Jurnal Pendidikan Indonesia
Volume 11, Number 3, 2022 pp. 480-491
P-ISSN: 2303-288X E-ISSN : 2541-7207
Open Access: https://doi.org/10.23887/jpiundiksha.v11i3.39560

Test Instrument to Measure Student's Level of Chemical Literacy in Macromolecules Topic

Oktrilina Amelia[1*], Risnita Vicky Listyarini [2]

[1,2] Chemistry Education Department, Universitas Sanata Dharma, Yogyakarta, Indonesia

ARTICLE INFO

Article history:
Received September 21, 2021
Revised September 22, 2021
Accepted July 30, 2022
Available online September 25, 2022

Kata Kunci:
Test Instruments, Chemical Literacy, Macromolecules.

Keywords:
Test Instruments, Chemical Literacy, Macromolecules.

ABSTRAK

Tingkat literasi sains siswa dapat ditentukan dengan menggunakan alat ukur yaitu instrumen tes yang mencakup empat aspek literasi kimia pada setiap itemnya. Penelitian ini bertujuan untuk mengembangkan instrumen tes yang digunakan untuk mengukur ketercapaian tingkat literasi kimia siswa. Metode penelitian yang digunakan adalah Model Educational Reconstruction (MER). Uji coba terbatas instrumen tes dilaksanakan pada dua belas siswa di SMA. Pengumpulan data dilakukan berdasarkan lembar validasi instrumen tes dari validasi ahli, pretest dan posttest. Analisis data instrumen tes didasarkan pada validasi instrumen tes menggunakan rumus Aiken's V. Berdasarkan analisis koefisien korelasi instrumen tes menghasilkan 20 butir soal yang valid dengan nilai koefisien korelasi 30,8% dengan kriteria validitas sedang, 7,7% dengan kriteria tinggi dan 12,8% dengan kriteria sangat tinggi. Uji reliabilitas menunjukkan bahwa instrumen tes reliabel dan memiliki korelasi yang tinggi dengan nilai Cronbach Alpha sebesar 0,920. Tingkat kesulitan soal dan diskriminatif soal ditentukan. Hasil angket respon siswa terhadap instrumen tes memiliki persentase rata-rata 79% dengan kriteria praktis. Penelitian ini menyimpulkan bahwa instrumen tes dapat digunakan untuk mengukur tingkat literasi kimia siswa. Tingkat literasi kimia siswa pada tingkat rendah dan tingkat literasi sains nominal merupakan tingkat tertinggi yang dapat dicapai siswa.

ABSTRACT

The science literacy level of students can be determined using a measuring instrument, namely a test instrument that includes four aspects of chemical literacy in each item. This study aims to develop a test instrument that is used to measure the achievement of a student's chemical literacy level. The research method used is the Model of Educational Reconstruction (MER). The limited trial of the test instrument was implemented for twelve students in the senior high schools. Data collection was carried out based on validation sheets of test instruments from expert validation, pretest and posttest. Data analysis for the test instrument is based on validation of the test instrument using Aiken's V formula. Based on the correlation coefficient analysis of the test instrument resulted in 20 valid question items with a correlation coefficient value of 30.8% having moderate validity criteria, 7.7% having high criteria and 12.8% having very high criteria. The reliability test shows that the test instrument is reliable and has a high correlation with the Cronbach Alpha value of of 0.920. The difficulty level of the question and the discriminatory of the question were determined. The results of the student response questionnaire to the test instrument have an average percentage of 79% with practical criteria. This research concludes that the test instruments can be used to measure the level of chemical literacy of students. The chemical literacy level of students at the low level and the nominal level of scientific literacy is the highest level the students can achieve.

