JPPIPA 10(1) (2024)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Development of An Integrated Assessment Instrument to Measure Students' Critical Thinking and Chemical Literacy Skills for Rate of Reaction Topic

Nurlaila Widyanarti Ariefiani[1*], Endang Widjajanti Laksono[2]

[1] Master Student of Chemistry Education, Faculty of Mathematics and Science, Yogyakarta State University, Yogyakarta, Indonesia.
[2] Department of Chemistry, Yogyakarta State University, Yogyakarta, Indonesia.

Received: August 4, 2024
Revised: September 27, 2024
Accepted: October 26, 2024
Published: October 31, 2024

Corresponding Author:
Nurlaila Widyanarti Ariefiani
nurlailawidyanarti.2021@student.uny.ac.id

DOI: 10.29303/jppipa.v10i10.9069

Abstract: Critical thinking is one of the 21st-century skills needed. Understanding chemical literacy can encourage students to be able to think more critically. Given the importance of critical thinking and chemical literacy in chemistry learning, there is a need for an assessment instrument that can measure both abilities. Based on this, the integrated assessment instrument was developed to determine the characteristics and feasibility of the integrated assessment instrument that can measure students' critical thinking skills and chemical literacy on reaction rate materials. The sample in this study is 154 students from two high schools in Makassar City. The feasibility and characteristics of the integrated assessment instrument were analyzed using the Rasch model with the Partial Credit Model-1 Logistic Parameter (PCM-1PL) approach. The results of the analysis show that the 10 items developed are valid and have a high reliability value. The integrated assessment instrument has the characteristics of fit with the model, and the difficulty level of the items is classified as good. The results of this study show that the integrated assessment instrument can be used as a good instrument to measure students' critical thinking skills and chemical literacy on reaction rate materials.

Keywords: Chemical Literacy; Critical Thinking; Integrated Assessment Instrument

Introduction

Along with the times that are leading to the Industrial Revolution 4.0, humans are required to have 21st-century life skills (Hayani, 2019). Critical thinking is one of the 21st-century life skills and is included in the classification of the *Higher-Order Thinking Skills* (HOTS). The ability to think critically is not only limited to the ability to memorize facts or concepts but how a person's ability to process the facts or concepts obtained (Anagün, 2018; Brookhart, 2010). Critical thinking is considered a complex thinking process that requires a

high level of cognitive reasoning in processing information (Choy & Cheah, 2009). This is a strong reason for the important role of critical thinking skills in supporting students' success in learning (Mahanal *et al.*, 2019). Therefore, efforts to improve critical thinking skills are something that really needs to be paid attention to (Sutiani, 2021).

Critical thinking is essential in chemistry learning. Chemistry is an important basic science to learn because it is closely related to various aspects of our daily lives. An understanding of chemistry leads to an understanding of the world, especially the phenomena that occur around us (Dori et al., 2013; Koballa et al., 2000). Understanding chemistry is certainly not obtained immediately but must be developed through literacy activities so that in chemistry learning, the achievement

How to Cite:
Ariefiani, N. W., & Laksono, E. W. (2024). Development of An Integrated Assessment Instrument to Measure Students' Critical Thinking and Chemical Literacy Skills for Rate of Reaction Topic. *Jurnal Penelitian Pendidikan IPA*, *10*(10), 7726–7734. https://doi.org/10.29303/jppipa.v10i10.9069

of chemical literacy is crucial to be reviewed as one of the achievements of chemistry learning outcomes (Sari et al., 2019).

Chemical literacy includes the chemical knowledge and skills needed to understand chemistry and socio-scientific issues. The understanding of chemistry includes an understanding of the basic concepts of chemistry as well as an understanding of placing chemistry in the real-world context. With chemical literacy, students can better understand the role of chemistry in life (Kohen et al., 2020). In addition, by having chemical literacy, students can apply the chemistry topics they have learned to solve relevant problems and can even be applied in integrating several chemical concepts that are interconnected with real situations in daily life (Jagger & Yore, 2012).

