JTC-RE: Journal of Tropical Chemistry Research and Education 6, 2 (2024): 105-118
Website: http://ejournal.uin-suka.ac.id/tarbiyah/index.php/jtcre
ISSN 2685-5690 **(online) ISSN** 2685-144X **(print)**

USING ITEM OPTION CHARACTERISTICS CURVE (IOCC) TO UNFOLD MISCONCEPTION ON CHEMICAL REACTION

Hilman Qudratuddarsi[1], Nurhikma Ramadhana[1], Nor Indriyanti[1], Ayu Indayanti Ismail[1,]

Universitas Sulawesi Barat
E-mail: hilman.qudratuddarsi@unsulbar.ac.id

ABSTRACT

Misconceptions can significantly hinder the learning process. To address this, various diagnostic instruments such as two-tier (2TMC), three-tier, and four-tier multiple-choice questions have been introduced. However, as the number of tiers increases, identifying misconceptions becomes more complex. Therefore, this study employs the Item Option Characteristics Curve (IOCC) to identify misconceptions by calculating the probability of each option being selected. This study is a quantitative study with survey design to directly test student abilities. The Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI) was administered to 185 pre-service teachers across three universities in Indonesia. The data was analyzed using Winstep software to generate the IOCC for each item. The analysis revealed that each item in the phenomenon and reasoning tiers contains distractors that could interfere with the option selected by pre-service chemistry teachers. While the alternative answers identified using traditional methods (commonly used since the introduction of 2TMC) were mostly similar to those identified by IOCC, the IOCC provided more detailed insights. Specifically, it highlighted unexpected curves after 0 logits, identified less effective distractors, and revealed inconsistencies in the most influential distractors. These findings suggest that the IOCC provides richer, more detailed information and can be a valuable alternative framework for analyzing 2TMC items to unfold misconceptions.

Keywords: assessment, chemical reaction, misconception, Rasch model

DOI: https://doi.org/10.14421/jtcre.2024.62-04

1. INTRODUCTION

Misconceptions in chemistry education are a significant issue because, once formed, they tend to be extremely persistent and difficult to correct (Qudratuddarsi et al., 2019).Misconceptions can interfere with students' learning processes, particularly when they are attempting to comprehend scientific phenomena. Furthermore, misconceptions can hinder students from achieving meaningful learning experiences, as the gaps in their knowledge impede their ability to connect new information with existing knowledge (Chew & Cerbin, 2021). The detrimental impact of misconceptions is that, even when they are believed to have been replaced by scientific understanding, they remain encoded in neural networks and continue to interfere with the acquisition of scientific knowledge. Using MediaLab v1.21 software to measure the speed and accuracy of respondents' thinking, it is found that the process of conceptual change merely suppresses misconceptions rather than completely replacing them (Masson et al., 2014; Shtulman & Valcarcel, 2012).

To reveal these misconceptions is important to provide teachers with the opportunity to tackle this issue. Interviews, essays, concept inventory and multiple-tiered multiple-choice questions are some of the tools that teachers could use to identify students' misconceptions (Liampa et al., 2017; Resbiantoro et al., 2022). Although there are many ways of identifying misconceptions, recent trends in literature point to the direction of multiple-choice diagnostic instruments. To increase its power to detect misconceptions, researchers have developed from two-tier to three and then four-tier multiple-choice questions. Adding more tiers into multiple-choice questions tend to complicate instrument users (i.e., school teachers or lectures) because it enhances the difficulties of using diagnostics instrument. The complexicity tend to be impractical in daily teaching and learning practices (Ardiansah et al., 2018).

Diagnostics power of original multiple-choice question can be enhanced using distractor analysis based on Rasch model as shown by (Herrmann-abell & Deboer (2016) and Wind & Gale (2015). The basic principle of the Rasch model is to measure latent traits, such as abilities, attitudes, or perceptions, by modeling the probability of a specific response (e.g., a correct answer) as a function of two factors: 1) person ability: the ability of the respondent or individual being assessed, 2) item difficulty: the difficulty of the question or item being answered (Qudratuddarsi et al., 2022). The Rasch model has been used for various purposes in science education, including validating instruments, analyzing students' responses, and evaluating survey results (Chan & Subramaniam, 2020).

