

Development of an integrated ability assessment instrument in reaction rate material

Wan Zahara Zaitu Hania¹ and Hari Sutrisno²

¹Graduate School of Yogyakarta State University, Indonesia

²Department of Chemistry Education, Universitas Negeri Yogyakarta, Indonesia

Jl. Colombo No. 1 Yogyakarta

Email: wanzaharazaituhania@gmail.com

Abstract: This research aimed to develop a valid and reliable integrated ability assessment instrument (IAAI) to measure the multiple representations and chemical literacy of senior high school students related to the reaction rates. The framework developed in this research was implemented based on a 4-D model (defining, designing, developing, and disseminating). The integrated assessment consisted of five items of essay questions. The preliminary version of the IAAI was initially piloted on reviews by experts, and the paper-pencil test was given to a group of students. The obtained data were analyzed using Item Response Theory (IRT). In addition, Generalized Partial Credit Model (GPCM) model was conducted to examine the parameter quality, which estimated the test items. The content validity, construct validity, and quality of the items suggest that the integrated assessment ability instrument has a high score in validity and reliability. The analysis results of the instrument trial data indicate that the integrated abilities of students are in the medium category.

Keywords: *development, integrated assessment, reaction rate*

How to cite (APA 7th Style): Hania, W., & Sutrisno, H. (2022). Development of an integrated ability assessment instrument in reaction rate material. *Jurnal Kependidikan*, 6(1). doi:<https://doi.org/10.21831/jk.v6i1.38390>

INTRODUCTION

The rapid development of science and technology requires an appropriate learning process to support students in exploring their abilities (Plush & Kehrwald, 2014, p. 2). One step that can be taken to prepare students to face global challenges is to improve the learning process quality. In the Indonesian National Education System Law (Law Number 20 of 2003 concerning National Education System, p. 3), it is stated that learning is a process of interaction between students, educators and learning resources in a learning environment. The learning process is divided into three sequential stages, which are planning, implementing, and evaluating the learning process. Assessment is the last agenda that becomes the spearhead in the learning process that provides information about students' knowledge in learning about things that are already known and still needed by students, and helps teachers in forming and directing students in the learning process (Giles & Earl, 2011, p. 25). The assessment instrument is an integral part of an assessment process in learning which includes tests and an assessment system (Juita, Zulva, & Edial, 2019, p. 101). However, the assessment used in some schools in Indonesia still measures the memorization aspect which results in the lack of students' competencies development monitoring during learning. Therefore, an instrument is needed to assess and train every skill to students in the field of science, especially chemistry. Chemical concepts that are presented in various levels of

representation are very important to be mastered by students, because the four levels are interrelated with one another and can construct students' understanding to link concepts to one and another.

According to the study conducted by Ainsworth (2004, p. 191), learning chemistry should be taught by linking all levels of multiple representations so that students can understand chemistry at all. The ability of multiple representations have four important aspects that are interconnected with each other, to understand the chemical concepts that will be applied in solving phenomena and problems in their daily lives, so that the learning process undertaken by students becomes more meaningful. Therefore, there is a relationship between multiple representations and chemical literacy, because the National Institute for Literacy (1992) defines literacy as the ability of individuals to read, write, speak, count, and solve problems according to the expertise level required in work, family and society (daily life). In order to increase students' understanding and practice the various abilities they have, especially in the multiple aspects of chemical representation and literacy, it is necessary to do a process of assessment of the aspects to be trained.

A good assessment process requires an assessment instrument that can measure and train students' understanding and ability (Sadhu & Laksono, 2017, p. 6), such as multiple representations and chemical literacy, so students can face daily life problems with chemistry concepts that they have understood by integrating both of these capabilities. Assessment is expected to be able to measure overall student competency using valid and reliable tests (Laksono, Rohaeti, Suyanta & Irwanto, 2017, p. 109). But, nowadays the cognitive aspects assessment on some abilities is still largely done separately. In fact, if there is a connection between these abilities, it would be better if assessed together (integrated) in an instrument, because it can minimize the work of teachers in making instruments and assessing (Sadhu & Laksono, 2017, p.7). The ability of students' multiple representations and chemical literacy have to be measured and trained on appropriate chemical content. Chemistry topics taught in secondary schools need to be packaged in a grading system that is able to combine these two abilities to be well expressed. One of the chemistry topics in high school that is closely related to daily life and also requires the ability of multiple representations to achieve a good understanding of chemical concepts is the reaction rate. Reaction rate materials include subtopics which are the concept of reaction rate, factors that affect reaction rate, collision theory, rate law and reaction order (reaction rate).