Critical thinking skills and chemical literacy students in Indonesia are still relatively low. The low level of critical thinking ability can be seen based on TIMSS data (*Trends in International Mathematics and Sciences Study*) in 1999, 2003, 2007, and 2011 which shows that Indonesia consecutively ranked 32nd out of 38 countries, 37th out of 46 countries, 35th out of 49 countries, and 40th out of 42 countries. Based on this, it shows that students have not been able to solve problems that require the use of high-level thinking processes (Suardana et al., 2018). In addition, for the achievement of chemical literacy which is part of the achievement of science literacy in general, by looking at survey data conducted by PISA, Indonesia's science literacy score in 2022, which is 366, has decreased compared to the science literacy score in 2018 which originally obtained a score of 383 (OECD, 2023). The data shows that science literacy skills among students in Indonesia are still relatively low and still need to be improved in learning (Pranomy *et al*., 2021).

Seeing the importance of critical thinking skills and chemical literacy in the success of chemistry learning, these two skills must be developed for students. Every teacher in Indonesia should be responsible for helping students develop their abilities.

One effort that teachers might make is to create learning assessment instruments that can be used to measure students' critical thinking skills and chemical literacy. The assessment instrument created must be able to be used to measure students' abilities objectively and can be used as an evaluation tool that can provide feedback to students (Koulaidis & Dimopoulos, 2003). With the development of questions in the learning assessment instrument, students will become accustomed to applying critical thinking skills and chemical literacy in solving the problems given (Davies, 2013).

Assessment instruments, which teachers commonly use to evaluate student learning, are typically limited to measuring a single ability. In fact, in learning, there may be several abilities that should be measured. To determine the success of learning in the classroom, teachers frequently assess the general thinking skills of students. However, the thinking ability assessment instrument used still only measures students' cognitive abilities in remembering (C1), understanding (C2), and applying (C3) (Yandriani et al., 2021). In addition, chemical literacy skills that are as important as thinking skills are never measured by teachers as an assessment of chemistry learning. Therefore, chemical literacy ability should be measured simultaneously (integrated) with other abilities. One of the abilities is to think critically. This integrated assessment of student learning outcomes can be used to measure several student abilities (Sumarni et al., 2016). However, in reality, there is little evidence about the existence of instruments used to measure students' critical thinking skills and chemical literacy in chemistry learning, especially in the lack of integrated instruments that can measure the critical thinking and chemical literacy skills simultaneously (Monk & Luxono, 2018).

Based on this, this study focuses on developing an integrated assessment instrument that can measure students' critical thinking skills and chemical literacy, especially in reaction rate materials.

Method

This research is a research and development (R&D) using a 4D development model by Thiagarajan et al. (1947). The stages of 4D development include Define, Design, Develop, and Disseminate. The stages of developing a fully integrated assessment instrument can be seen in Figure 1.

The subject of validation is 154 students in 12th grade from two high schools in Makassar City. The sample was determined using the purposive sampling technique, with the criteria of schools that have implemented learning that encourages students to think critically and increase their chemical literacy. The data collection techniques carried out are interview techniques and questionnaire techniques. The data collection instrument uses interview guidelines and questionnaires on the validity of question items. Interview guides are used to collect data related to the availability of assessment instruments that can measure students' critical thinking abilities and chemical literacy abilities to determine students' learning characteristics.

The question items in the integrated assessment instrument are developed based on the content in the 2013 curriculum related to reaction rate material. The stem of each question item contains each aspect of critical thinking and chemical literacy, which is developed in such a way that it can encourage students to think critically and increase their chemical literacy skills at the same time. The critical thinking framework is synthesized by several experts following Sarigoz (2012), Ennis (1987), Facione (1990), and Sadhu & Laksono (2018). Aspects of critical thinking based on the results of synthesis include (1) identifying problems, (2) building arguments, (3) analyzing problems, (4) evaluating problems and (5) drawing conclusions. The chemical literacy framework adopts the literacy framework by Shwartz et al. (2006) which is limited to the chemical aspect in the context which includes (1) explaining phenomena using chemical concepts, (2) using chemical understanding in solving problems, and (3) analyzing strategies and benefits of chemical application. The aspects of critical thinking are then integrated with aspects of chemical literacy to form an aspect called the integrated ability aspect. Based on this, there are five aspects of integrated abilities of critical thinking and chemical literacy skills that can be measured using the integrated assessment instruments developed, which are presented in Table 1.