The Rasch model has been widely used to analyze misconceptions in science education. Wind and Gale (2015) conducted a study focused on identifying misconceptions related to physics concepts among eighth-grade students. Similarly, Herrmann-Abell and DeBoer (2016) examined misconceptions about energy across a much larger sample of approximately 20,000 students from various grade levels in the United States. To date, the only study applying the Rasch model to the field of chemistry, to our knowledge, is Mulyani et al. (2021), who explored misconceptions among first-year high school students in the topic of electrolysis. These studies highlight the utility of the Rasch model in uncovering persistent misconceptions that can hinder student understanding of key scientific concepts. While research in physics and general science education is more abundant, further studies in chemistry and other specific scientific disciplines are needed.

In the current study, the analysis of misconceptions will be based on distractor analysis, which relies on the percentages of students selecting a distractor. The method, known as the Item Option Characteristics Curve (IOCC), was introduced by Herrmann-Abell and DeBoer (2011) to analyze ordinary multiple-choice questions. To our knowledge, this study is novel study in terms of the uses of two-tier multiple choice questions (2TMC) and comparisons of two methods. The basic idea of IOCC analysis is to examine trace lines for alternative choices (Ding & Beichner, 2009). The distractor analysis plots were developed 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). Basically, the relative popularity of each answer choice for students (y-axis) and different levels of achievement (*x*-axis). In details, after Rasch estimates of student achievement on the logit scale were obtained from the Winstep computer program and 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). Therefore, this study will analyze 2TMC by using IOCC to unfold misconceptions of pre-service chemistry teachers in the topic of chemical reactions.

2. RESEARCH METHODS

This quantitative study collected data from a test designed to quantify the number of students selecting each option in a 2TMC format (Hidayat, Idris, et al., 2021). Before data collection, the researcher informed the test-takers that the examination would not affect their grades but could be used by their lecturers to enhance the learning process. A paper-and-pencil-based test was chosen as the data collection method due to its ability to allow observation of the testing process, its higher response rate, and its affordability for the respondents.

1. Participants

The sample was selected using stratified random sampling. First, the population was divided into groups based on academic year: first, second, and third year. Then, 65% of students from each group were randomly chosen using SPSS 25. The final sample consisted of 185 pre-service teachers (19 males, 166 females) aged 18-21 years. These students were from three different universities: University A (61.62%), University B (27.56%), and University C (10.81%). They were distributed across three academic years: first year (40.19%), second year (32.06%), and third year (27.75%).

2. Instruments

The instrument namely representational systems and chemical reactions diagnostic instrument (RSCRDI) was adapted from (Chandrasegaran, Treagust, & Mocerino, 2007) and the whole instrument was available at (Chandrasegaran, Treagust, & Mocerino, 2011). The topic is chemical reactions which emphasized on involving multiple representations. It was selected because of the opportunity of analyzing the aspect of distractor driven

misconception in the options of phenomenon tier and reasoning tier. The instrument was then translated to avoid error in testing because of misinterpretation of questions from respondents and the efficient time of administration because of the reduction of required time for students to interpret the question (Wild et al., 2009). Method of RSCRDI translation was back-translation with the following procedures: 1) The instrument was first translated from English as its original language to Indonesian language by its researchers and reviewed by some graduate chemistry students who study at Thailand, Japan, and Australia 2) The raw translation document was sent to two experts in Chemistry who are good at Indonesian language and English. 3)The third step was back translation of Indonesian language version of the instrument to English language by a chemistry lecture. 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.

Since 2TMC was a test instrument, the face and content validity were estimated. Content validity was based on the expert judgment, while face validity required the respondent to make such judgment about the validity of the instrument (Delgado-Rico et al., 2012; Shultz et al., 2014). To measure the content validity of the instrument, the draft of the translated version is reviewed by three chemistry lecturers in Chemistry (regarded as subject matters expert (SME) (Shultz et al., 2014). They have to evaluate the appropriateness of the instrument to measure misconception concerning chemical reaction that involved multiple representations (Raykov & Marcoulides, 2011). Generally, all panels agreed that all questions are appropriate to measure chemical reactions using multiple representations and its distractors look well-functioned and tend to distract students well.