Johnstone (2000, p. 34) states that the difficulties often faced by students in learning Chemistry actually combine three levels of Chemistry representation consisting of macroscopic, microscopic, and symbolic. Fahriyah and Wiyarsi (2017, p. 198) also added that these three abilities have to also be supported by mathematical abilities so that students can understand the concept of chemistry at all. According to Smith and Nakhleh (2011, p. 35), representation is "the meaningful structure of something: a word for an object, a sentence for a state of things, a diagram for an arrangement of things, an image for a scene" so that representation can be defined as something used to represent things, objects, circumstances, and phenomena. To design a good instrument, it requires learning indicators that refer to aspects raised by the scientists developing the theory. These aspects are then concluded to produce an indicator of learning. Johnson (2000, p. 33) suggests that there are four aspects of the ability of multiple chemical representations, which are the macroscopic, submicroscopic, symbolic and mathematical aspects.

The quality parameter of a nation's progress is greatly influenced by the level of literacy that is owned by students and the human resources in it. One supporting factor is the ability of chemical literacy. Chemical literacy has become an important issue to be discussed in recent decades, making students aware of the benefits of literacy, becoming a major goal for educators, scientists, and curriculum policy makers (Martinez-Hernandez, K., Ikpeze, C., & Kimaru, 2015, p. 8). National Institute for Literacy (1992), defines literacy as the ability of individuals to read, write, speak, count and solve problems based on the expertise level required in work, family and society. Schwartz, Ben-Zvi, and Hofstein (2006, p. 206) suggested that there are four aspects in the field of chemical literacy, which are knowledge of chemical materials and scientific ideas, chemistry in context, higher learning skills, and affective aspects. Furthermore, strengthening of literacy in students has been proven to strengthen students' understanding of concepts so that they can be easily applied in their lives (Avikasari, Rukayah, & Indriayu, 2018, p. 231).

Item response theory (IRT) is a study of questions based on the test parameters and the ability of test takers (Rudner, 2001, p. 231). IRT has 3 measurement models, which are: one-parameter logistics model, two-parameter logistics model, and three-parameter logistics model. Item response theory (IRT) is a review of questions by using answer questions. IRT explains the relationship between opportunities to answer a problem correctly with the ability of the test takers underlying it in mathematical functions. The concept of item response theory is based on two postulates. *First*, the test participant's achievement in a problem can be predicted by a group of factors called latent trait or ability. *Second*, the relationship between the test participant's achievement on a problem with the underlying ability described by a monotonous ascending functions called Item Characteristic Function or item characteristic curves/ICC. There are three assumptions that have to be fulfilled before continuing the analysis by modeling using item response theory, which are the uni-dimensional assumption, local independence and parameter invariance. Some theoretical response models in the form of politomus items that are often used by experts include the *Graded Response Model (GRM)*, *Partial Credit Model (PCM)*, and *Generalized Partial Credit Model (GPCM)*.

GPCM is a common form of the PCM model, with a 2 PL approach for scoring politomus data. Various studies have been conducted to determine the accuracy of the use of these models. One of them is Dodd (1991, p. 6) who analyzes the GRM and GPCM models for attitude measurement (de Ayala, 1993, p. 233). These models are used to analyze teacher attitudes towards communication skills and analysis of non verbal reasoning abilities. In addition, the application of GPCM is suitable with the condition of scoring the response of test items in the form of description in the field, the score of the points is given based on the number of steps that are answered correctly regardless of the sequence of steps.

METHOD

Students who participated in this study consisted of students in class XI IPA high school (for the measurement phase) and class XII (for the instrument trial phase). The trial phase of the integrated assessment instrument was conducted on 370 students of class XII IPA students in Pekanbaru City, Riau, Indonesia who had the highest, medium and lowest national exam scores. They are SMAN 8, SMAN 9, SMAN 10, SMAN 6, and SMAN 15. The choice of instrument trial subjects of 370 test takers due to the use of a relatively small population size

would be able to produce an inaccurate parameter estimate value, even for a population size that was too small, it was not possible to conduct the parameter estimated value.

The items were constructed based on the contents of Chemistry Curriculum 2013 applied in Indonesia. In order to elicit the desired multiple representation and chemical literacy from the students, each item of indicator is designed in such way that provoked multiple representation and chemical literacy together and describes in Tabel 1.