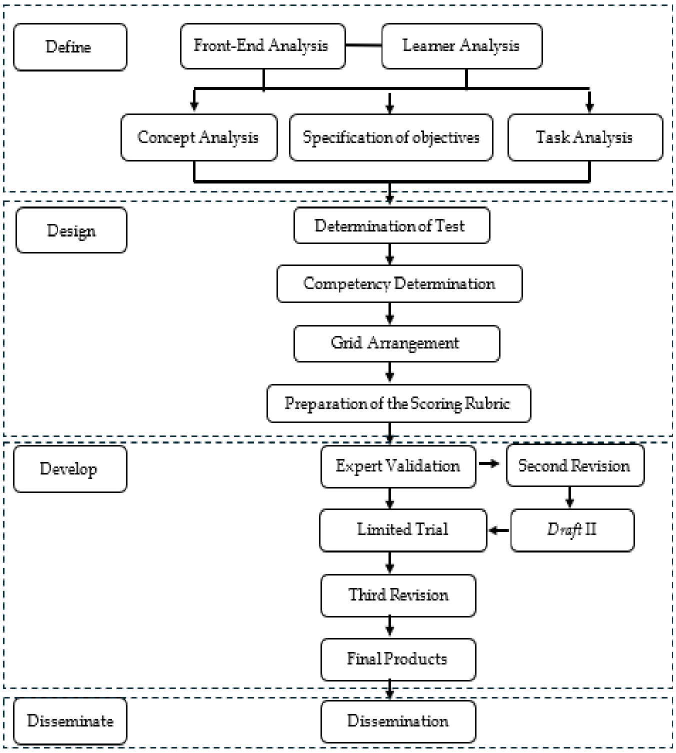

Figure 1. Development Procedure

The analysis carried out includes content validity analysis, construct validity, and item characteristics analysis. To determine the content validity of the integrated assessment instrument, an expert assessment (two chemistry lecturers and four chemistry teachers) was carried out based on the substance, construct, language, and product appearance aspects, which was then analyzed using Aiken's V formula. The analysis of the Rasch model was carried out with the help of the Winstep and SPSS programs. Before testing with the Rasch model, several assumption tests must be met, including unidimensional tests (construct validity), local independence tests, and parameter invariance tests (Hambleton & Swaminathan, 1985). Furthermore, for the analysis of construct validity, the exploratory factor analysis (EFA) approach is used to find out how many factors will appear. To determine whether the data is suitable for factor analysis, there is a test of the requirements for factor analysis that must be met. The test of the requirements consists of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO-MSA) test, the Barlett Sphericity test, and the anti-image correlation test.

Table 1. Integrated Aspects

Aspects of Critical Thinking	Aspects of Chemical Literacy	Integrated Aspects
Identifying problems	Analyze the strategies and benefits of chemical application	Identify problems by analyzing the benefits of chemical application
Building arguments	Explaining phenomena using chemical concepts	Building an argument by utilizing chemical concepts in explaining a phenomenon
Analyze the problem	Using chemical understanding	Analyze problems using chemical understanding
Evaluate the issue		Evaluate problems using chemical understanding
Drawing conclusions		Draw conclusions using chemical understanding

Result and Discussion

The integrated assessment instrument developed consists of ten items of description questions with the subject matter of reaction rate, each item contains aspects of critical thinking and aspects of chemical literacy. The scores for each item in the integrated assessment instrument have the same score weight, starting from a score range of 0 to 5, depending on the level of accuracy of the student's responses. A total of 6 experts assessed the validity of the content by providing assessments on several aspects including aspects of substance, construct, language, and product appearance of the integrated assessment instrument. The results of the assessment by the experts were then analyzed using Aiken's V formula. The content validity results were analyzed by comparing the Aiken's V index calculation result for each item with the value in the Aiken's V reference table (Kriswantoro et al., 2021). The validity criteria of the content were declared to be met, namely for the assessment of items with four scales of assessment categories, with the number of raters (assessors) as many as 6 people, and with the significance level used of 0.05, the value of the Aiken's V index obtained for each question item must reach the limit of the table reference value for Aiken's V of 0.78 (Aiken, 1985). The results of Aiken's V index value for each item are presented in Table 2.

Table 2. Aiken's V Index Value

No. Item	Aiken's V Value
1	1
2	0.889
3	0.778
4	0.778
5	0.944
6	0.833
7	0.833
8	1
9	0.778
10	1

The results presented in Table 2 show that all items meet the content validity testing criteria because Aiken's V index value obtained reaches the table reference value limit of 0.78. It can be concluded that all items in the integrated assessment instrument are declared valid in content. In addition, the assessments provided by experts also yield qualitative data in the form of improvement suggestions to refine the developed integrated assessment instrument. After obtaining the results that all items are valid in content, the next step is to analyze the construct validity and characteristics of the integrated assessment instrument items. The characteristics of items were analyzed using the Rasch model with the PCM-1PL approach. Several assumptions must be met before using the Rasch model with the PCM 1-PL approach.

