s41029-015-0005-x

Furtak, E. M., Morrison, D., & Kroog, H. (2014). Investigating the link between learning progressions and classroom assessment. *Science Education*, *98*(4), 640–673. https://doi.org/10.1002/sce.21122

Gabel, D. (1999). Improving teaching and learning through chemistry education research: A look to the future. *Journal of Chemical Education*, *76*(4), 548–554.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2009). *Educational research: Competencies for analysis and applications* (9th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). *Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. Review of Educational Research* (Vol. XX). https://doi.org/10.3102/0034654317726529

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education : A review of the literature. *Computers & Education*, *57*(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004

Given, L. M. (2008). *The Sage encyclopedia of qualitative research methods*. Los Angeles, Calif.: Sage Publications.

Gjersing, L., Caplehorn, J. R., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology*, *10*. https://doi.org/10.1186/1471-2288-10-13

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159–1176. https://doi.org/10.1002/tea.20442

Gracia, S. (2005). Analyzing CSR implementation with the Rasch Model. *Faculty Publication, Rhode Island College*.

Green, S., Lissitz, R., & Mulaik, S. (1977). Limitations of coefficient alpha as an index of test unidimensionlity. *Educational Psychological Measurement.*, *37*, 827–838.

Greenwood, B. K. R. (2017). The development of a two-tiered diagnostic instrument to assess acid / base misconceptions in organic chemistry students. *Unpublished Master Thesis, Arizona State University*.

Griffard, P. B., & Wandersee, J. H. (2001). The two-tier instrument on photosynthesis: What does it diagnose? *International Journal of Science Education*, *23*(10), 1039–1052. https://doi.org/10.1080/09500690110038549

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, *61*(2), 29–45. https://doi.org/10.1027/1901-2276.61.2.29

Gültepe, N. (2016). What do data mean for pre-service chemistry teachers? *Journal of Education and Training Studies*, *4*(7), 100–110. https://doi.org/10.11114/jets.v4i7.1449

Gurcay, D., & Gulbas, E. (2015). Development of three-tier heat, temperature and internal energy diagnostic test. *Research in Science & Technological Education*, 1–21. https://doi.org/10.1080/02635143.2015.1018154

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students‘ misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

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

Haidar, A. H. (1997). Prospective chemistry teachers‘ conceptions of the conservation of matter and related concepts. *Journal of Research in Science Teaching*, *34*(2), 181–197. https://doi.org/10.1002/(SICI)1098-2736(199702)34:2<181::AID-TEA5>3.0.CO;2-P

Hair, J. F., Hult, G. T., Ringle, C. M., & Sarstedt, M. (2014). *A primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks, CA.: Sage Publications.

Haladyna, T. M. (2016). *Item analysis for selected-response test items In S. Lane, M. Raymond, & T. Haladyna (Eds.), Handbook of test development (2nd ed., pp. 392– 409)*. New York, NY: Routledge.

Haladyna, T. M, & Rodriguez, M. C. (2013). *Developing and validating test items New*. New York, NY: Routledge.

Haladyna, Thomas M, Downing, S. M., & Rodriguez, C. (2002). A review of multiple- choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309–333. https://doi.org/10.1207/S15324818AME1503

Hamid, Y., & Mahmood, S. (2010). Understanding constructive feedback: A commitment between teachers and students for academic and professional development. *Journal of the Pakistan Medical Association*, *60*(3), 224–227.

Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for educational research. *American Journal of Physics*, *64*(10), 1316-1325.

Handhika, J., Cari, C., Suparmi, A., Sunarno, W., & Purwandari, P. (2018). Development of diagnostic test instruments to reveal level student conception in kinematic and dynamics. *Journal of Physics: Conference Series*, *983*(1), 0–6. https://doi.org/10.1088/1742-6596/983/1/012025

Haryono, H. E. (2018). The effectiveness of science student worksheet with cognitive conflict strategies to reduce misconception on heat concept. *Jurnal Pena Sains*, *5*(2), 79. https://doi.org/10.21107/jps.v5i2.4510

Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the Certainty of Response Index (CRI). *Physics Education*, *34*(5), 294–299. https://doi.org/10.1088/0031-9120/34/5/304

Haslam, F., & Treagust, D. F. (1987). Diagnosing secondary students misconceptions about photosynthesis and respiration in plants using a two-tier multiple-choice instrument. *Journal of Biological Education*, *21*(3), 203–211.

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

Helmi, H., Rustaman, N. Y., Tapilow, F. S., & Hidayat, T. (2019). Preconception analysis of evolution on pre-service biology teachers using certainty of response index. *Journal of Physics: Conference Series*, *1157*(2), 0–6. https://doi.org/10.1088/1742-6596/1157/2/022033

Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A., Isjoni, I., Johan, H., … Safitri, D. (2017). Constructing and implementing a four tier test about static electricity to diagnose pre-service elementary school teacher‘ misconceptions. *Journal of Physics: Conference Series*, *895*, 1–6. https://doi.org/10.1088/1742-6596/895/1/012167

Herrmann-abell, C. F., & Deboer, G. E. (2016). Using rasch modeling and option probability curves to diagnose students‘ misconceptions. *Paper Presented at the 2016 AERA Annual Meeting Washington, DC*, 1–12.

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

Hestenes, D, Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141–158.

Hestenes, David, & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, *30*(3), 159–166. https://doi.org/10.1119/1.2343498

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

Howitt, S. T., Anderson, M., Hamilton, S., & Wright, T. (2008). A concept inventory for molecular life sciences: How will it help your teaching practice. *Australian Biochemist*, *39*(3), 14–17. https://doi.org/10.1109/LSP.2017.2714578

Hrast, Š., & Savec, F. V. (2017). The integration of submicroscopic representations used in chemistry textbook sets into curriculum topics. *Acta Chimica Slovenica*, 959–967. https://doi.org/10.17344/acsi.2017.3657

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons Inc.

Hudson, R. D., & Treagust, D. F. (2013). Which form of assessment provides the best information about student performance in chemistry examinations ? *Research in Science & Technological Education*, *31*(1), 49–65. https://doi.org/10.1080/02635143.2013.764516

Ikenna, I. A. (2015). Remedying students‘ misconceptions in learning of chemical bonding and spontaneity through intervention discussion learning model (IDLM). *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, *8*(10), 3251–3254.

Imaduddin, M. (2018). Analisis miskonsepsi submikroskopik konsep larutan pada calon guru kimia. *EduSains: Jurnal Pendidikan Sains & Matematika*, *6*(2), 1–12.

Iqbal, M. H. (2016). The development and validation of a one tier diagnostic assessment to test premedical students‘ misconceptions about traumatic brain injury. *Wright State University Theses*.

Irsyad, M., & Linuwih, S. (2018). Learning cycle 7e model-based multiple representation to reduce misconseption of the student on heat theme. *Journal of Innovative Science Education*, *7*(1), 45–52.