Table 1. The distribution of representation and literacy aspects in the test indicators

Number of Indicator	Content	Aspect of Literacy				Aspect of Representation			
		L1	L2	L3	L4	Ma	Sub	Sim	Mat
1	Collision Theory	√	√	√		√	√	√	√
2	The concept of reaction rate	√	√	√	√	√	√	√	
3	Factor affecting the rate of reaction	√	√	√	√	√	√	√	
4	Reaction order	√	√	√	√	√	√	√	√
5	Constants and equations of reaction rates	√	√	√		√	√	√	√

Explanation:

L1 = Literacy Aspect 1

L4 = Literacy Aspect 4

Sim = symbolic

L2 = Literacy Aspect 2

Ma = macroscopic

Mat = mathematics

L3 = Literacy Aspect 3

Sub = sub-microscopic

Data collection process for the development of integrated assessment instruments was in accordance with the predetermined product preparation procedures based on reviews and recommendations from expert judgment. Data collection on the characteristics of integrated assessment instruments in this study was conducted through a trial of instruments involving 370 students of class XII IPA SMA in five high schools in Pekanbaru City, Riau Province, Indonesia. Data collection techniques used in this study were in the form of tests and non-tests, which included questionnaires (item validation sheets), interviews, and written tests.

Data analysis techniques used were qualitative and quantitative analysis. Qualitative analysis was used to describe the product development process. Meanwhile, quantitative analysis was used to determine the integrated instruments produced characteristics, where the data to be analyzed was the data of instrument validation and trial results. The data that had been obtained regarding the process of developing integrated assessment instruments was descriptive data that was described according to the product development stages undertaken. Qualitative analysis included consideration of content validity and constructs by experts. In addition, quantitative data analysis included an analysis of the instrument construct validation results, the results of an instrument trial that included the assumptions results of the item response theory assumptions and the results of the item characteristics, and the results of integrated capability measurements. The process of analyzing data validation and testing of instruments used the Winstep application (for test assumption), SPSS (for test assumption), and the R Program (for item parameter). While the process of analyzing data was the result of integrated capability measurement using Microsoft Excel application.