The next measure was construct validity and reliability by analyzing data of pilot study (Fraenkel, Wallen, & Hyun, 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 aged 18-20 years old from student at university A as in the real study. They are 69 students (10 males, 59 females) in their first year (39%), second-year (30.5%) and third-year (30.5%) of their study. From this step, the administration time is 45 minutes after the observation of pilot study administration time. In estimate construct validation using the Rasch model to find reliability, separation and psychometric properties. The rule of scoring system as follows:

Table 1. 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, Chu, Treagust, & Neumann, 2015; Park & Liu, 2019; Sadhu & Laksono, 2018; Xiao, Han, Koenig, Xiong, & Bao, 2018)

As the proof of construct validation, analysis of items in the current study utilized Rasch model. There are some fit statistics to measure such as mean square (MNSQ), tolerated Z-Standard (ZSTD) and Correlation Points (Pt Mea Corr). This study found that all items follow the criteria of 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 Mea Corr): 0.4 <Pt Mea Corr <0.85 (Boone, Staver, & Yale, 2014).

The next measure is reliability, the degree to which an instrument consistently give a similar result among numerous administration (Qudratuddarsi et al., 2022) According to (Hidayat, Qudratuddarsi, et al., 2021), person reliability elicits the stability of student responses in each instrument, while item reliability elicits the stability of item score. In the current study, it is found that person reliability 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). 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.

Table 2. Reliability and separation of the instrument

	Item	Person
Reliability	0.76	0.60
Item Reliability	1.80	1.23
Cronbach's Alpha	0.65	

3. *Data Analysis*

Analysis of data for this study was based on two methods namely traditional method and using item option characteristics curve (IOCC). In this paper, 2TMC has two tiers, where tier 1 or phenomenon tier will be written as Phen-1 till Phen-15, while tier 2 or reasoning tier will be written Rea-1 to Rea-15. In the traditional method, the analysis was similar to many previous studies using RSCRDI such as (Chandrasegaran et al., 2009, 2011). For misconception type-1, the answer on phenomenon tier was firstly determined, and if it was correct, then continued to see the alternative answer at reasoning tier. In simple manner, this analysis can be said as "correct at phenomenon tier only, while reasoning tier was incorrect". For instance, in item 1, there were 99 participants (53.51%) answer correctly, and 71.71% of them chose incorrect reasoning tier or categorized as misconception type-1. 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. This analysis was also conducted for misconception type-2 with the same methods as misconception type-1. Misconception type-2 is correct at reasoning tier, incorrect at phenomenon tier. This method has been applied since the introduction of the instrument by (Fetherstonhaugh & Treagust, 1992).

Another analysis is to reveal misconception by using IOCC based on guidelines from (Herrmann-Abell & DeBoer, 2011) which started by conducting ANOVA using SPSS version 25 and then followed by Winstep 3.73 to draw IOCC. Firstly, we conducted one-way ANOVA to ascertain there was a difference of ability in phenomenon and reasoning tier between first year, second year and third year of pre-service chemistry teachers. Before conducting

this analysis, raw data (ordinal scale) were firstly transformed into log odd unit or logit (interval scale) employing Winstep version 3.7.3. After considering assumptions test namely normality, homogeneity, one-way ANOVA was conducted. An analysis of variance showed that there is a difference of student achievement in phenomenon tier among the first year, second year and third-year students, $F(2,182)=70.624$, $p = 0.000$. The analysis showed that there is a difference of student achievement in phenomenon tier among the first year, second year and third-year students, $F(2,182)=27.897$, $p = 0.000$, and there is a difference of student achievement in phenomenon tier among the first year, second year and third-year students, $F(2,182)=27.897$, $p = 0.000$.