The first is a unidimensional assumption test which aims to test whether each item only measures one variable or one ability. The unidimensional assumption test is also known as the instrument construct validity test. The construct validity aims to determine whether the items in an integrated assessment instrument are valid or not based on empirical data. Factor analysis is used to prove unidimensional assumptions by looking at the relationship between variables in the results of the calculation of eigen values in the variance-covariance matrix (Reckase, 1979). The principle used in factor analysis is to group data based on the intercorrelation among the items. The plot where the slope of the line begins to change (turning point) is used as the limit for the number of factors that can be taken (Daryono et al., 2020). Before conducting factor analysis, the data was first analyzed through the Kaiser-Mayer-Olkin Measure of Sampling Adequacy (KMO-MSA) and Barlett's sphericity test, to find out whether the data was worth analyzing with factor analysis. If both test criteria are met, then the data is declared feasible for further analysis with factor analysis. The results of the KMO-MSA and the Barlett sphericity test are presented in Table 3.

Table 3. KMO-MSA and Barlett Test Results

Kaiser-Mayer-Olkin Measure of Sampling Adequacy (KMO-MSA) Sample Size Suitability		0.776
Barlett sphericity Test	Chi-Square	1038.532
	Df	45
	Sig.	0.000

The KMO-MSA test aims to see the adequacy of sampling, while the Barlett sphericity test aims to determine whether there is a correlation between variables. The test criteria are met if the KMO-MSA test value obtained is greater than 0.5 (Leech et al., 2015) and the significance value of the Barlett sphericity test less than 0.05 (Beavers et al., 2013). The test results showed that the KMO-MSA value obtained was greater than 0.5 (0.776>0.5), which shows that the sample used in this study is sufficient. In addition, the significance value of the Barlett sphericity test is less than 0.05 (0.00<0.05), which showed that the variables in this study were correlated with each other. The results prove that the data obtained is worthy of further testing with factor analysis. Following Yilmaz's et al (2011) research state that the factor analysis test can be continued when the KMO-MSA test value is greater than 0.5 and the significance value of the Barlett sphericity test is less than 0.05.

Another condition that must be considered is to see a strong correlation between variables, which can be indicated by the anti-image correlation value. The anti-image correlation values can determine which variables are suitable for analysis in factor analysis. The criteria in this test is the anti-image correlation value must be greater than 0.5 (Pett et al., 2003). The value of the anti-image correlation is presented in Table 4. The results show that the anti-image correlation value for the entire item is greater than 0.5. Following the results obtained in the Sadhu & Laksono study (2018), which states that an item that has an anti-image correlation value greater than 0.5, then the item contributes highly to the factor structure formed so the item deserves further analysis with factor analysis. Therefore, the 10 items in the integrated assessment instrument are worthy of further use in factor analysis.

Table 4. Anti-Image Correlation Value

Anti-Image Correlation Value	Results
0.753	Saved/Used
0.834	Saved/Used
0.71	Saved/Used
0.67	Saved/Used
0.748	Saved/Used
0.778	Saved/Used
0.915	Saved/Used
0.858	Saved/Used
0.746	Saved/Used
0.775	Saved/Used

The next analysis is factor analysis. The principle in factor analysis is to group data based on intercorrelation between items. This test is carried out by paying attention to the eigen values obtained and the variance ratio that interprets the first factor (Al-Shirawia & Tashtoush, 2023). The condition for forming the factor of the variable being analyzed must have a eigen values greater than 1. The criteria for factor analysis are also indicated by the presence of the dominant factors produced (Costello & Osborne, 2005). The results of the factor analysis are presented in Figure 2.

Based on the scree plot results, it can be seen that 3 factors were formed, which have eigen values greater than 1. In addition, Component 1 shown in the scree plot has quite a big difference from Component 2, while Component 2 and Component 3 have a close difference. Following the research conducted by Pandra et al. (2021), which states that if there is one component that has a greater distance difference from the other components, then that component is referred to as the dominant factor and the other component is a good contributor to variance. With the formation of the dominant factors shown in the scree plot, then it can be concluded that the assumption test is said to be fulfilled.