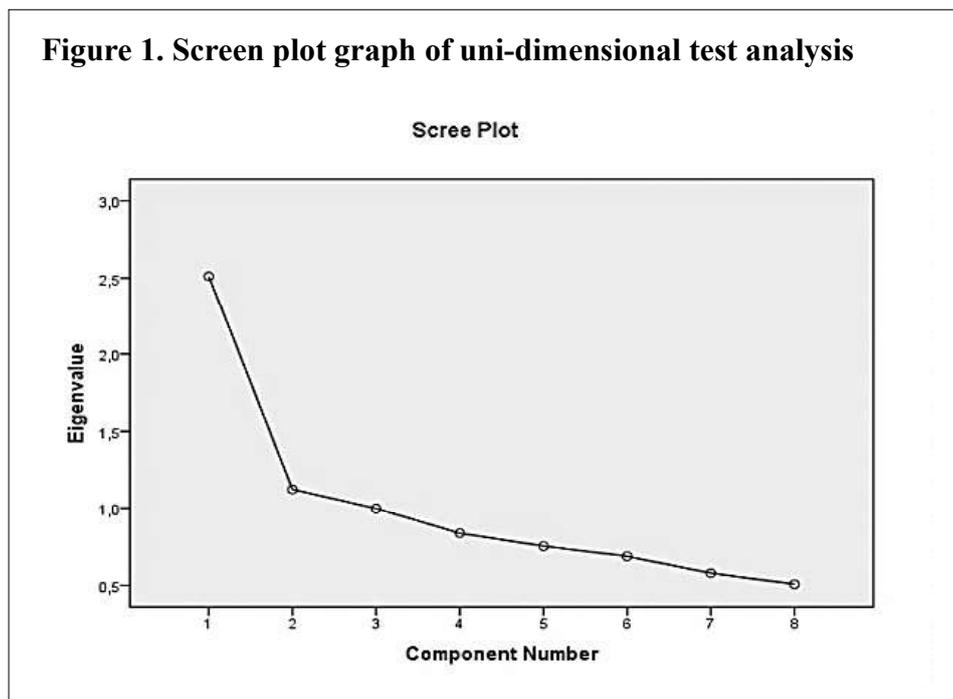
FINDING AND DISCUSSION

The expert judgment trial was aimed at producing valid instruments in terms of content, so that it can be used to measure chemical literacy skills and multiple chemical representations of students. Qualitative study was conducted before the trial of instruments and measurements involving expert lecturers and high school chemistry teachers. The aspects assessed at this stage included the substance, construction, language and appearance of the initial product of the integrated assessment instrument. The results of the validation from the experts were then described in a qualitative descriptive way to be concluded in order to improve to the next stage. Suggestions that had been obtained from the results of expert studies on each item were improved to obtain a valid initial integrated assessment product. According to the evaluation results of the items obtained questions that are feasible to be used, there were 8 items to be tested with the time allocation of 2 hours of learning (90 minutes) on a small scale test. Meanwhile, five valid items would be used to be tested in the measurement test phase.

The assumption test of item response theory was conducted before conducting the analysis using GPCM modeling with the 2-PL approach. The response model in item response theory has the meaning that the subject's probability of answering the item correctly depends on the subject's ability to answer the item correctly based on the subject's ability and item characteristics. This can be interpreted that students with high abilities will have a greater probability of answering correctly when compared to students who have lower abilities. There are three main assumptions that underlie item response theory, the three basic assumptions tests that have to be met are uni-dimensional assumptions, local independence, and parameter invariance (Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991).

Uni-dimensional assumption test was performed by exploratory factor analysis using the Statistical Package for the Social Science (SPSS) Version 17.0 program, taking into account the Eigen value and scree plot in the variance of the analysis results in the SPSS program. If the difference in Eigenvalues between components 1 and 2 is 4 or 5, it can be said that the test instrument measures a single or uni-dimensional ability. Meanwhile, the number of components measured in the test instrument is seen from an Eigenvalue which is equal to 1. The results of the factor analysis indicate the validity of a construct or empirical validity. However, there are requirements that have to be met before conducting a factor analysis, which are by testing a sample eligibility using the Kaiser-Mayer-Olkin Measure of Sampling Adequacy (KMO-MSA) test and the Bartlett test. Factor analysis conducted in this study uses exploratory factor analysis approach by looking at the value of KMO.

If the KMO value is lower than 0.50, then the exploratory factor analysis approach cannot be conducted (Yilmaz, Altinkurt, & Cokluk, 2011, p. 347). Meanwhile, the use of the Barlett test serves to determine whether or not there is a correlation between the variables of the instrument being developed. The KMO-MSA value obtained was 0.735 and the significance value of the Barlett test <0.05 . This indicates that the sample used was sufficient and the exploratory factor analysis approach could be conducted for the next stage, which is the construct validity or unidimensional test. Furthermore, the uni-dimensional test assumption was conducted using factor analysis. Factor analysis aims to identify the relationship between variables by looking at the eigen value in the inter-covariant variance matrix. The results of the uni-dimensional assumption test analysis can be strengthened by the scree plot graph shown in Figure 1.



The screen plot is able to clarify the eigen value visualization with the number of components retained as a factor. Figure 1 shows the amount of steepness based on eigenvalues. According to the figure, the uni-dimensional assumption has been fulfilled. Besides that, the second assumption that has to be fulfilled is the local independence assumption. The analysis results of local independence tests can be calculated using a variance-covariance matrix of the person measure based on the students ability from each of the test instruments tested (Greiff et al., 2013, p. 368). Local independence test is considered fulfilled if the value below the diagonal line in the variance-covariance matrix is zero. The zero value indicates that the students 'skills in answering one item does not affect the students' skills in answering another item. In other words, each item is independent. The third assumption is the parameter invariance assumption. The parameter invariance assumption test consists of item parameter and capability invariance. The results of the analysis can be concluded that students' abilities are invariant.

Item analysis in this study used the GPCM model (PCM – 2 Parameter Logistic). The model is used to estimate the item parameters in the instrument which includes the item suitability, item difficulty level, the test information function and the test reliability of the product that has been developed. Item analysis was conducted to determine items fit with the test participant response pattern based on the GPCM model. The items developed were 8 items consisting of multiple representation and chemical literacy indicators. The item conformity index determines whether an item is functioning optimally and meets the requirements as a good measurement tool. The item is said to be fit with the model if it meets the criteria for the value of means-square outfit, z-standard outfit and point measure correlation (Bond & Fox, 2015, p. 213).