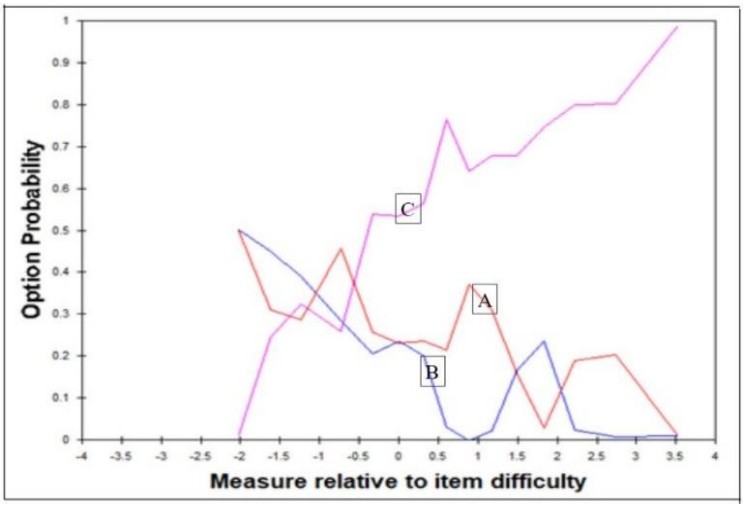

Figure 1. IOCC item Phen-1

To create IOCC, firstly respondents' 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 concept of students holding misconception. To analyze alternative answer based on the figure, the distractor which was consistently had a high probability for being selected along the curve was considered as alternative answer. For instance, the figure below is the IOCC of item phen1. 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 considered as alternative answer.

4. RESULTS AND DISCUSSION

Findings of analysing alternative answer based on the traditional method and IOCC are depicted in Table 3. It is interesting to note that some items in misconception type-2, we cannot decide the alternative answer for item Phen-11, Phen-12, Phen-14, Phen-15 due to the similar number of students who chosen the answers. However, by analyzing its curve along IOCC graph, we can find one stronger distractor. It means that IOCC analysis can inform more detailed analysis by providing student's performance along IOCC graphs.

Table 3. Alternative Answer

Item	Misconception type-1 (correct phenomenon, incorrect reasoning)		Misconception type-2 (correct reasoning, incorrect phenomenon)	
	Traditional method	IOCC	Traditional method	IOCC
1	3	3	A	A
2	1	1	A	A
3	3	3	C	C
4	2	2	B	B
5	2	2	C	C
6	3	3	B	B
7	3	3	B	B
8	2	2	C	C
9	1	1	B	B
10	2	2	C	C
11	2	2	B / C	B
12	3	3	B / C	C
13	3	3	B	B
14	3	3	A / B	A
15	2	2	A / B	A

Some other vital information, as the strengths of the analysis, after carefully observing IOCC of each item can be revealed. Some striking points from the IOCC analysis comprising the curve for 2-options (question Phen-2, Phen-9, and Phen-13), unexpected curve after 0 logits (Phen-4, Phen-6, Phen-15, Rea-1), the inconsistency of strongest distractor (Phen-8) and unworking distractor (Phen-3 and Rea-4).

Unexpected curve after 0 logits.

From this study, some unexpected curves were observed in item Phen-4, item Phen-6, item Phen-15 for phenomenon tier and item Rea-1 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 Phen-4, 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.

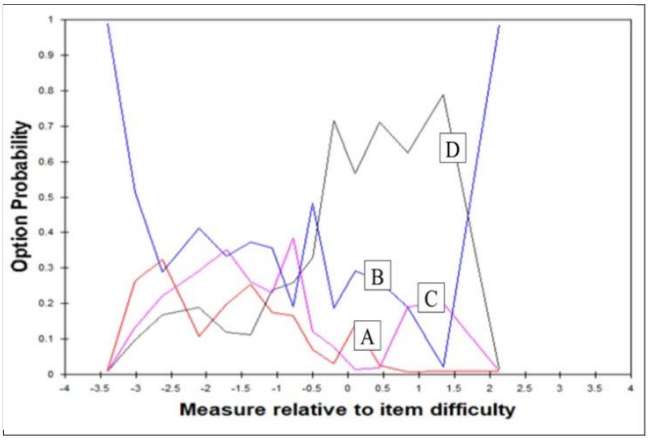

Figure 2. IOCC item Phen-4

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 Phen-15 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 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 & option B). Possibly, option A looks more scientific because they consider copper oxidation which refers to the loss of electrons.