Jaffar, S. M., & Dindyal, J. (2011). Language-related misconceptions in the study of limits. *AAMT -MERGA Conference Paper*, (July), 390–397. Retrieved from https://www.merga.net.au/documents/RP_JAFFAR&DINDYAL_MERGA34-AAMT.pdf

Jang, N. H. (2003). Developing and validating a chemical bonding instrument for korean high school students. *Unpublished Doctoral Dissertation, University of Missouri-Columbia*, 165. https://doi.org/10.1016/j.jorganchem.2010.09.058

Johnstone, A. H. (2000). Teaching of chemistry - logical or psychological? *Chem. Educ. Res. Pract.*, *1*(1), 9–15. https://doi.org/10.1039/A9RP90001B

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2009). *Data analysis: A model comparison approach* (2nd ed.). New York, NY, US: Routledge/Taylor & Francis Group.

Kahveci, A. (2009). Exploring chemistry teacher candidates‘ profile characteristics, teaching attitudes and beliefs, and chemistry conceptions. *Chemistry Education Research and Practice*, *10*, 109–120.

Kajander, A., & Lovric, M. (2009). Mathematics textbooks and their potential role in supporting misconceptions. *International Journal of Mathematical Education in Science and Technology*, *40*(2), 173–181.

Kalas, P., O‘Neill, A., Pollock, C., & Birol, G. (2013). Development of a meiosis concept inventory. *CBE Life Sciences Education*, *12*(4), 655–664. https://doi.org/10.1187/cbe.12-10-0174

Kalin, B., & Arikil, G. (2010). Misconceptions possessed by undergraduate students about the topic ―solutions .‖ *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *4*(2), 177–206.

Kaltacki, D. (2012). Development and application of a four-tier test to assess pre-service physics teachers‟ misconceptions about geometrical optics. *Published Dissertation Middle East Technical University*.

Kaltakci-gurel, D., Eryilmaz, A., & Mcdermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers ‘ misconceptions about geometrical optics. *Research in Science & Technological Education*, *5143*(April), 238–260. https://doi.org/10.1080/02635143.2017.1310094

Kaltakçi, D., & Didiç, N. (2007). Identification of pre-service physics teachers‘ misconceptions on gravity concept: A study with a 3-tier misconception test. *AIP Conference Proceedings*, 499–500. https://doi.org/10.1063/1.2733255

Kaltakci, D., & Eryilmaz, A. (2010). Identifying pre-service physics teachers‘ misconceptions with three-tier tests. *GIREP-ICPE-MPTL Conference 2010 in Reims, France.*, 1–8. Retrieved from http://www.univ-reims.fr/site/evenement/girep-icpe-mptl-2010-reims-international-conference/gallery_files/site/1/90/4401/22908/29321/29329.pdf

Kanli, U. (2015). Using a two-tier test to analyse students‘ and teachers‘ alternative concepts in astronomy. *Science Education International*, *26*(2), 148–165.

Kanli, Uygar. (2014). A study on identifying the misconceptions of pre-service and in-service teachers about basic astronomy concepts. *Eurasia Journal of Mathematics, Science & Technology Education*, *10*(5), 471–479. https://doi.org/10.12973/eurasia.2014.1120a

Karsli, F., & Çalik, M. (2012). Can freshman science student teachers‘ alternative conceptions of ―electrochemical cells‖ be fully diminished? *Asian Journal of Chemistry*, *24*(2), 485–491.

Kartiko, D. C. (2018). Revealing physical education students‘ misconception in sport biomechanics. *Journal of Physics: Conference Series*, *1006*(1), 0–6. https://doi.org/10.1088/1742-6596/1006/1/012040

Keig, P. ., & Rubba, P. . (1993). Translations of the representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge. *Journal of Research in Science Teaching*, *30*, 883–903.

Kind, V. (2004). Beyond ppearances: Students‘ misconceptions about basic chemical ideas (2nd ed.). Durham: Durham University.

Kind, V. (2014). A Degree is not enough : A quantitative study of aspects of pre-service science teachers ‘ chemistry content knowledge A Degree Is Not Enough : A quantitative study of aspects of pre-service science teachers ‘ chemistry content knowledge. *International Journal of Science Education*, *0*(0), 1–33. https://doi.org/10.1080/09500693.2013.860497

Kind, V. (2017). Development of evidence-based , student-learning- oriented rubrics for pre-service science teachers ‘ pedagogical content knowledge knowledge. *International Journal of Science Education*, 1–34. https://doi.org/10.1080/09500693.2017.1311049

King, K. V, Gardner, D. a, Zucker, S., & Jorgensen, M. a. (2004). The distractor rationale taxonomy : enhancing multiple-choice items in reading and mathematics. *Learning*, (July 2004), 1–14. Retrieved from http://www.pearsonassessments.com/NR/rdonlyres/D7E62EC6-CC3F-47B6-B1CB-A83F341AD768/0/Distractor_Rationales.pdf

Kirbulut, Z. D, & Geban, O. (2014). Using three-tier test to assess students‘ misconceptions of states of matter. *Eurasia Journal of Mathematics, Science & Technology Education*, *10*(5), 509–521.

Kirbulut, Zubeyde Demet, & Geban, O. (2014). Using three-tier diagnostic test to assess students ‘ misconceptions of states of matter. *Eurasia Journal of Mathematics, Science & Technology Education*, *10*(5), 509–521. https://doi.org/10.12973/eurasia.2014.1128a

Kirschner, S., Borowski, A., Fischer, H. E., Kirschner, S., Borowski, A., Fischer, H. E., … Aufschnaiter, C. Von. (2016). physics teachers ‘ pedagogical content knowledge Developing and evaluating a paper-and-pencil test to assess components of physics teachers ‘ pedagogical content knowledge. *International Journal of Science Education*, (June), 1–30. https://doi.org/10.1080/09500693.2016.1190479

Kılıç, D., & Sağlam, N. (2009). Development of a two-tier diagnostic test to determine students??? understanding of concepts in genetics. *Eurasian Journal of Educational Research*, (36), 227–228.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences*, *17*, 161–173.

Klymkowsky, M. W., & Garvin-Doxas, K. (2008). Recognizing student misconceptions through Ed‘s tools and the biology concept inventory. *PLoS Biology*, *6*(1), 0014–0017. https://doi.org/10.1371/journal.pbio.0060003

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*, 298–310. https://doi.org/10.1187/05

Koriat, A., S., Lichtenstein, L., & Fischhof, B. (1980). Reasons for confidence. *Journal of Experimental Psychology*, *6*(2), 107–118.

Korur, F. (2015). Exploring seventh-grade students‘ and pre-service science teachers‘ misconceptions in astronomical concepts. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 1041–1060. https://doi.org/10.12973/eurasia.2015.1373a

Köse, S. (2008). Diagnosing student misconceptions: Using drawings as a research method. *World Applied Sciences Journal*, *3*(2), 283–293. Retrieved from http://idosi.org/wasj/wasj3(2)/20.pdf

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching*, *34*(9), 949–968. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1098-2736(199711)34:9%3C949::AID-TEA7%3E3.0.CO;2-U/epdf

Kozma, R., & Russell, J. (2005). *Students becoming chemists: Developing representational competence*. In Gilbert, J. K. (Ed.), Visualization in science education (pp. 121–146). Dordrecht, Netherlands: Springer.

Kruse, R. A., & Roehrig, G. H. (2005). A comparison study: Assessing teachers‘ conceptions with the chemistry concepts inventory. *Journal of Chemical Education*, *10*, 1246–1250.