Quantitative analysis was conducted through GPCM modeling using the Winstep program. The item calibration has been conducted to determine whether the items are proper

with the GPCM or PCM-2PL model. According to Sumintono and Widhiarso (2015) if the test items meet two of the three criteria for acceptance of the MNSQ, ZSTD, or Pt Mean Corr limits, then the items can be considered as suitable for the model and fit for further analysis. Outfit-MNSQ (mean square) received outfit values range from 0.5 to 1.5. Outfit Z-STD received values range from -2.0 to +2.0. And the value of Pt Mean Corr had ranges from 0.4 to 0.85. The results of the conformity of the items with the partial credit model of two logistical parameters are presented in Table 2. According to Table 2, it appears that each item has fulfilled two of the three criteria for acceptance of the MNSQ, ZSTD, or Pt Mean Corr limits. According to this it can be concluded that each item is fit with the model, then all items can be said to be the final product of integrated assessment ability instruments (IAAI) that have been empirically valid.

Table 2. Item fit analysis results

Item Number	Outfit MNSQ	Outfit ZSTD	Pt Measure Corr	Criteria
1.	0.83	-2.6	0.53	<i>Fit</i>
2.	0.89	-1.7	0.57	<i>Fit</i>
3.	0.97	-0.4	0.55	<i>Fit</i>
4.	1.10	1.4	0.57	<i>Fit</i>
5.	1.21	2.6	0.51	<i>Fit</i>
6.	0.97	-0.4	0.51	<i>Fit</i>
7.	1.04	0.6	0.51	<i>Fit</i>
8.	1.09	1.3	0.55	<i>Fit</i>

The items characteristics to be analyzed next are the difficulty level parameters and the different power of the questions. Difficulty level parameters describe the estimated difficulty level of an item stated in a logit. The item difficulty level is one of the factors that affect the probability of students' answers in responding to certain items. Items that have a high degree of difficulty require a high ability to be answered correctly, while items that considered easy have a low level of difficulty which required a low ability to be answered correctly. Quantitative analysis is conducted through GPCM modeling using the Winstep program. The results of the item difficulty level and the different power with the GPCM model (partial credit 2 logistical parameters) are presented in Table 3.

According to the analysis of the items using the Winstep program, the items obtained were feasible to be used in the measurement test phase. It is due to all items have a good level of difficulty, which ranges from -2.0 to +2.0 logit. The difficulty level analysis results based on the R program output that have been obtained can be clarified in Table 3. Hambleton and Swaminathan (1985) and Baker (2001), an item is considered good if it has a difficulty index ranging from -2.0 to +2.0 logit. This indicates that the items in the integrated assessment instrument tend to have medium level questions. In addition, the results of the analysis of the different power indexes described in Table 3 (symbolized by a), overall located at 0.323-0.692 indicate all items have good differences. This refers to the criteria for good items according to Ebel and Frisbie (1991) which states that the power parameters differ (a) items with different power ranges from 0.39 to 0.40 or > 0,25 have very good difference power.

Table 3. Analysis results of the R program for difficulty & difference parameters

Item Number	a1	b1	b2	b3	b4	C
B1	0.646	-3.22	-1.39	0.443	1.565	0
B2	0.692	-1.47	-0.58	-0.45	1.601	0
B3	0.576	-1.35	-1.93	-0.46	0.185	0
B4	0.505	-1.26	-0.77	-0.33	0.004	0
B5	0.348	-1.31	-1.82	-0.74	-1.32	0
B6	0.403	-0.79	-2.58	0.734	1.788	0
B7	0.323	-1.34	-0.90	0.581	2.025	0
B8	0.423	-1.00	-1.11	0.435	0.582	0

The items that have the highest difference in power are point 2 and the lowest is item 7. The analysis results of item difficulty level is shown in Tabel 4. Therefore, based on the program output R of the difficulty level parameter and the different power of the items, all items in the integrated assessment instrument are included in the good category.

Table 4. Analysis results of item difficulty level (b)

Item Number	Value of Measure	Description
1	-0,52276	Medium
2	-0,179	Medium
3	-0,70849	Easy
4	-0,48814	Medium
5	-1,08894	Easy
6	-0,18726	Medium
7	0,061404	Difficult
8	-0,23603	Medium