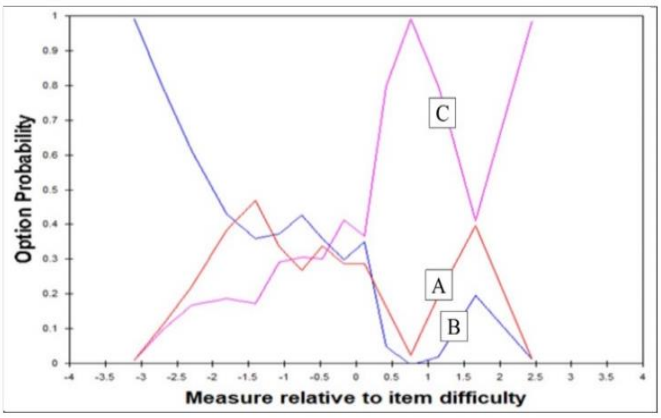

Figure 3. IOCC item Phen-15 (Phenomenon tier)

Less-functioned distractor

According to the IOCC analysis, there were two distractors can be identified as less-functioned distractor namely option A item Phen-3 and option 1 item Rea-4. In the figure, as an instance for 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.

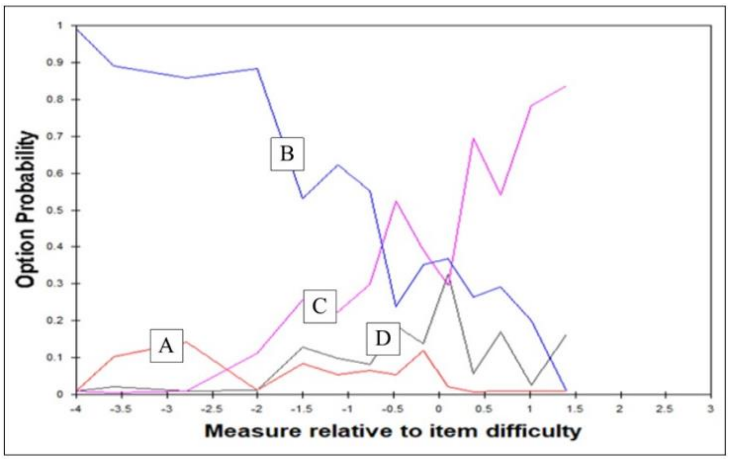

Figure 4. IOCC item Phen-15

From the result of 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. This result confirms the finding from (Wind & Gale, 2015) and indicating that the item functioned well to measure student abilities to explain and describe chemical reaction using multiple representations. 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 theoretically by guessing. This study can be considered as the extension of using IOCC in the diagnostics test. In the study from (Herrmann-Abell & DeBoer, 2011b; Herrmann-abell & Deboer, 2016), IOCC was used only to detect misconception. This application is further utilized by (Wind & Gale, 2015) to analyze data of difference of ability in pre-test and post-test. To continue, this framework is used to analyze distractors to show student conception.

Advantages of using Item Option Characteristics Curve

Utilizing IOCC to analyze 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. 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.

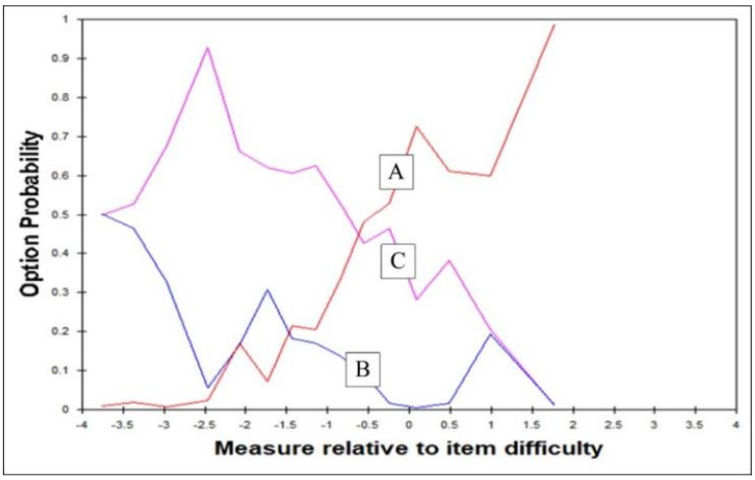

Figure Error! No text of specified style in document.**. IOCC item Phen-1**

The next advantage of using IOCC is in the determination of alternative answer, there were some items based on traditional method could not be decided one alternative answer. These were the alternative answer for item Rea-7, item Phen-11, item Phen-12, item Phen-14, and item Phen-15 because there were a fair number of selected two different options. For this difficulty, IOCC could decide which option could be stronger alternative answer. For instance, item Phen-15 with option A and option B. Based on IOCC at Figure, it was clear that the alternative answer was option A which has higher probability along the graph compared to option B. The same case was evident in some items such as item Phen-11, Phen-12, Phen-14 and Rea-7 in which traditional method get difficulties to measure one alternative answer, while IOCC can straight to choose single alternative answer for all the items.