Kuhn, D. (2000). Metacognitive development. *Current Directions In Psychological Science*, *9*(5), 178–181. https://doi.org/10.1111/1467-8721.00088

Kumpha, P., Suwannoi, P., & Treagust, D. F. (2014). Thai grade 10 students conceptual understanding of chemical bonding. *Procedia - Social and Behavioral Sciences*, *143*, 657–662. https://doi.org/10.1016/j.sbspro.2014.07.458

Kusairi, S., Alfad, H., & Zulaikah, S. (2017). Development of web-based intelligent tutoring (iTutor) to help students learn fluid statics. *Journal of Turkish Science Education*, *14*(2), 1–11. https://doi.org/10.12973/tused.10194a

Kutluay, Y. (2005). Diagnosis of eleventh grade students‘ misconceptions about geometric optic by a three-tier test. *Unpublished Master Thesis, Middle East Technical University*.

Kutnick, P. (2000). Girls, boys and school achievement: Critical comments on who achieves in schools and under what economic and social conditions achievement takes place-a Caribbean perspective. *International Journal of Educational Development*, *20*(1), 65–84.

Larkin, D. (2012). Misconceptions about ―misconceptions": Preservice secondary science teachers‘ views on the value and role of student ideas. *Science Education*, *96*(5), 927–959. https://doi.org/10.1002/sce.21022

Larson, L. M., Stephen, A., Bonitz, V. S., & Wu, T. F. (2014). Predicting science achievement in India: Role of gender, self-efficacy, interests, and effort. *Journal of Career Assessment*, *22*(1), 89–101.

Lastri, L., Kusumo, E., & Susilaningsih, E. (2018). Inovasi bahan ajar redoks dengan pendekatan multirepresentasi dan contextual teaching learning. *Chemistry in Education*, *7*(1), 91–97.

Lee, H., Feldman, A., & Beatty, I. D. (2012). Factors that affect science and mathematics teachers' initial implementation of technology-enhanced formative assessment using a classroom response system. *Journal of Science Education and Technology*, *21*(5), 523–539. https://doi.org/10.1007/s10956-011-9344-x

Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing*, *12*(3), 180–198. https://doi.org/10.1016/j.asw.2008.02.003

Leonard, M. J., Kalinowski, S. T., & Andrews, T. C. (2014). Misconceptions yesterday, today, and tomorrow. *CBE Life Sciences Education*, *13*(2), 179–186. https://doi.org/10.1187/cbe.13-12-0244

Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education - Is it still a matter of number right scoring or negative marking? *Studies in Educational Evaluation*, *39*(3), 188–193. https://doi.org/10.1016/j.stueduc.2013.07.001

Liampa, V., Malandrakis, G. N., Papadopoulou, P., & Pnevmatikos, D. (2017). Development and evaluation of a three-tier diagnostic test to assess undergraduate primary teachers ' understanding of ecological footprint. *Research in Science Education*. https://doi.org/10.1007/s11165-017-9643-1

Lin, J., Chu, K., & Meng, Y. (2010). Distractor rationale taxonomy : Diagnostic assessment of reading with ordered multiple-choice items. *American Education Research Association*, 1–15.

Lin, J. W. (2016). Development and evaluation of the diagnostic power for a computer-based two-tier assessment. *Journal of Science Education and Technology*, *25*(3), 497–511. https://doi.org/10.1007/s10956-016-9609-5

Lin, S.-W. (2004). Development and application of a two-tier diagnostic test for high school students' understanding of flowering plant growth and development. *International Journal of Science and Mathematics Education*, *2*(2), 175–199. https://doi.org/10.1007/s10763-004-6484-y

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, *3*(2), 103-122.

Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education*, *7*(1), 51–62. https://doi.org/10.1177/1469787406061148

Linenberger, K. J., & Bretz, S. L. (2015). Biochemistry students' ideas about how an enzyme interacts with a substrate. *Biochemistry and Molecular Biology Education*, *43*(4), 213–222. https://doi.org/10.1002/bmb.20868

Liu, O. L., Lee, H., Linn, M. C., & Liu, O. L. (2011). An investigation of explanation multiple-choice items in science assessment. *Educational Assessment*, *16*(3), 164–184. https://doi.org/10.1080/10627197.2011.611702

Liu, T. C. (2010). Developing simulation-based computer assisted learning to correct students' statistical misconceptions based on cognitive conflict theory, using

–correlation” as an example. *Educational Technology and Society*, *13*(2), 180–192.

Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*. Charlotte, NC: Information Age.

Liu, X. (2007). Elementary to high school students‘ growth over an academic year in understanding the concept of matter. *Journal of Chemical Education*, *84*, 1853–1856.

Liu, Xiufeng, & Lesniak, K. (2006). Progression in children‘s understanding of the matter concept from elementary to high school. *Journal of Research in Science Teaching*, *43*(3), 320–347. https://doi.org/10.1002/tea.20114

Liu, Xiufeng, & Lesniak, K. M. (2005). Students ‘ progression of understanding the matter concept from elementary to high school. *Science Education*, 433–450. https://doi.org/10.1002/sce.20056

Lu, S., & Bi, H. (2016). Development of a measurement instrument to assess students‘ electrolyte conceptual understanding. *Chemistry Education Research and Practice*, (1). https://doi.org/10.1039/C6RP00137H

Ludlow, L. H., & Haley, S. M. (1995). Rasch model logits: interpretation, use and transformation. *Educational and Psychological Measurement*, *55*(6), 967–975.

Luxford, C. J., & Bretz, S. L. (2014). Development of the bonding representations inventory to identify student misconceptions about covalent and ionic bonding representations. *Journal of Chemical Education*, *91*(3), 312–320. https://doi.org/10.1021/ed400700q

Madu, B. C., & Orji, E. (2015). Effects of cognitive conflict instructional strategy on students‘ conceptual change in temperature and heat. *SAGE Open*, *5*(3). https://doi.org/10.1177/2158244015594662

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers and Education*, *95*, 85–98. https://doi.org/10.1016/j.compedu.2015.12.002

Malik, U., Angstmann, E. J., & Wilson, K. (2019). Learning and conceptual change in thermal physics concepts: An examination by gender. *International Journal of Innovation in Science and Mathematics Education*, *27*(1), 37–46.

Maloney, D. P., O‘Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students‘ conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*(S1), S12–S23. https://doi.org/10.1119/1.1371296

Mann, M., & Treagust, D. (1998). A pencil and paper instrument to diagnose students‘ conceptions of breathing, gas exchange and respiration. *Australian Science Teachers Journal*, *44*(March), 55–59. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+pencil+and+paper+instrument+to+diagnose+students‘+conceptions+of+breathing,+gas+exchange+and+respiration#0

Maskiewicz, A. C., & Lineback, J. E. (2013). Misconceptions are –so yesterday!”

CBE Life Sciences Education, *12*(3), 352–356. https://doi.org/10.1187/cbe.13-01-0014

Masson, S., Potvin, P., Riopel, M., & Foisy, L. M. B. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, *8*(1), 44–55. https://doi.org/10.1111/mbe.12043

McMillan, J. H., & Schumacher, S. (2010). *Research in education : evidence-based inquiry* (7th ed.). Boston : Pearson.