1. INTRODUCTION

Chemistry is one part of science that studies the structure, properties, and changes of a substance that tends to be abstract so students need to have a high level of thinking ability (Mutakinati et al., 2018; Rasmawan, 2018; Rosmani & Halim, 2017). The ability of students in understanding the concept of chemical learning and to be able to associate it with various phenomena in daily life is related to the level of scientific literacy of students. This level involves not only the understanding of science but also the level

*Corresponding author
E-mail addresses: oktrilinamelia@gmail.com (Oktrilina Amelia)

of applying it in real conditions of problems faced in daily life (Camel et al., 2020; Giri & Ardiawan, 2018; Rostikawato, 2016). The achievement of students' scientific literacy has a close relationship with the world of science so it becomes the main goal in the success of education in schools (Hottecke & Allchin, 2020; Tesi Muskania & Wilujeng, 2017). However, the achievement of scientific literacy level in Indonesia is still relatively low. The low level of scientific literacy in Indonesia is recorded in the results of the Program for International Student Assessment (PISA). The PISA result data showed that Indonesia is still in the bottom group of countries that are participating in the assessment (El Islami & Nuangchalerm, 2020; Nugrahanto & Zuchdi, 2019). PISA is a program incorporated into the Organization for Economic Cooperation and Development to assist countries in preparing human resources to achieve the competencies expected by international markets (Anandari & Iswandari, 2019; Hidayah et al., 2021; Hong & Talib, 2018). In line with this fact, other data from showed that Indonesia is in the lowest rank in terms of ability literacy. Therefore, there is an urgency to solve this problem. Several factors are causing low literacy in Indonesia one of them is the school environment. As previously mentioned, low literacy is related to the ability of students in understanding a concept so that it affects the way students learn (Oktariani & Febliza, 2019; Pratiwi et al., 2019). Previous study explained that learning methods and teaching materials used in schools do not facilitate the teaching of reading comprehension (Hottecke & Allchin, 2020). Reading activities become uninteresting because of the monotonous learning model thus impacting the evaluation of the level of learning understanding results. It is found that the low level of students' scientific literacy might be caused by the fact that the students have never worked or are not familiar with the scientific literacy evaluation question. Evaluation instruments to measure the level of scientific literacy of students adopted from PISA already exist but the evaluation instruments are made on a large/international scale (Kholis et al., 2020; Pratiwi et al., 2019). There needs to be a test instrument to measure the level of scientific literacy of students in scope or on a small scale. The results of students' scientific literacy can be used as evaluation material for the learning process. There is a need for the development of a scientific literacy evaluation instrument on a smaller scale integrated with learning activity evaluation at school. In line with that, chemical literacy can be defined as a part of scientific literacy (Mozeika & Bilbokaite, 2010; Thummathong & Thathong, 2018). The level of chemical literacy of students can be measured using instruments containing learning evaluations in the form of test instruments. This study will focus on developing a test instrument to measure the level of students' chemical literacy which is made regarding the four domains. Test instrument to measure student's level of chemical literacy in macromolecules topic contains four domains that cover chemical literacy are knowledge of chemical materials and scientific idea, chemistry in context, high level of learning ability, and affective aspects. Previous research state that there are five levels of scientific literacy that can be achieved by students, namely: 1) Scientific illiteracy; 2) Nominal scientific literacy; (3) Functional scientific literacy; Conceptual scientific literacy; and (5) Multidimensional scientific literacy (Shwartz et al., 2006). Based on the results of an interview with a chemistry teacher in one of the senior high school SMA in Bantul, Yogyakarta, the teacher has never developed or used test instruments to measure students' levels of chemical literacy in particular on macromolecules topic. Macromolecules topic is one of the chemical concepts that are closely related in daily life. Therefore, it is necessary to develop test instruments to determine the level of chemical literacy of students in macromolecules topic. Previous research to measure the level of chemical literacy has been carried out (Fahmina et al., 2019; Muntholib et al., 2020; Yusmaita & Nasra, 2018). Model of Educational Reconstruction (MER) is used for the development of the test instrument because the results were good. This research aims to develop a test instrument to measure students' level of chemical literacy on macromolecules topic. The measurement of students' chemical literacy for instance in macromolecules topics can bring advantages for teachers to be used as an evaluation to design a better learning process and able to improve the understanding of the chemical concepts of students.