Characteristics of the items seen also based on the analysis results of the test information function. The test information function describes the accuracy of an item on the test instrument and illustrates how much the item's contribution in expressing the latent ability of students (latent trait) is measured using a test instrument that has been developed (Retnawati, 2014, p. 32). Latent abilities are hidden abilities that cannot be seen directly but have the potential to emerge. Gruijter and van der Camp (2014) state that the value of the test information function depends on latent skills. Therefore, the value of the information function will vary according to students' skills. The test information function is the sum of the item information functions in the test (Hambleton & Swaminathan, 1985, P. 335). Linn and Gronlund (Rosaroso, 2015, p. 372) state that generally reliability is defined as measurement consistency. Reliability provides consistency that makes the validity of an instrument said feasible. In this study, the reliability consistency approach is used in the estimation of reliability, which is conducted with one test given once to a group of subjects at the trial stage

or empirical instrument validation. The advantage of this approach is practical and efficient, because it is only conducted once. According to the results of data input on the Winstep program, reliability estimations and tests can be obtained directly. The item reliability value obtained was 0.95 (high) and person reliability was 0.69 (medium).

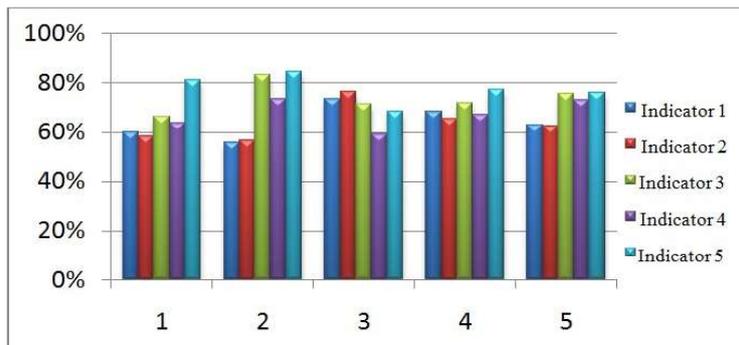
The instruments that have been tested on students in the instrument test phase produce instruments that are valid and reliable empirically, and have known the characteristics of the instrument through the previous stages. The next stage is the preparation of the instrument into the final product that is ready to be trialed at the measurement stage to the subject, which are class XI IPA students in five high schools in Pekanbaru in 2019 academic year. At this stage, measurement data obtained based on students' responses in answering items on the integrated assessment instrument are analyzed according to the assessment rubric that has been made.

The answer analysis is conducted by giving a value to each answer that has been written by students on the answer sheet that has been provided. The measurement results are then analyzed descriptively to find out the students achievement level integrated to chemistry skills mastery. The entire score of measurement results in each school is then analyzed descriptively and converted to a category of integrated skills mastery achievement level in the reaction rate material as well as the percentage of students mastery in understanding the material on each competency indicator that is reflected in each item. The percentage of achievement of integrated skills in each school is different, but if the score is categorized it results in the same achievement category, which is considered as medium category. In addition, Figure 2 outlines the level of achievement of integrated skills for each indicator at each school level.

Overall, it can be seen that in schools with low, medium and high level of learners' abilities the lowest integrated skills achieved in almost the same with competency indicators, which are in competency indicators 1 and 2. Schools, in order, are listed in Figure 2, they are SMAN 6, SMAN 8, SMAN 15, SMAN 10, and SMAN 9. The two lowest indicators are explaining the real phenomena that can be captured by the five senses using the concept of collision theory owned (indicator 1) and explaining the real phenomena that can be captured by the five senses using the concept of reaction rates that are owned (indicator 2). These two indicators are extensions of the collision theory and the reaction rate concept in the reaction rate chapter. These two indicators respectively have an average for each school in general of 65.85% and 69.4%, and considered as medium category. In the first indicator, students are required to be able to analyze the cause of phenomena in daily life using the collision theory that is owned by linking it to the representation image. In the text of the questions presented two different types of collisions and students are asked to analyze the collision that is effective and the factors that cause the collision of the two collision illustrations. According to question number one, it is known that most students have errors in answering the factors that affect collisions, as shown in the Figure 3.

According to Figure 3, it can be seen that the students' answers at points b and c have errors. At point b, the students answer the command about the factors that affect collisions with the answers to the factors that affect the reaction rate, and show students lack of literacy aspect 1 mastery and aspects of macroscopic representation. Besides that, at point c students are also not quite right in determining the concentration of NO and O₂ after reacting, and this indicates that the students lack of mathematical representation aspect mastery, which is one aspect of representation that is needed to be mastered by students at most chemical material.