Meijer, M. (2011). *Macro-meso-micro thinking with structure-property relations for chemistry. An explorative design-based study. Faculty of Sciences, Freudenthal Institute*. Retrieved from http://www.fisme.science.uu.nl/toepassingen/20065/

Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016a). Development of a three-tier test as a valid diagnostic tool for identification of misconceptions related to carbohydrates. *Journal of Chemical Education*, *93*(9), 1514–1520. https://doi.org/10.1021/acs.jchemed.6b00261

Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016b). Identification of misconceptions through multiple choice tasks at municipal chemistry competition test. *Journal of Subject Didactics*, *1*(1), 3–12. https://doi.org/10.5281/zenodo.55468

Milner-Bolotin, M. (2015). Learning Physics Teaching Through Collaborative Design of Conceptual Multiple -Choice Questions. *The Canadian Journal of Action Research*, *16*(1), 22–30.

Ministry of Education and Culture. (2016). *Silabus Kimia SMA/MA*.

Monteiro, A., Nóbrega, C., Abrantes, I., & Gomes, C. (2012). Diagnosing portuguese students‘ misconceptions about the mineral concept. *International Journal of Science Education*, *34*(17), 2705–2726. https://doi.org/10.1080/09500693.2012.731617

Morrison, J. A., & Lederman, N. G. (2003). Science teachers‘ diagnosis and understanding of students‘ preconceptions. *Science Education*, *87*(6), 849–867.

Muchson, M., & Su, M. (2015). Menggali pemahaman konsep dan identifikasi miskonsepsi mahasiswa prodi pendidikan kimia FMIPA UM angkatan 2013 pada topik ikatan ionik menggunakan instrumen tes diagnostik two tier. *Media Komunikasi Pembelajaran Kimia*, *1*(1), 99–107.

Mulford, D., & Robinson, W. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education*, *79*(6), 739–744.

Murti, P. R., Aminah, N. S., & Harjana. (2019). The identification of high school students‘ knowledge of newton‘s law of science literacy using a test based on nature of science (NOS). *Journal of Physics: Conference Series*, *1153*(1), 0–6. https://doi.org/10.1088/1742-6596/1153/1/012122

Mutlu, A., & Sesen, B. A. (2015). Development of a two-tier diagnostic test to assess undergraduates‘ understanding of some chemistry concepts. *Procedia - Social and Behavioral Sciences*, *174*, 629–635.

https://doi.org/10.1016/j.sbspro.2015.01.593

Mutlu, A., & Sesen, B. A. (2016). Evaluating of pre- service science teachers ' understanding of general chemistry concepts by using two tier diagnostic test. *Journal of Baltic Science Education*, *15*(1), 79–96.

Muttaqin, T., Wittek, R., Heyse, L., & van Duijn, M. (2019). The achievement gap in Indonesia? Organizational and ideological differences between private Islamic schools. *School Effectiveness and School Improvement*, *0*(0), 1–31. https://doi.org/10.1080/09243453.2019.1644352

Naah, B. M. (2012). Identifying students' misconceptions in writing balanced equations for dissolving ionic compounds in water and using multiple-choice questions at the symbolic and particulate levels to confront these misconceptions. *Unpublished Doctoral Dissertation, Middle Tennessee State University*.

Naah, B. M., & Sanger, M. J. (2012). Student misconceptions in writing balanced equations for dissolving ionic compounds in water. *Chemistry Education Research and Practice*, *13*(3), 186–194. https://doi.org/10.1039/c2rp00015f

Nehm, R. H., Ha, M., Rector, M., Opfer, J., Perrin, L., Ridgway, J., & Mollohan, K. (2010). Scoring guide for the open response instrument (ORI) and evolutionary gain and loss test (EGALT). *Technical Report of National Science Foundation REESE Project 0909999*, 1–40.

Nelson, P. G. (2003). Basic chemical concepts. *Chemistry Education Research and Practice*, *4*(1), 19–24. https://doi.org/10.1039/B2RP90033E

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

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Nwafor, C. E. O. A. C. O. G. (2015). Effect of Self Regulated Learning Approach on Junior Secondary School. *Journal of Education and Practice*, *6*(5), 45–53.

Nyachwaya, J. M., Mohamed, A. R., Roehrig, G. H., Wood, N. B., Kern, A. L., & Schneider, J. L. (2011). The development of an open-ended drawing tool: An alternative diagnostic tool for assessing students' understanding of the particulate nature of matter. *Chemistry Education Research and Practice*, *12*(2), 121–132. https://doi.org/10.1039/c1rp90017j

Oberoi, M. (2017). Review of literature on student's misconceptions in science. *International Journal of Scientific Research and Education*, *05*(03), 6274–6280. https://doi.org/10.18535/ijsre/v5i03.05

Odom, A. L., & Barrow, L. H. (1995). The development and application of a two-tiered diagnostic test measuring college biology students' understanding of diffusion and osmosis following a course of instruction. *Journal of Research in Science Teaching*, *32*(I), 45–61.

Önder, İ. (2018). The Effect of Conceptual Change Texts Supplemented Instruction on Students' Achievement in Electrochemistry. *International Online Journal of Educational Sciences*, (December). https://doi.org/10.15345/iojes.2017.04.006

Orbanić, N. D., Dimec, D. S., & Cencič, M. (2016). The effectiveness of a constructivist teaching model on students' understanding of photosynthesis. *Journal of Baltic Science Education*, *15*(5), 575–587.

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

Osman, K. (2017). Addressing secondary school students' misconceptions about simple current circuits using the learning cycle approach. In *Overcoming Students' Misconceptions in Science: Strategies and Perspectives from Malaysia* (pp. 223–242). Singapore: Springer. https://doi.org/10.1007/978-981-10-3437-4

Osman, S A, Badaruzzaman, W. H. W., & Hamid, R. (2011). Assessment on students performance using rasch model in reinforced concrete design course examination. *Recent Researches in Education*, 193–198.

Osman, Siti Aminah, Naam, S. I., Jaafar, O., Badaruzzaman, W. H. W., & Rahmat, R. A. A. O. K. (2012). Application of rasch model in measuring students' performance in civil engineering design II course. *Procedia - Social and Behavioral Sciences*, *56*(Ictlhe), 59–66. https://doi.org/10.1016/j.sbspro.2012.09.632

Othman, J., Treagust, D. F., & Chandrasegaran, A. L. (2008). An investigation into the relationship between students' conceptions of the particulate nature of matter and their understanding of chemical conding. *International Journal of Science Education*, *30*(11), 1531–1550. https://doi.org/10.1080/09500690701459897

Ozmen, H. (2011). Turkish primary students ' conceptions about the particulate nature of matter. *International Journal of Environmental & Science Education*, *6*(1), 99–121.

Ozmen, H., & Ayas, A. (2003). Students ' difficulties in understanding of the conservation of matter in open and closed-system chemical reactions. *Chemistry Education: Research and Practice*, *4*(3), 279–290.

Pabuçcu, A., & Geban, Ö. (2006). Remediating misconceptions concerning chemical through conceptual change text. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, (30), 184–192.

Pagano, R. R. (2009). *Understanding statistics in the behavioral sciences*. Belmont, California: Wadsworth, Cengage Learning.

Park, M., & Liu, X. (2019). An investigation of item difficulties in energy aspects across biology , chemistry , environmental science , and physics. *Research in*

Science Education.