2. METHOD

The approach used in this research is a mixed-method which combines quantitative and qualitative methods. Qualitative studies will explain and describe data based on the results of interviews, and analysis results from quantitative studies. Qualitative studies are based on results of validation sheets, student response questionnaires, pretest and posttest, and test instruments. The study was conducted in April 2021 at one of the senior high school SMA in Bantul, Yogyakarta. The research sample is a class XII IPA consisting of twelve students. The method used in this research was the Model of Educational Reconstruction (MER) which consisted of three components that affect each other and are recursive (Duit et al., 2012; Nursa'adah et al., 2018) i.e. research on teaching & learning; clarification and analysis of science content; design and evaluation of teaching and learning. Data collection was carried out based on validation sheets of test instruments from expert validation, pretest and posttest, student response questionnaires, and the results of testing the test instrument on students. Before the test instrument is given to students, students are given a pretest to determine their initial chemical literacy ability. The pretest given to students is 20 statements (true/false) related to macromolecular topics. Furthermore, a test instrument is given which contains 20 essay questions containing polymer, carbohydrates, protein and fat materials (5 questions for each material). Posttest is given after the test instrument testing to determine the effectiveness of the test instrument. The number and types of questions on the posttest are the same as those of the pretest. Students' response questionnaires were given to determine the practicality of the developed instrument. It shows the attractiveness of the test instrument, the ease of use of the test instrument and the clarity of the instructions for working on the test instrument. Data analysis for the test instrument is based on validation of the test instrument using Aiken's V formula (Akbar, 2013; Azwar, 2013). The validity of the criteria is defined as from 0 to 1 (very valid). The item validity in the test instrument is analyzed using the correlation coefficient according to Guilford. The reliability using Cronbach Alpha. The reliability value obtained can be interpreted into several criteria according to Guilford. The difficulty level was calculated based on with criteria 0 - 1.0 (very difficult to very easy). Discrimination power was calculated based on and with criteria 0 - 1.0 (bad to very good). Analysis of the science literacy level of students is done by calculating the total score obtained by students. Interpretations of students' chemical literacy levels were determined based on with criteria very good to very poor (scale 0 – 100%). Practicality criteria of the test instrument based on the student response questionnaire. The criteria are based on with criteria not practical to very practical (0 – 100%).

3. RESULT AND DISCUSSION

Result

In this study, the test instrument was used as a tool to determine and measure student learning outcomes in cognitive terms, such as knowledge, understanding, analysis and evaluation. The results of the validation of the test instrument by four validators are presented in Table 1.

Table 1. Test Instrument Validation Results

Aspects	Validator			
	V1	V2	V3	V4
Graphic component	100%	93%	93%	87%
Graphic	95%	95%	90%	80%
Effectiveness	87%	87%	100%	80%
Average Percentage	94%	92%	94%	82%
Criteria	Very feasible	Very feasible	Very feasible	Feasible
Average Percentage	91%			
Validator recommendation	Applicable with little revision			

Base on Table 1, the average percentage value is 91% indicates that the results of the validation of the test instrument are within the very feasible criteria. The average percentage value that is above 85% is included in the very feasible criteria. It can be concluded that based on the average percentage results obtained, the design of the test instrument is considered very feasible to be used in limited trials with revision based on validators' suggestions. *Question Item Validation* The results of the validation of each item along with the criteria are presented in Table 2.

Table 2. Results of Item Validation in Test Instruments

Question number	Average V	Criteria
1	0.896	Very high validity
2	0.934	Very high validity
3	0.906	Very high validity
4	0.885	Very high validity
5	0.899	Very high validity
6	0.885	Very high validity
7	0.892	Very high validity
8	0.882	Very high validity
9	0.868	Very high validity

Question number	Average V	Criteria
10	0.868	Very high validity
11	0.899	Very high validity
12	0.878	Very high validity
13	0.854	Very high validity
14	0.878	Very high validity
15	0.875	Very high validity
16	0.875	Very high validity
17	0.892	Very high validity
18	0.851	Very high validity
19	0.872	Very high validity
20	0.889	Very high validity
Average V	**0.884**	**Very high validity**

Based on Table 2, it can be seen that the validity of each question item based on the assessment of each aspect of the item shows very high validity criteria because the value obtained exceeds 0.800 with the average result of the Aiken V scale validity test being 0.884. The validity value of the Aiken V scale which is above 0.800 is included in the very high validity criteria and can be used for limited trials. Based on this statement and the results of the validity of the Aiken V scale which shows results above 0.800, then 20 question items are considered and can be used for a limited trial with repair based on the validator's suggestions.