Figure 2. Integrated skills achievement diagram of competency indicators based on school level (The indicators according to Table 1)



Figures 3. Examples of students' answers to question number 1

a) Jenis tumbukan I, karena tumbukan tersebut menghasilkan gas reaksi
 $2NO(g) + O_2(g) \rightarrow 2NO_2(g)$
 b) 1. Pengaruh konsentrasi & luas permukaan
 2. pengaruh suhu
 3. konsentrasi
 4. Pengaruh katalis
 c) $v = k \cdot [NO]^2 \cdot [O]$
 $12 = k \cdot (x)^2 \cdot (x)$
 $12 = k \cdot x^3$
 $2 \cdot 20 = x$
 d) Jenis tumbukan I, karena semakin tinggi temperatur maka semakin tinggi energi kinetik dari partikel reaktannya sehingga frekuensi tumbukan & energi tumbukan meningkat. Oleh karena itu, semakin tinggi suhu, laju reaksi juga makin cepat.

(a)

a) Jenis tumbukan I, karena akan terbentuk zat baru
 atau $O = N - O$ atau NO_2
 b) Konsentrasi dan Jumlah energi, arah, dan frekuensi
 c) $12 = (2)^2 \cdot (3)$ $[NO] = 2$
 $[O_2] = 3$
 d) Gambar a, karena jika suhu tinggi partikel zat akan bertambah cepat karena jika suhu tinggi maka energi kinetik bertambah

(b)

This is supported by research conducted by Tima and Sutrisno (2018, p. 11) that the study of chemistry must involve macroscopic, submicroscopic and symbolic representations, so when students learn, it will connect the three representations. Fahriyah and Wiyarsi (2017, p. 199) also states that the macroscopic, microscopic, symbolic and mathematical aspects in multiple chemical representations have to be supported by mathematical aspects so that students can understand chemistry more thoroughly. In addition, according to Figure 3a it can be seen that the students have been right in answering the four question commands contained in question number 1. Different from the answers in Figure 3a, students who answered the question items in Figure 3b have correctly answered items b and c, which is indicated by the answers to factors that affect collisions that are answered by students is influenced by energy, direction and frequency of collisions, but these answers have not been accompanied by a more complete explanation. Correspondingly, the students' answers to item c have

been answered correctly, that is the increase in NO and O₂ concentrations is 2 and 3 times initially. This indicates that the students have mastered the mathematical, macroscopic and literacy aspects 1 which are integrated in the question number 1.

In line with the integration of several aspects of chemical literacy and multiple representations in a question instrument, it is important to conduct the ability to train chemical literacy and multiple representations of students (Odja & Payu, 2014, p. 46), and can increase students' understanding, especially in some sufficient chemical materials which is difficult to understand. Furthermore, the broader of context would allow the students to make a better connection among the chemistry concept to their daily life that brings an opportunity to investigate their own hypothesis perceived by themselves (Wiyarsi, Damanhuri, & Fitriyana, 2020, p. 192). This is in line with the statement of Ainsworth (2004, p. 191), which states that chemistry learning should be taught by linking all levels of multiple representations so that students can comprehend chemistry as at all and chemical learning that has been taken by students in the learning process becomes more meaningful, which certainly can be more easily applied by students in daily life in order to balance the progress of science and technology which is rapidly developing.

CONCLUSION

The importance of training the students' ability of multiple representations and chemical literacy, especially on some chemical materials, requires the existence of a valid and reliable assessment instrument. Integrated assessment instruments to identify the ability of chemical literacy and multiple chemical representations of High School students in class XI IPA have gone through a trial process to expert judgment and a set of instruments to be used in this study have been deemed feasible by all three expert lecturers or expert judgment with various suggestions or improvements. The quality characteristics of the integrated assessment instruments of this study results were reviewed from the construct validity results of the uni-dimensional assumptions, followed by two other assumptions which are local independence and parameter invariance, as conditions for the use of IRT. IRT modeling used in this study is the GPCM with a 2 PL approach. According to the criteria of the lowest and highest limit of the mean-square (MNSQ) outfit obtained, 5 test items were proven to be in accordance with the GPCM (2-PL) model. In addition, the results of the parameter analysis show that the level of items' difficulty was classified as good because it is in the range of -2.0 - +2.0 logit and is spread evenly on each item. The difference in power overall lies in 0.323 - 0.692 indicating that all items have good difference because it is > 0.25. The results of the measurement of chemical literacy skills and the multiple chemical representations of class XI high school students using an integrated assessment instrument showed that the students' skills were classified as moderate with a percentage of 69.28%.