Pentecost, T. C., & Langdon, L. S. (2008). Development of a covalent bonding and molecular struc- ture learning progression and concept inventory for first semester general chemistry. *Paper Pre- Sented at the 20th Biennial Conference on Chemical Education, Bloomington, Indiana*.

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

Peşman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *Journal of Educational Research*, *103*(3), 208–222. https://doi.org/10.1080/00220670903383002

Peterson, R. F, Treagust, D. F., & Garnett, P. (1986). Identification of secondary students' misconceptions of covalent bonding and structure concepts using a diagnostic test instrument. *Research in Science Education*, *16*, 40–48.

Peterson, Raymond F, & Treagust, D. F. (1989). Conceptual difficulties experienced by senior high school students of electrochemistry : Electrochemical (galvanic) and electrolytic cells. *Journal of Research in Science Teaching*, *26*(4), 301–314. https://doi.org/10.1002/tea.3660291006

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*(2), 167–199.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science & Education*, *66*, 211–227.

Potvin, P., Skelling-desmeules, Y., & Sy, O. (2015). Exploring secondary students ' conceptions about fire using a two- tier , true / false , easy-to-use diagnostic test. *Journal of Education in Science, Environment and Health (JESEH)*, *1*(2), 63–78.

Prastiwi, A., Kholiq, A., & Setyarsih, W. (2018). Implementation of ECIRR model based on virtual simulation media to reduce students' misconception on kinetic theory of gases. *Journal of Physics: Conference Series*, *1006*(1), 12005. https://doi.org/10.1088/1742-6596/1006/1/012005

Prodjosantoso, A. K., Hertina, A. M., & Irwanto, I. (2019). The misconception diagnosis on ionic and covalent bonds concepts with three tier diagnostic test. *International Journal of Instruction*, *12*(1), 1477–1488.

Putri, L. O. L., Rahman, T., & Priyandoko, D. (2016). Analyzing concepts mastery and misconceptions about evolution. *Journal of Physics: Conference Series*, *755*(1), 0–6. https://doi.org/10.1088/1742-6596/755/1/011001

Rahayu, S. R. I., & Kita, M. (2010). An analysis of Indonesian And Japanese students ' understandings of macroscopic and submicroscopic. *International Journal of Science and Mathematics Education*, *8*(2010), 667–688. https://doi.org/10.1007/s10763-009-9180-0

Rahman, M. N. A., Zamri, S. N. A. S., & Leong, K. E. (2018). Kajian Meta-Analisis

Pengetahuan Guru Matematik di Malaysia. *Jurnal Kurikulum & Pengajaran Asia Pasifik*, *6*(2), 11–22.

Rahmawati, A. (2015). Pengembangan modul kimia dasar berbasis multipel representasi untuk meningkatkan kemampuan berpikir kritis mahasiswa. *Phenomenon Jurnal Pendidikan MIPA*, *5*(2).

Rappoport, L. T., & Ashkenazi, G. (2008). Connecting levels of representation: Emergent versus submergent perspective. *International Journal of Science Education*, *30*(12), 1585–1603.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory* (1st ed.). Madison Avenue, New York: Routledge/Taylor & Francis Group.

Rode, N. (2005). Translation of measurement instruments and their reliability : An example of job-related affective well-being scale. *Metodološki Zvezki*, *2*(1), 15–26.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta- analysis of 80 years of research. *Educational Measurement: Issues & Practice*, *24*(2), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Rodriguez, M. C., Kettler, R. J., & Elliott, S. N. (2014). Distractor functioning in modified items for test accessibility. *SAGE Open*, *4*(4). https://doi.org/10.1177/2158244014553586

Rollnick, M., & Mahoona, P. P. (1999). A quick and effective way of diagnosing student difficulties: two tier from simple multiple choice questions. *International Journal of Conflict Management*, *4*(52), 161–164. https://doi.org/10.1086/250095

Romine, W. L., Schaffer, D. L., & Barrow, L. (2015). Development and application of a novel rasch-based methodology for evaluating multi-tiered assessment instruments: validation and utilization of an undergraduate diagnostic test of the water cycle. *International Journal of Science Education*, *37*(16), 2740–2768. https://doi.org/10.1080/09500693.2015.1105398

Romine, W. L., Todd, A. N., & Clark, T. B. (2016). How do undergraduate students conceptualize acid–base chemistry ? Measurement of a concept progression. *Science Education*, *100*(6), 1150–1183. https://doi.org/10.1002/sce.21240

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifyng content validity : Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94–104. https://doi.org/10.1093/swr/27.2.94

Saat, R. M., Hidayah, M. F., Aziz, N. A. A., Haron, K., Rashid, K. A., & Shamsuar, N. R. (2016). Development of an online three-tier diagnostic test to assess pre-university students ‘ understanding. *Journal of Baltic Science Education*, *15*(4), 532–546.

Sackes, M. (2013). The role of cognitive, metacognitive, and mo- tivational variables in conceptual change: preservice early child- hood teachers‘ conceptual understanding of the cause of lunar phases. *Unpublished Doctoral Dissertation, Columbus: Ohio State University.*

Sadhu, S., & Laksono, E. W. (2018). Development and validation of an integrated assessment for measuring critical thinking and chemical literacy in chemical equilibrium. *International Journal of Instruction*, *11*(3), 557–572. https://doi.org/10.12973/iji.2018.11338a

Sadhu, S., Tima, M. T., Cahyani, V. P., Laka, A. F., Annisa, D., & Fahriyah, A. R. (2017). Analysis of acid-base misconceptions using modified certainty of response index (CRI) and diagnostic interview for different student levels cognitive. *International Journal of Science and Applied Science: Conference Series*, *1*(2), 91. https://doi.org/10.20961/ijsascs.v1i2.5126

Sadler, P. M., & Sonnert, G. (2016). Understanding misconceptions teaching and learning in middle school physical science. *American Educator*, *40*(1), 26–32. Retrieved from https://eric.ed.gov/?id=EJ1094278

Saidfudin, M., Azrilah, A., Rodzo'an, N., Omar, M., Zaharim, a, & Basri, H. (2010). Easier learning outcomes analysis using Rasch model in engineering education research. *Latest Trends on Engineering Education*, 442–447. Retrieved from http://dl.acm.org/ft_gateway.cfm?id=1864264&type=pdf

Saifullah, A. M., Sutopo, S., & Wisodo, H. (2017). Senior high school students' difficulties in solving impulse and momentum problems. *Jurnal Pendidikan IPA Indonesia*, *6*(1), 1–10. https://doi.org/10.15294/jpii.v6i1.9593

Salirawati, D. (2011). Pengembangan instrumen pendeteksi miskonsepsi kesetimbangan kimia pada peserta didik SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *15*(2), 232–249.

Salta, K., & Tzougraki, C. (2011). Conceptual versus algorithmic problem-solving: Focusing on problems dealing with conservation of matter in chemistry. *Research in Science Education*, *41*(4), 587–609. https://doi.org/10.1007/s11165-010-9181-6

Samsudin, A., Liliawati, W., Sutrisno, A. D., Suhendi, E., & Kaniawati, I. (2014). The Use of Computer Simulation in Cooperative Learning to Minimize Students' Misconceptions of Momentum and Impulse. *International Conference on Advances in Education Technology (ICAET 2014) The*, (Icaet), 72–74. https://doi.org/10.2991/icaet-14.2014.16

Sana, S., Adhikary, C., & Chattopadhyay, K. N. (2018). Evolutionary paradigm shift in the instructional strategies of chemical concepts. *Bhatter College Journal of Multidisciplinary Studies*, *8*(1), 72–79.