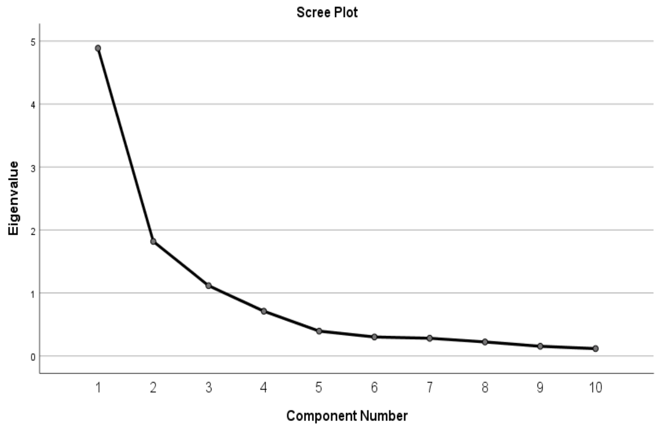

Figure 2. The Scree Plot of Factor Analysis

The second assumption test is the assumption of local independence. This test aims to prove whether students' answers to one item will not affect their answers to other items. The results of the assumption of local independence are presented in Table 5. This test criterion is said to be met if the covariance value between the ability intervals is small or close to zero (Pandra et al., 2021). Based on Table 5, all the values of the covariance variants between the groups of students' skill intervals that form a diagonal line on the matrix are all close to zero (0.00). This shows that the assumption of local independence of the integrated assessment instrument is fulfilled.

Table 5. Results of Local Independence Assumption

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
K1	0.02									
K2	0.01	0.01								
K3	0.00	0.00	0.00							
K4	0.00	0.00	0.00	0.00						
K5	0.00	0.00	0.00	0.00	0.00					
K6	0.00	0.00	0.00	0.00	0.00	0.00				
K7	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
K8	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.03		
K9	0.02	0.02	0.01	0.00	0.00	0.01	0.01	0.03	0.06	
K10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The last assumption test is the parameter invariance assumption test. This test aims to find out whether the characteristics of the item do not depend on the characteristics of the students answering the item. Testing the parameter invariance assumption includes the item invariance parameter and student ability invariance parameter. The assumption of the item invariance parameter is used to determine the consistency of item characteristics when answered by different groups of students. Meanwhile, the assumption of student ability invariance parameter is used to determine the consistency of students' abilities even though they answer items with different difficulty levels. The assumption of parameter invariance can be analyzed by making a scree plot of the item invariance parameter and a scree plot of the student ability invariance parameter. The test criteria were said to be met if each dot was relatively close to the slope line, which showed no variation in the estimated parameters in both groups. The item parameter invariance assumption test results can be seen in the scree plot presented in Figure 3.

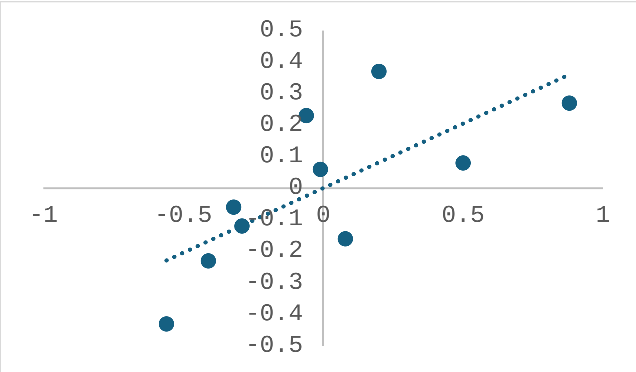

Figure 3. Scree Plot of Item Invariance Parameters

It can be seen in Figure 3, that almost all dots on the scree plot are relatively close to the slope line. The scree plot shows that the characteristics of the items in the integrated assessment instrument have met the criteria for the item parameter invariance assumption. This illustrates that despite being responded to by several groups of students, the items' characteristics remained the same. In addition, testing the assumption of the student ability invariance parameter was also carried out by looking at the results of the scree plot presented in Figure 4. Figure 4 shows that almost all dots on the scree plot are relatively close to the slope line. The scree plot showed that the characteristics of the students who responded meet the criteria for the student ability invariance parameter assumption. The results obtained are in accordance with Retnawati (2014) research, which states that the assumption of item parameter invariance is met when the plots on the scree plot approach the slope line. The results showed that even though students answer items with different difficulty levels, the characteristics of students' abilities do not change at all.