Limited Trial Results of Test Instrument

The limited trial was used to determine the quality of the test instrument and to measure the science literacy level of students. The test instrument can be categorized as good if it is valid, reliable, has distinguishing power and has a good level of difficulty. The limited trial will produce data to determine the quality of the test instrument design to measure the chemical literacy level of students. Test instrument can be said well if it is valid, reliable, has distinguishing power and has a good level of difficulty. The results of the limited test analysis of the test instrument to determine the quality of the test instrument are described in several aspects below.

Correlation coefficient analysis

The question items contained in the test instrument can be said to be valid and capable of measuring what should be measured if they have a validity value that is following the provisions. The validity of each item is determined using item analysis and is done by correlating the score of each item with the total score which is the number of each item. The Pearson Product Moment correlation coefficient is the formula used to test the correlation. The test of the validity of the correlation coefficient (r_{xy}) aims to determine the validity of each question item contained in the test instrument to measure the level of science literacy of students. This validity test will analyze the validity of each item by correlating the score of each item with the total score of all items. There are 20 questions contained in the test instrument, there are 19 numbered questions which contain 2 questions and 1 question number which contains only 1 question so the total number of questions contained in 20 numbered questions is 39 questions. According to Guilford in very high validity has a range of r_{xy} values above 0.90; high validity between 0.70 to 0.90; moderate validity between 0.40 to 0.70; low validity between 0.20 to 0.40; and very low validity under 0.20. The results of the validity test using the correlation coefficient (r_{xy}) along with the criteria are presented in Table 3.

Table 3. Validity Test Results using Correlation Coefficient (r_{xy})

Question Item	Topic	R_{xy}	Validity Criteria
1a		0.634	High
1b		0.841	Very high
2a		0.502	Moderate
2b		0.797	High
3a	Polymer	0.224	Low
3b		0.526	Moderate
4a		0.857	Very high
4b		0.498	Moderate
5a		0.481	Moderate

Question Item	Topic	R_{xy}	Validity Criteria
5b		0.505	Moderate
6a		0.890	Very high
6b		0.433	Moderate
7a		0.187	Very low
7b		0.314	Low
8a	Carbohydrate	0.433	Moderate
8b		0.371	Low
9a		0.259	Low
9b		0.204	Low
10a		0.303	Low
10b		0.268	Low
11a		0.650	High
11b		0.364	Low
12a		0.433	Moderate
12b		0.433	Moderate
13a		0.433	Moderate
13b		0.433	Moderate
14a		0.189	Very low
14b	Protein	0.394	Low
15a		0.212	Low
15b		0.211	Low
16a		0.303	Low
16b		0.298	Low
17a		0.184	Very low
17b		0.252	Low
18	Fat	0.204	Low
19A		0.890	Very high
19B		0.927	Very high
20A		0.432	Medium
20B		0.308	Low

Base on Table 3, the validity of the items is part of the content validity because the items are used as a measuring instrument to determine learning outcomes. A measuring instrument used to determine learning outcomes can be said to be valid if it is following the curriculum content to be measured. So it can be concluded that based on the results of the analysis of the validity test using the correlation coefficient (r_{xy}), there are a total of 20 question items which are included in the criteria of moderate validity, high validity and very high validity which are suitable for use in limited trials as a measuring tool to determine the level of students' chemical literacy.

Reliability analysis

Reliability analysis was carried out to show the consistency of the measurement results even though it was carried out several times on the same object. The reliability value will indicate that the instrument can be used as a tool to generate data because the instrument developed is good. Reliability is used to indicate the level of consistency, accuracy and precision of the test instrument used. This consistency means that although the items given are different and the correctors are different, the characteristics of the measurement results will remain the same. Based on the results of the analysis of the validity test using the correlation coefficient (r_{xy}), there are 20 valid question items so the reliability value is calculated only on the valid questions. Table 4 shows the reliability value of the test instrument for the 20 question items.