Sanger, M. J., & Greenbowe, T. J. (1992). An analysis of college chemistry textbooks as sources of misconceptions and errors in electrochemistry. *Journal of Chemical Education*, *76*(6), 853–860.

Santos, V. C., & Arroio, A. (2016). The representational levels: Influences and contributions to research in chemical education. *Journal of Turkish Science Education*, *13*(1), 3–18. https://doi.org/10.12973/tused.10153a

Sari, D. R., Sopandi, W., Koesbandiyah, H., & Arviana, R. (2018). Analysis of primary school student's misconception through six tire diagnostic test abput the concept of water characteristics. *Jurnal Dimensi Pendidikan Dan Pembelajaran*, *6*(3).

Schaffer, D. (2012). An analysis of science concept inventories and diagnostic tests: Commonalities and differences. *Annual International Conference of the National Association for Research in Science Teaching*, (April).

Schaffer, D. L. (2013). The development and validation of a three-tier diagnostic test measuring pre-service elementary education and secondary science teachers' understanding of the water cycle. *Unpublished Doctoral Dissertation, University of Missouri*, (May), 178.

Schmidt, H. (1997). Students ' Misconceptions — Looking for a Pattern. *Science & Education*.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on compherension. *Journal of Educational Psychology*, *82*(3), 498–504.

Schroeder, J., Murphy, K. L., & Holme, T. A. (2012). Investigating factors that influence item performance on ACS exams. *Journal of Chemical Education*, *89*(3), 346–350. https://doi.org/10.1021/ed101175f

Schultz, M., Lawrie, G. A., Bailey, C. H., Simon, B., Dargaville, T. R., Brien, G. O., … Thompson, C. D. (2017). Evaluation of diagnostic tools that tertiary teachers can apply to profile their students ' conceptions. *International Journal of Science Education ISSN:* https://doi.org/10.1080/09500693.2017.1296980

Schwartz, P., & Barbera, J. (2014). Evaluating the content and response process validity of data from the chemical concepts inventory. *Journal of Chemical Education*, *91*, 630–640.

Sen, S., & Yilmaz, A. (2017). The development of a three-tier chemical bonding concept test. *Journal of Turkish Science Education*, *14*(1), 110–126. https://doi.org/10.12973/tused.10193a

Şenel, S., Pehlivan, E. B., & Alatlı, B. (2015). Effect of correction-for-guessing formula on psychometric characteristics of test. *Procedia - Social and Behavioral Sciences*, *191*, 925 – 929.

Şenol, Ş. (2018). Examination of the preservice chemistry teachers ' understanding of periodic table concepts. *SHS Web of Conferences*, *48*.

Sesli, E., & Kara, Y. (2012). Development and application of a two-tier multiple-choice diagnostic test for high school students' understanding of cell division and reproduction. *Journal of Biological Education*, *46*(4), 214–225. https://doi.org/10.1080/00219266.2012.688849

Shaughnessy, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, *13*, 111–119.

Shi, J., Wood, W. B., Martin, J. M., Guild, N. A., Vicens, Q., & Knight, J. K. (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE-Life Sciences Education*, *9*, 453–461. https://doi.org/10.1187/cbe.10

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209–215. https://doi.org/10.1016/j.cognition.2012.04.005

Shultz, K. S., Whitney, D. J., & Zickar, M. J. (2014). *Measurement theory in action:*

Case studies and exercises (2nd ed.). Third Avenue, New York: Routledge/Taylor & Francis Group.

Singh, C., & Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *American Journal of Physics*, *71*(6), 607–617.

Sirhan, G. (2007). Learning difficulties in chemistry: An overview. *Journal of Turkish Science Education*, *4*(2), 2–20.

Siswaningsih, W., Firman, H., Zackiyah, Z., & Khoirunnisa, A. (2016). Development of two-tier diagnostic test pictorial-based for identifying high school students misconceptions on the mole concept. *Journal of Physics: Conference Series*, *755*(1), 0–7. https://doi.org/10.1088/1742-6596/755/1/011001

Soeharto, S. (2016). Implementation of text transformation in physics education to reduce students' misconception. *JETL (Journal Of Education, Teaching and Learning)*, *1*(2), 56. https://doi.org/10.26737/jetl.v1i2.38

Sopapun, S. (2002). Development of a two-tiered multiple choice test to measure misconceptions in physics among high school students in Thailand. *Unpublished Doctoral Dissertation, Oregon State University*, 165. https://doi.org/10.1016/j.jorganchem.2010.09.058

Sreenivasulu, B., & Subramaniam, R. (2013). University students' understanding of chemical thermodynamics. *International Journal of Science Education*, *35*(4), 601–635. https://doi.org/10.1080/09500693.2012.683460

Sreenivasulu, B., & Subramaniam, R. (2014). Exploring undergraduates' understanding of transition metals chemistry with the use of cognitive and confidence measures. *Research in Science Education*, *44*(6), 801–828. https://doi.org/10.1007/s11165-014-9400-7

Stains, M., Escriu-Sune, M., Santizo, M. L. M. A. de, & Sevian, H. (2011). Assessing secondary and college students' implicit assumptions about the particulate nature of matter: Development and validation of the structure and motion of matter survey. *Journal of Chemical Education*, *88*, 1359–1365.

Stevens, A. M., Smith, A. C., Marbach-ad, G., Balcom, S. A., Buchner, J., Daniel, S. L., … Yarwood, S. A. (2017). Using a concept inventory to reveal student thinking associated with common misconceptions about antibiotic resistance. *Journal of Miscrobiology and Biology Education*, *18*(1), 1–6.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Madison Avenue, New York: Routledge/Taylor & Francis Group.

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

Su, C. T., & Parham, L. D. (2002). Generating a valid questionnaire translation for cross-cultural use. *American Journal of Occupational Therapy*, *56*(5), 581–585. https://doi.org/10.5014/ajot.56.5.581

Subayani, N. W. (2016). The profile of misconceptions among science subject student-teachers in primary schools. *International Journal of Education and*

Literacy Studies, *4*(2). https://doi.org/10.7575/aiac.ijels.v.4n.2p.54

Suliyanah, Putri, H. N. P. A., & Rohmawati, L. (2018). Identification student‘s misconception of heat and temperature using three-tier diagnostic test. *Journal of Physics: Conference Series*, *997*(1), 0–10. https://doi.org/10.1088/1742-6596/997/1/012035

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Cimahi: Trim Komunikata.

Sun, L., Bradley, K. D., & Akers, K. (2012). A multilevel modelling approach to investigating factors impacting science achievement for secondary school students: PISA Hong Kong sample. *International Journal of Science Education*, *34*(14), 2107–2125.

Sungur, S., Tekkaya, C., & Geban, O. (2001). The contribution of conceptual change texts accompanied by concept mapping to students‘ understanding of the human circulatory system. *School Science & Mathematics*, *101*(2), 91–101.

Svec, M. ., Boone, W. ., & Olmer, C. (1995). Changes in a pre-service elementary teachers‘ physics course. *Journal of Science Teacher Education*, *6*(2), 79–88.