REFERENCES

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representation. *Learning and Instruction*, 3(16), 183-198. <https://doi.org/10.1016/j.learninstruc.2006.03.001>
- Avikasari, Rukayah, & Indriayu, M. (2018). Keefektifan penggunaan bahan ajar science literacy terhadap peningkatan prestasi belajar. *Jurnal Kependidikan*, 2(2), 221-234. <https://doi.org/10.21831/jk.v2i2.19167>

- Bond, T. G., & Fox, C. M. (2015). *Applying the GPCM model fundamental measurement in the human science* (3th ed.). Routledge.
- De Aayala, R. J. (1993). An introduction to polytomus item response theory models. *Measurement and Evaluation in Counseling and Development*, 25, 172-189.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Publications.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essential of educational measurement* (5th ed.). Prentice-Hall.
- Fahriyah, A. R., & Wiyarsi, A. (2017, 12-13 September). Multiple representation skill of high school students on reaction rate material. Paper presented at *Conference: International Seminar on Chemical Education* (2nd ISCE), UII, Yogyakarta, Indonesia.
- Gilbert, J. K. (2006). On the nature of "context" in chemical education. *International Journal of Science Education*, 28(9), 957-976.
- Giles, D., & Earl, K. (2014). Being "in" assessment: The ontological layer(ing) of assessment practice. *Journal of Applied Reaserch in Higher Education*, 6(1), 22-29. <https://doi.org/10.1108/jarhe-01-2012-0001>.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item respon theory*. Kluwer Inc.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item respon theory*. Sage Publication, Inc.
- Martinez-Hernandez, K., Ikpeze, C., & Kimaru, I. (2015). Perspectives on Science Literacy: A comparative study of United States and Kenya. *Educational Research International*, 4(2), 25.
- Johnstone, A. (2000). Chemical education reaserch: Where from here? *University Chemistry Education*, 4(1), 34-48.
- Juita, E., Zulva, R., & Edial, H. (2019). Pengembangan perangkat asesmen pembelajaran geografi bencana untuk meningkatkan kemampuan berpikir kritis. *Jurnal Kependidikan*, 3(1), 85-105. DOI: <https://doi.org/10.21831/jk.v3i1.18623>.
- Laksono, E. W., Rohaeti, E., Suyanta, & Irwanto. (2017). The evaluation instrument of analytical thinking and science process skill in chemistry subject. *Jurnal Kependidikan*, 1(1). 100-110. <https://doi.org/10.21831/jk.v1i1.8205>
- Ministry of Education and Culture Indonesia. (2013). *Regulation of ministry of education and culture number 66 in 2013 regarding the standard for education assessment*. Ministry of Education and Culture Indonesia.
- Plush, S. E., & Kehrwald, B. A. (2014). Supporting new academics' use of student centred strategies in traditional university teaching, *Journal of University Teaching & Learning Practice*, 11(1), 1-16. Retrieved from <http://ro.uow.edu.au/jutlp/vol11/iss1/5>.
- Rosaroso, R. C. (2015). Using reliability measure in test validation. *European Scientific Journal*, 11(8), 367-377. <https://doi.org/10.1166/asl.2017.990>
- Rudner, L. M. (2001). *Item respon theory*. Educational Resources Information Center, USA.
- 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>
- Schwartz, Y., Ben-Zvi, R., & Hofstein, A. (2006). The use of scientific literacy taxonomy for assessing the development of chemical literacy among high-school student. *Chemistry Education Reaserch and Practice*, 7(4), 203-225. <https://doi.org/10.1039/B6RP90011A>.

- Smith, K. C., & Nakhleh, M. B. (2011). University students' conceptions of bonding in melting and dissolving phenomena. *Chemistry Education Research and Practice*, 12(4), 398-408. <https://doi.org/10.1039/c1rp90048j>
- Tima, M. T., & Sutrisno, H. (2018). Influence of problem solving based on multiple representations model on teaching and learning of chemistry on student's academic self-efficacy and student's cognitive achievement. *American Journal of Educational Research*, 6(7), 887-892. Retrieved from <http://pubs.sciepub.com/education/6/7/1>.
- Wiyarsi, A., Damanhuri, M. I. M., & Fitriyana, N. (2020). Exploring pre-service chemistry teachers' problem solving skills on vocational context-based learning. *Jurnal Kependidikan*, 4(2), 185-199. Retrieved from journal.uny.ac.id/index.php/jk/article/view/35302/pdf.
- Yilmaz, K., Altinkurt, Y., & Cokluk, O. (2011). Developing the educational belief scale: The validity and reliability study. *Educational Sciences: Theory & Practice*, 11(1), 343-350.