Table 4. Test Instrument Reliability

Cronbach Alpha	N Question Item
0.920	20

If the Cronbach Alpha value is between 0.60 and 1, then the instrument is reliable and has a high correlation. Table 4 shows that the Cronbach Alpha value on the test instrument is 0.920. The reliability value indicates that the instrument is reliable and has a high correlation. In the interpretation scale by

Guilford, the reliability value which is on a scale of 0.80 to 1 is included in the high criteria. Based on this statement, it can be concluded that the 20 items contained in the test instrument are reliable and have high-reliability criteria so that they can show the consistency of measurement results even though they are carried out several times.

Difficulty Level Analysis

Analysis of the difficulty level of the question was performed on 20 question items with the valid criteria because it is considered able to be used as a measuring tool to measure the level of students' chemical literacy on macromolecules topic. Difficulty level analysis aims to find out the difficulty level of the question. The results of the difficulty level and the criteria are presented in Table 5.

Table 5. Test Results of Item Difficulty Level

Question Item	Topic	Difficulty Level	Criteria
1a		0.92	Easy
1b		0.50	Moderate
2a		0.33	Moderate
2b		0.83	Easy
3b	Polymer	0.50	Moderate
4a		0.42	Moderate
4b		0.28	Difficult
5a		0.67	Moderate
5b		0.39	Moderate
6a		0.50	Moderate
6b	Carbohydrate	0.17	Difficult
8a		0.17	Difficult
11a		0.59	Moderate
12a	Protein	0.17	Difficult
12b		0.17	Difficult
13a		0.17	Difficult
13b		0.17	Difficult
19A		0.33	Moderate
19B	Fat	0.47	Moderate
20A		0.50	Moderate

The proportion of the level of difficulty of the questions should be spread out to obtain good learning achievement. These proportions can be arranged as follows: 1) difficult questions 25%, moderate questions 50%, easy questions 25%; 2) 20% difficult questions, 60% moderate questions, 20% easy questions; or 3) 15% difficult questions, 70% moderate questions, 15% easy questions. The proportions shown in Table 5 compared with the proportions according to have in common, that is the proportion of questions with moderate criteria is higher than the difficult and easy criteria. This shows that the proportion of questions with moderate criteria has a good proportion. For questions with difficult and easy criteria, because the proportions between difficult and easy criteria are not balanced, they do not comply with good proportions according to the proportion. Based on these results, it can be concluded that the proportion of questions with difficult and easy criteria does not meet the proportion of the level of difficulty of the questions that should be spread out to obtain good learning achievement.

Discriminatory Power Analysis

The discriminatory power analysis aims to show the ability of the items in distinguishing students belonging to high and low ability groups. The discriminatory power of questions will compare students who have mastered or have not mastered the material. Distinguishing power can be determined by dividing the sample into two equally large classes, namely the upper class and the lower class. The results obtained in the limited trial are then sorted from the students who get the highest results to the lowest results. As many as 50% of students with the highest results will enter the upper class while as many as 50% of students with low results will enter the lower class. Each item has different criteria for differentiating questions; poor, satisfactory, good and excellent. The results of the discriminatory power and the criteria are presented in Table 6.

Table 6. Distinguishing Power Results

Question item	Topic	DP	Criteria
1a	Polymer	0.06	Poor
1b		0.33	Satisfactory
2a		0.00	Poor
2b		0.33	Satisfactory
3b		0.20	Satisfactory
4a		0.83	Excellent
4b		0.56	Good
5a	Carbohydrate	0.67	Good
5b		0.78	Excellent
6a		1.00	Excellent
6b		0.33	Satisfactory
8a		0.33	Satisfactory
11a	Protein	0.26	Satisfactory
12a		0.33	Satisfactory
12b		0.33	Satisfactory
13a		0.33	Satisfactory
13b		0.33	Satisfactory
19A	Fat	0.67	Good
19B		0.94	Excellent
20A		0.33	Satisfactory