Tabachnick, B. G., & Fidell, L. . (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education Inc.

Taber, K. S., & Tan, K. C. D. (2011). The insidious nature of ‗hard-core‘ alternative conceptions: Implications for the constructivist research programme of patterns in high school students‘ and pre-service teachers‘ thinking about ionisation energy. *International Journal of Science Education*, *33*, 259–297.

Taber, Keith 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

Taber, Keith S. (2018). The use of cronbach ‘ s alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Talanquer, V. (2006). Commonsense chemistry: A model for understanding students‘ alternative conceptions. *Journal of Chemical Education*, *83*(5), 811. https://doi.org/10.1021/ed083p811

Talanquer, V. (2011). Macro, submicro, and symbolic: The many faces of the chemistry ―triplet.” *International Journal of Science Education*, *33*(2), 179–195. https://doi.org/10.1080/09500690903386435

Tamir, P. (1989). Some issues related to the use of justifications to multiple-choice answers. *Journal of Biological Education*, *23*, 285-292.

Tan, K. C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2002). Development and application of a two-tier multiple-choice diagnostic instrument to assess high school students‘ understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching*, *39*(4), 283–301.

Tan, K. C. D., & Treagust, D. F. (1999). Evaluating students‘ understanding of chemical bonding. *School Science Review*, *81*(294), 75–83.

Taskin, V., Bernholt, S., & Parchmann, I. (2015). An inventory for measuring student teachers‘ knowledge of chemical representations: design, validation, and psychometric analysis. *Chemistry Education Research and Practice*, *16*(3), 460–477. https://doi.org/10.1039/c4rp00214h

Taslidere, E. (2016). Development and use of a three-tier diagnostic test to assess high school students ‘ misconceptions about the photoelectric effect. *Research in Science & Technological Education*. https://doi.org/10.1080/02635143.2015.1124409

Tastan, Ö., Yalcinkaya, E., & Boz, Y. (2010). Pre-service chemistry teachers‘ idea about reaction mechanism. *Journal of Turkish Science Education*, *7*(1), 47–60.

Tatli, Z., & Ayas, A. (2013). Effect of a virtual chemistry laboratory on students ‘ achievement. *Journal of Educational Technology & Society*, *16*(1), 159–170.

Taufiq, L., Sriyati, S., & Priyandonko, D. (2018). Students‘ conceptual change on human reproduction concept using scientific approach. *International Journal of Science and Applied Science: Conference Series*, *2*(1), 216. https://doi.org/10.20961/ijsascs.v2i1.16714

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach ‘ s alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Taylor, A. K., & Kowalski, P. (2004). Naive psychological science: The prevalence, strenght, and sources of misconceptions. *The Psychological Record*, *54*, 15–25. https://doi.org/10.1007/BF03395459

Tellinghuisen, J., & Sulikowski, M. M. (2008). Does the Answer Order Matter on Multiple-Choice Exams? *Journal of Chemical Education*, *85*(4), 572. https://doi.org/10.1109/20.800885

Teo, T. W., Goh, M. T., & Yeo, L. W. (2014). Chemistry education research trends: 2004-2013. *Chemistry Education Research and Practice*, *15*, 470–487. https://doi.org/10.1039/b000000x

Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton‘s Laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics ·*, *66*(4), 338–352. https://doi.org/10.1119/1.18863

Tien, L. T., & Osman, K. (2017). Misconceptions in electrochemistry : How do pedagogical agents help? In *Overcoming Student's Misconceptions in Science* (pp. 91–110). Singapore: Springer. https://doi.org/10.1007/978-981-10-3437-4

Tongchai, A., Sharma, M. D., Johnston, I. D., Arayathanitkul, K., & Soankwan, C. C. (2009). Developing , evaluating and demonstrating the use of a conceptual survey in mechanical waves. *International Journal of Science Education*, *31*(18), 2437–2457. https://doi.org/10.1080/09500690802389605

Touli, E. H., Talbi, M., & Radid, M. (2015). Barriers to learning physics in moroccan secondary college: the case of student‘s representations in electricity. *Procedia - Social and Behavioral Sciences*, *197*(July), 278–280.

https://doi.org/10.1016/j.sbspro.2015.07.136

Treagust, D. (1986). Evaluating students' misconceptions by means of diagnostic multiple choice items. *Research in Science Education*, *16*, 199–207.

Treagust, D.F. (2006). Diagnostic assessment in science as a means to improving teaching, learning and retention. *UniServe Science Assessment Symposium Proceedings*, 1–9. Retrieved from http://openjournals.library.usyd.edu.au/index.php/IISME/article/view/6375

Treagust, David F., Chittleborough, G., & Mamiala, T. L. (2003). The role of submicroscopic and symbolic representations in chemical explanations. *International Journal of Science Education*, *25*(11), 1353–1368. https://doi.org/10.1080/0950069032000070306

Treagust, David F, Chandrasegaran, A. L., Crowley, J., Yung, B. H. W., Cheong, I. P. A., & Othman, J. (2010). Evaluating students ' understanding of kinetic particle theory concepts relating to the states of matter , changes of state and diffusion : A cross- national study. *International Journal of Science and Mathematics Education*, *8*, 141–164. https://doi.org/10.1007/s10763-009-9166-y

Treagust, David F, & Duit, R. (2009). Multiple perspectives of conceptual change in science and the challenges ahead. *Journal of Science and Mathematics Education in Southeast Asia*, *32*(2), 89–104. Retrieved from http://www.recsam.edu.my/R&D_Journals/YEAR2009/dec2009vol2/multiplepe rspectives(89-104).pdf%5Cnpapers2://publication/uuid/EF6D2C76-CF06-489F-9FBA-F69FC3749413

Trumper, R. (2000). A cross-college age study about physics students' conceptions of force in pre-service training for high school teachers. *Curriculum Matters*, 227–238.

Trumper, R., & Gorsky, P. (1997). A survey of biology students' conceptions of force in pre-service training for high school teachers. *Research in Science & Technological Education*, *15*(2), 133–147.

Tsui, C. Y., & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, *32*(8), 1073–1098. https://doi.org/10.1080/09500690902951429

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

Tyson, L., Treagust, D. F., & Bucat, R. B. (1999). The complexity of teaching and learning chemical equilibrium. *Journal of Chemical Education*, *76*, 554–558.

Uzuntiryaki, E., & Geban, Ö. (2005). Effect of conceptual change approach accompanied with concept mapping on understanding of solution concepts. *Instructional Science*, *33*(4), 311–339. https://doi.org/10.1007/s11251-005-2812-z

Velde, G. V. A. N. D. E. R., Beaton, D., Hogg-johnston, S., Hurwitz, E., & Tennant,

A. (2009). Rasch analysis provides new insights into the neck disability index. *Arthritis & Rheumatism (Arthritis Care & Research)*, *61*(4), 544–551. https://doi.org/10.1002/art.24399

Vilppu, H., Mikkilä-erdmann, M., & Ahopelto, I. (2011). The role of regulation and processing strategies in understanding science text among university students. *Scandinavian Journal of Educational Research*, *57*(3), 246–262. https://doi.org/10.1080/00313831.2011.637229

Vitharana, P. R. K. A. (2015). Student misconceptions about plant transport – a Sri Lankan example. *European Journal of Science and Mathematics Education*, *3*(3), 275–288.