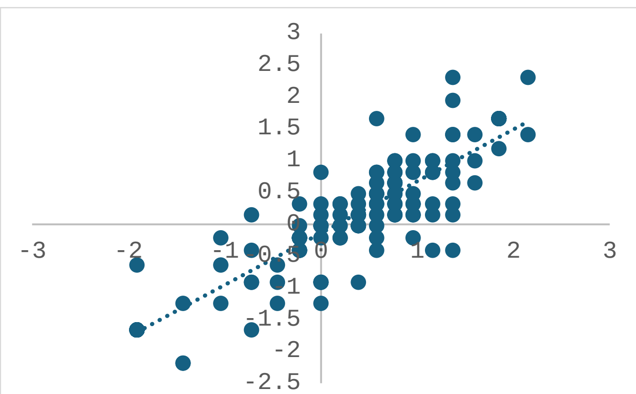

Figure 4. Scree Plot of Students' Ability Invariance Parameters

After the three Rasch modeling assumption tests are met, the next step is to analyze the reliability and characteristics of the integrated assessment instrument that has been developed. The characteristics of the items include the item fit analysis, the item difficulty, and the information function. The reliability test of the instrument is carried out to measure how reliable or consistent an instrument is in measuring a characteristic. Reliability analysis consists of reliability to the subject and reliability to the item. The results of the reliability of the integrated assessment instrument are presented in Table 6.

Table 6. Integrated Assessment Instrument Reliability

Reliability Parameter (N)	Reliability Coefficient	Categories Liability
Subject Reliability (154)	0.85	High
Item Reliability (10)	0.91	High

According to Table 6, the subjects' reliability coefficient was 0.85, whereas the items' reliability coefficient was 0.91. The results reveal that the reliability of the subject and the reliability of the integrated assessment instrument items generated are both in the high category. The reliability coefficient of the items obtained at 0.91 shows that the entire item in the integrated assessment instrument was stated to be 91% consistent in measuring the same thing repeatedly. Meanwhile, the reliability coefficient of the subject obtained of 0.85 shows that all students in this study are stated to be 85% consistent in responding to the items in the integrated assessment instrument. Following the research of Pandra et al. (2021), the reliability score with a high category indicates that the integrated assessment instrument is suitable for measurement. Thus, the integrated research instruments developed in this study can be further used to measure students' integrated critical thinking and chemical literacy skills.

The first characteristic analysis is the item fit analysis. The item fit analysis aims to determine whether the items can function well in measuring. Three criteria can be used to analyze the items fit with the model, consisting of the value of the outfit mean square (MNSQ) is accepted if the value is in the range of 0.5 < MNSQ < 1.5; the value of outfit Z-standard (ZTSD) is accepted if the value ranges from -2.0 < ZTSD < +2.0; and the value of the correlation point (PT Mean Corr) is accepted if the value is in the range of 0.4 < PT Mean Corr < 0.85 (Boone et al., 2014). The results of the item fit analysis are presented in Table 7.

Table 7. Item Fit Result

No. Item	Outfit MNSQ Value	Outfit ZSTD Value	PT Measure Corr. Value	Conclusion
1	1.02	0.3	0.68	Item Fit
2	0.83	-1.6	0.7	Item Fit
3	1.03	0.3	0.68	Item Fit
4	1.12	1.1	0.63	Item Fit
5	0.54	-4.8	0.79	Item Fit
6	0.71	-2.8	0.74	Item Fit
7	0.74	-2.5	0.68	Item Fit
8	0.71	-2.8	0.72	Item Fit
9	1.07	0.6	0.68	Item Fit
10	1.26	2.1	0.62	Item Fit

Based on Table 7, all items meet at least two criteria for item fit analysis, so that all items are declared fit (suitable) to the PCM model used. Thus, no items are discarded in the integrated assessment instrument, and all the items in the integrated assessment instrument can be said to be suitable for use as final products, which can be used to measure the integrated ability of critical thinking and chemical literacy skills of students.

The next characteristic analysis is the analysis of the difficulty level of the items. This test aims to determine the index difficulty of the items developed, whether they belong to moderate, easy, or difficult categories. The results of the index difficulty of the items are presented in Table 8. According to Hambleton et al. in Yanto's research et al. (2019), the criteria for a good item include an index difficulty range of no less than -2.0 or more than +2.0 logit. Based on this, Table 8 shows that all items in the integrated assessment instrument are good questions, as their difficulty indexes range from -0.48 to +0.52 logit. In addition, Table 8 shows that the items in the integrated assessment instrument have difficulty levels ranging from the item with easy, medium, to difficult levels. Because all items are in a good category, the integrated assessment instrument can be used further to measure students' critical thinking and chemical literacy skills.