The findings of these studies demonstrate that the Item-Objective Congruence Coefficient (IOCC) can be effectively used to analyze student misconceptions, as shown by previous research, including the studies by Wind and Gale (2015), Herrmann-Abell and DeBoer (2016), and Mulyani et al. (2021). These studies highlight the advantages of IOCC, particularly in comparison to traditional methods, when assessing students' understanding

of scientific concepts. The IOCC provides a more nuanced and powerful analytical tool for identifying gaps in students' knowledge, making it easier for educators to pinpoint specific areas of misconception. The effectiveness of this approach is theoretically underpinned by Item Response Theory (IRT), which enhances the accuracy and reliability of the analysis by considering the interaction between students' abilities and item difficulties. This combination makes IOCC a valuable method for improving the assessment of student understanding in various educational settings.

5. CONCLUSION

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. This study can be a new framework of analysing science test instruments.

BIBLIOGRAPHY

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.

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

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

Chan, M., & Subramaniam, R. (2020). Validation of a Science Concept Inventory by Rasch Analysis. In *Rasch Measurement* (pp. 159–178). Springer Singapore. https://doi.org/10.1007/978-981-15-1800-3_9

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

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.

Chew, S. L., & Cerbin, W. J. (2021). The cognitive challenges of effective teaching. *Journal of Economic Education*, *52*(1), 17–40. https://doi.org/10.1080/00220485.2020.1845266

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.

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

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

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

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

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

Herrmann-Abell, C. F., & DeBoer, G. E. (2011a). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chem. Educ. Res. Pract.*, *12*(2), 184–192. https://doi.org/10.1039/C1RP90023D

Herrmann-Abell, C. F., & DeBoer, G. E. (2011b). 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

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.

Hidayat, R., Idris, W. I. W., Qudratuddarsi, H., & Rahman, M. N. A. (2021). Validation of the Mathematical Modeling Attitude Scale for Malaysian Mathematics Teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(12). https://doi.org/10.29333/EJMSTE/11375

Hidayat, R., Qudratuddarsi, H., Mazlan, N. H., & Mohd Zeki, M. Z. (2021). EVALUATION OF A TEST MEASURING MATHEMATICAL MODELLING COMPETENCY FOR INDONESIAN COLLEGE STUDENTS. *Journal of Nusantara Studies (JONUS)*, *6*(2), 133–155. https://doi.org/10.24200/jonus.vol6iss2pp133-155

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

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.

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

Masson, S., Potvin, P., Riopel, M., & Foisy, L. 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

Mulyani, S., Haniza, N., Ramadhani, D. G., & Mahardiani, L. (2021). Rash model approach for analysis of misconception on chemistry learning with distractor analysis. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *6*(1), 98-107.

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

Qudratuddarsi, H., Hidayat, R., Shah, R. L. Z. binti R. M., Nasir, N., Imami, M. K. W., & Nor, R. bin M. (2022). Rasch Validation of Instrument Measuring Gen-Z Science, Technology, Engineering, and Mathematics (STEM) Application in Teaching during the Pandemic. *International Journal of Learning, Teaching and Educational Research*, *21*(6), 104–121. https://doi.org/10.26803/ijlter.21.6.7

Qudratuddarsi, H., Sathasivam, R. V, & Hutkemri, A. (2019). *Difficulties and Correlation between Phenomenon and Reasoning Tier of Multiple-Choice Questions: A Survey Study* (Vol. 3).

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

Resbiantoro, G., Setiani, R., & Dwikoranto. (2022). A Review of Misconception in Physics: The Diagnosis, Causes, and Remediation. *Journal of Turkish Science Education*, *19*(2), 403–427. https://doi.org/10.36681/tused.2022.128

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

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.). Routledge/Taylor & Francis Group.

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

Wild, D., Eremenco, S., Mear, I., Martin, M., Houchin, C., Gawlicki, M., Hareendran, A., Wiklund, I., Chong, L. Y., Cohen, L., & 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

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

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

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