Voska, K. W., & Heikkien, H. W. (2000). Identification and analysis of student conceptions used to solve chemical equilibrium problems. *Journal of Research in Science Teaching*, *37*(2), 160–176.

Vosniadou, S. (2012). Reframing the classical approach to conceptual change: Preconceptions, misconceptions and synthetic models. . . *In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), Second International Handbook of Science Education (Pp. 119–130). Dordrecht, The Netherlands: Springer*.

Vrabec, M., & Prokša, M. (2016). Identifying misconceptions related to chemical bonding concepts in the slovak school system using the bonding representations inventory as a diagnostic tool. *Journal of Chemical Education*, *93*(8), 1364–1370. https://doi.org/10.1021/acs.jchemed.5b00953

Vyas, R., & Supe, A. (2008). Multiple choice questions: A literature review on the optimal number of options. *The National Medical Journal of India*, *3*, 130–133.

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. *In D. L. Gabel (Ed.), Handbook of Research on Science Teaching and Learning (Pp.177-210). New York: Macmillan.*

Wang, J. R. (2004). Development and validation of a two-tier instrument to examine understanding of internal transport in plants and the human circulatory system. *International Journal of Science and Mathematics Education*, *2*, 131-157.

Wartono, W., Batlolona, J. R., & Putirulan, A. (2018). Cognitive Conflict Strategy and Simulation Practicum to Overcome Student Misconception on Light Topics. *Journal of Education and Learning (EduLearn)*, *12*(4), 747. https://doi.org/10.11591/edulearn.v12i4.10433

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

Widarti, H. R., Permanasari, A., & Mulyani, S. (2016). Student misconception on redox titration (A challenge on the course implementation through cognitive dissonance based on the multiple representations). *Jurnal Pendidikan IPA Indonesia*, *5*(1), 56–62. https://doi.org/10.15294/jpii.v5i1.5790

Widarti, Hayuni Retno, Permanasari, A., & Mulyani, S. (2017). Undergraduate Students' Misconception on Acid-Base and Argentometric Titrations: a

Challenge To Implement Multiple Representation Learning Model With Cognitive Dissonance Strategy. *International Journal of Education*, *9*(2), 105. https://doi.org/10.17509/ije.v9i2.5464

Wijaya, C. P., Supriyono Koes, H., & Muhardjito. (2016). The diagnosis of senior high school class X MIA B students misconceptions about hydrostatic pressure concept using three-tier. *Jurnal Pendidikan IPA Indonesia*, *5*(1), 14–21. https://doi.org/10.15294/jpii.v5i1.5784

Wijayanti, M. D., Raharjo, S. B., Saputro, S., & Mulyani, S. (2018). Investigation to reduce students' misconception in energy material. *Journal of Physics: Conference Series*, *1013*(1), 0–6. https://doi.org/10.1088/1742-6596/1013/1/012080

Wild, D., Eremenco, S., Mear, I., Martin, M., Houchin, C., Gawlicki, M., … Molsen, E. (2009). Multinational trials—recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: The ISPOR patient-reported outcomes translation and linguistic. *Value in Health*, *12*(4), 430–440. https://doi.org/https://doi.org/10.1111/j.1524-4733.2008.00471.x

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translating adaptation. *Value Health*, *8*(2), 94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x

William, D. (2013). Assessment : The bridge between teaching and learning. *Voices from the Middle*, *21*(2), 15–20.

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

Winarni, S., & Syahrial, S. (2011). Analisis kesalahan konsep level mikroskopis yang dialami mahasiswa tingkat akhir program studi pendidikan kimia FKIP Unsyiah pada materi persamaan reaksi. *Unpublished Laporan Penelitian Dosen Muda Universitas Syiah KUala*.

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

Wren, D., & Barbera, J. (2014). Psychometric analysis of the thermochemistry concept inventory. *Chemistry & Biodiversity Education Research and Practice*, *15*, 380–390. https://doi.org/10.1039/b000000x

Wright, B. D. (1977). Solving Measurement Problems with the Rasch Model. *Journal of Educational Measurement*, *14*(2), 97–116.

Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel Rasch modeling of two-tier multiple choice test: A case study using Lawson's classroom test of scientific reasoning. *Physical Review Physics Education Research*, *14*(2), 020104. https://doi.org/10.1103/PhysRevPhysEducRes.14.020104

Yalcin, F. A. (2012). Pre-service primary science teachers‘ understandings of the effect of temperature and pressure on solid-liquid phase transition of water. *Chemistry Education Research and Practice*, *13*(3), 369–377. https://doi.org/10.1039/c2rp20021j

Yalcinkaya, E., Tastan, O., & Boz, Y. (2009). High school students‘ conceptions about energy in chemical reactions. *Pamukkale Universitesi Egitim Fakultesi Dergisi*, *26*, 1–11.

Yan, Y. K., & Subramaniam, R. (2018). Using a multi-tier diagnostic test to explore the nature of students‘ alternative conceptions on reaction. *Chemistry Education Research and Practice*, *19*, 213–226. https://doi.org/10.1039/c7rp00143f

Yang, D. C., Li, M. N., & Lin, C. I. (2008). A study of the performance of 5th graders in number sense and its relationship to achievement in mathematics. *International Journal of Science and Mathematics Education*, *6*(4), 789–807. https://doi.org/10.1007/s10763-007-9100-0

Yang, D. C., & Lin, Y. C. (2015). Assessing 10- to 11-year-old children‘s performance and misconceptions in number sense using a four-tier diagnostic test. *Educational Research*, *57*(4), 368–388. https://doi.org/10.1080/00131881.2015.1085235

Yasthophi, A., & Ritonga, P. S. (2017). Miskonsepsi mahasiswa mengenai ikatan ion dalam senyawa NaCl. *Konfigurasi*, *1*(2), 195–202.

Yetisir, M. İ. (2014). The multilevel effects of student and classroom factors on the science achievement of eighth graders in turkey. *Education and Science*, *39*(172), 108–120.

Yezierski, E. J., & Birk, J. P. (2006). Misconceptions about the particulate nature of matter using animations to close the gender gap. *Research: Science and Education*, *83*(6), 954–960.

Young, T. A., Yang, Y., Brazier, J. E., & Tsuchiya, A. (2011). The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation : The Case of moving from AQLQ to AQL-5D, 195–210. https://doi.org/10.1177/0272989X10364846

Yumusak, A., Maras, İ., & Sahin, M. (2015). Effects of computer-assisted instruction with conceptual change texts on removing the misconceptions of radioactivity. *Journal for the Education of Gifted Young Scientists*, *3*(2), 23–23. https://doi.org/10.17478/jegys.2015214277

Zheng, C., Fu, L., & He, P. (2014). Development of an instrument for assessing the effectiveness of chemistry classroom teaching. *Journal of Science Education and Technology*, *23*(2), 267–279. https://doi.org/10.1007/s10956-013-9459-3

Zuidgeest, M., Hendriks, M., Koopman, L., Spreeuwenberg, P., & Rademakers, J. (2011). Comparison of a postal survey and mixed-mode survey using a questionnaire on patients‘ experiences with breast care. *Journal of Medical Internet Research*, *13*(3).