Table 8. Results of Item Difficulty Level

No. Item	Index Difficulty (logit)	Categories of Difficulty
1	-0.48	Easy
2	-0.17	Medium
3	0.1	Medium
4	0.29	Medium
5	-0.31	Medium
6	-0.19	Medium
7	0.03	Medium
8	-0.05	Medium
9	0.26	Medium
10	0.52	Difficult

The last analysis is related to the characteristics of the item, namely the analysis of the information function test. The analysis of the information function test aims to provide an overview of the contribution of an item to the estimation of students' latent abilities measured using an integrated assessment instrument. The results of the information function test can be seen in the graph presented in Figure 5.

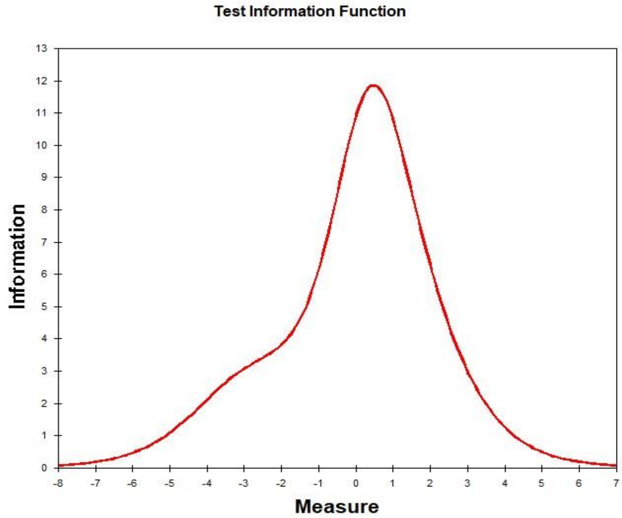

Figure 5. The Result of the Information Function Test

Figure 5 shows that the measurement information obtained is quite low for a low level of ability. Similarly, for a high level of ability, the measurement information obtained is also quite low. However, unlike for the medium level of ability, the measurement information obtained is very high, so based on this information function, it can provide an overview that the integrated assessment instrument developed can be used to measure the integrated ability of students at low, medium to high levels.

Conclusion

The integrated assessment instrument, which consists of ten items on the topic of reaction rates, is relatively valid and reliable. The reliability of the item and subject was both found in high category. An integrated assessment instrument can be used to measure five integrated aspects of critical thinking and chemical literacy skills.

Acknowledgments
This research could be completed thanks to the contributions of various parties. On this occasion, the author would like to express gratitude to the principals of several high schools in Makassar City who allowed the author to conduct research at their schools. Special thanks are also extended to the chemistry teachers who were willing to provide the opportunity for this research.

Author Contributions
Prof. Dr. Endang Widjajanti L.F.X., M.S. as the supervisor who provided guidance, input, and valuable knowledge to the author.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

References

Aiken, L.R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*(1), 131-142. http://doi.org/10.1177/0013164485451012.

Al-Shirawia, N., & Tashtoush, M. (2023). Differential item functioning analysis of an emotional intelligence scale for human resources management at Sohar University. *Information Sciences Letters*, *12*(11), 2937–2952. https://www.naturalspublishing.com/Article.asp?ArtcID=27843.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research and Evaluation*, *18*(6), 1-13. https://doi.org/10.7275/qv2q-rk76.

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

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, *10*(7), 1-9. https://doi.org/10.7275/jyj1-4868.

Daryono, R. W., Hariyanto, V. L., Usman, H., & Sutarto, S. (2020). Factor analysis: Competency framework for measuring student achievements of architectural engineering education in Indonesia. *Research and Evaluation in Education*, *6*(2), 98-108. https://doi.org/10.21831/reid.v6i2.32743.

Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research and Development*, *32*(4), 529-544. https://doi.org/10.1080/07294360.2012.697878.

Dori, Y. J., Rodrigues, S., & Schanze, S. (2013). How to promote chemistry learning through the use of ICT. *Teaching Chemistry – A Studybook,* 213-240. https://doi.org/10.1007/978-94-6209-140-5_8.

Ennis, R. H. (1987). *A taxonomy of critical thinking disposition and abilities. In J. B. Baron & R. J. Sternberg (Eds.), Teaching thinking skills: Theory and practice*. New York: W. H. Freeman.

Facione, P. A. (1990). Critical Thinking : A Statement of expert consensus for purposes of educational assessment and instruction executive summary the Delphi report. *The California Academic Press*, *423*(c), 1-21.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory principles and applications*. MA: Kluwer Nijhoff Publishing.

Koballa, T., Gräber, W., Coleman, D. C., & Kemp, A. C. (2000). Prospective gymnasium teachers' conceptions of chemistry learning and teaching. *International Journal of Science Education*, *22*(2), 209-224. https://doi.org/10.1080/095006900289967

Koulaidis, V., & Dimopoulos, C. (2003). An analysis of the discursive transitions across different modalities of the pedagogic discourse. *International Journal of Learning*, *10,* 3263-3274.

Kriswantoro, Kartowagiran, B., & Rohaeti, E. (2021). A critical thinking assessment model integrated with science process skills on chemistry for senior high school. *European Journal of Educational Research*, *10*(1), 285-298. https://doi.org/10.12973/EU-JER.10.1.285.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2015). *IBM SPSS for intermediate statistics: Use and interpretation (5th ed.)*. New York and London: Routledge.

OECD. (2023). *PISA 2022 results (volume II); Learning during and from disruption*. OECD. https://doi.org/10.1787/a97db61c-en.

Pandra, V., Kartowagiran, B., & Sugiman. (2021). Mathematics test development by item response theory approach and its measrument on elementary school students. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(5), 464-483. https://doi.org/10.17762/turcomat.v12i5.994.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oak, California: Sage Publication, Inc.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, *4*(3), 207-230. https://doi.org/10.3102/10769986004003207.

Retnawati, H. (2014). *Teori respon butir dan penerapannya; Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.

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.

Sari, S., Safitri, I., & Farida, I. (2019). Design of educational games oriented to chemical literacy on petroleum material. *Journal of Physics: Conference Series*. *1402*(5), 1-6. https://iopscience.iop.org/article/10.1088/1742-6596/1402/5/055032/meta.

Sarigoz, O. (2012). Assessment of the high school students' critical thinking skills. *Procedia - Social and Behavioral Sciences*, *46,* 5315-5319. https://doi.org/10.1016/j.sbspro.2012.06.430.

Shwartz, Y., Ben-Zvi, R., & Hofstein, A. (2006). The use of scientific literacy taxonomy for assessing the development of chemical literacy among high-school students. *Chemistry Education Research and Practice*, *7*(4), 203-225. https://doi.org/10.1039/B6RP90011A.

Suardana, I. N., Redhana, I. W., Sudiatmika, A. A. I. A. R., & Selamat, I. N. (2018). Students' critical thinking skills in chemistry learning using local culture-based 7E learning cycle model. *International Journal of Instruction*, *11*(2), 399-412. https://doi.org/10.12973/iji.2018.11227a.

Sumarni, W., Sudarmin, S., Wiyanto, W., & Supartono, S. (2016). Preliminary analysis of assessment instrument design to reveal science generic skill and chemistry literacy. *International Journal of Evaluation and Research in Education (IJERE)*, *5*(4), 331-340. https://doi.org/10.11591/ijere.v5i4.5961.

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1947). *Instructional Development for Training Teachers of Exceptional Children A Sourcebook*. Bloomington, Indiana: Indiana University.

Yandriani, Rery, R. U., & Erna, M. (2021). Developing and validating the assessment instruments to measure students' analytical thinking ability and chemical literacy on colligative properties. *Journal of Physics: Conference Series*, *1788*(1), 1-11. https://doi.org/10.1088/1742-6596/1788/1/012027.

Yanto, B. E., Subali, B., & Suyanto, S. (2019). Measurement instrument of scientific reasoning test for biology education students. *International Journal of Instruction*, *12*(1), 1383-1398. https://doi.org/10.29333/iji.2019.12188a.

Yilmaz, K., Altinkurt, Y., & Cokluk, O. (2011). Developing the educational belief scale: The validity and reliability study. *Educational Science: Theory and Practice*, *11*(1), 343-350. Retrieved from https://files.eric.ed.gov/fulltext/EJ919905.pdf.