The questions that may be used in the trial do not have to be in the category of good discriminatory criteria. Questions that are in the sufficient criteria with a large coefficient of 0.2 to 0.4 are still feasible to use. Two question items are in the criteria of poor discriminating power. Decision-making about items that must be rejected, revised or accepted can be known based on the distinguishing power coefficient obtained. Table 6 shows that the 20 items have criteria from poor, satisfactory, good and excellent. Decisions in the selection of questions to be rejected, revised or accepted are based on the discriminating power coefficient. Based on the decisions in **Table 6,** it can be seen that two question items are in the decision to be revised and two question items are rejected. If the data collection is carried out again, the two questions must be revised because the coefficient is below 0.29, namely question item 3b with a discriminating coefficient of 0.20 and question item 11a with a discriminating coefficient of 0.26. Two items that were rejected because the discriminatory coefficient was below 0.10 were question item 1a with a discriminatory coefficient of 0.06 and question item 2a with a discriminatory coefficient of 0.00. If there are items that are rejected, the items can be discarded or replaced with new items. The distinguishing power of a good question can be seen from the proportion of the question criteria. From the results of the discriminatory test of questions, the questions that are in the excellent criteria are 20% and the good criteria are 15% and the satisfactory criteria are 55% which in total 90% and can be considered for instrument test. It can be concluded that the test instrument can be used to determine students' chemical literacy.

Students' Chemical Literacy Level

The chemical literacy level of students can be known based on the final score of the test instrument work. The total score obtained is then calculated and converted into a percent value. Interpretation of students' chemical literacy level criteria based on the percent value obtained is presented in Table 7.

Table 7. Results of Analysis and Achievement of Students' Chemical Literacy Level

Student Code	Final Score	Percentage	Level
1A	35	58%	Low
2A	22	38%	Low
3A	29	48%	Low
4A	29	48%	Low
5A	20	33%	Low
6A	28	47%	Low
7A	6	10%	Low

Student Code	Final Score	Percentage	Level
8A	7	12%	Low
9A	10	17%	Low
10A	3	5%	Low
11A	2	3%	Low
12A	3	5%	Low
Average percentage		27%	Low

Students' Pretest and Posttest Results

The developed test instrument is equipped with pretest and posttest sheets. Pretest and posttest sheets can be used to determine the effectiveness of the test instrument. In line with this statement, the pretest and posttest sheets in this study will show the effectiveness of the test instrument being tested. The effectiveness of the test instrument used is not only to determine learning achievement but also to provide new information as a process of developing student knowledge, known as assessment as learning. Assessment as learning is the use of assessment to support and develop students' metacognition. Based on this statement, the effectiveness of the test instrument as an assessment of learning can be seen from the normalized N-gain. The normalized N-gain analysis is an analysis of the results of the pretest and posttest to determine the increase in student learning outcomes. The pretest and posttest given to students are 20 statements related to macromolecular material. The results of the pretest and posttest analysis using normalized N-gain along with the criteria are presented in Table 8.

Table 8. Results of Pretest and Posttest Analysis Using Normalized N-gain

Students Code	Score		Normalized N-gain	Criteria
	Pretest	Posttest		
1A	15	19	0,8	High
2A	16	19	0,8	High
3A	16	19	0,8	High
4A	15	19	0,8	High
5A	14	19	0,8	High
6A	15	16	0,2	Low
7A	16	17	0,3	Low
8A	14	19	0,8	High
9A	12	16	0,5	Intermediate
10A	14	15	0,2	Low
11A	18	18	0,0	Low
12A	15	17	0,4	Intermediate

Based on Table 8, the increase in student learning outcomes is on various criteria; high, medium and low. The pretest and posttest scores showed an increase except for students with code 11A. The pretest score of students with code 11A is the highest score of 12 students, which is 18 out of a maximum score of 20. The pretest score shows the extent to which the material is mastered by students so that when viewed based on the results of the pretest participants. Students with code 11A showed that their initial ability was good even though the scores between the pretest and posttest did not increase. Other students experienced various improvements. The most normalized N-gain distribution is in the high criteria, which is 50%. There are two students in the medium criteria and four students in the low criteria. The highest percentage that falls into the high N-gain criteria indicates that students experience a high ability increase in terms of pretest and posttest results. Based on this statement, it can be concluded that students who get the low N-gain criteria indicate that these students experience an increase in abilities that are classified as low. The pretest and posttest scores which showed an increase after the implementation of assessment as learning in the competency test showed that assessment as learning was effective in improving the competence of students. Based on this research and the results of the Normalized N-gain which also showed an increase based on the results of the pretest and posttest scores, it can be concluded that the test instrument is effectively used as an assessment of learning and can increase the students' initial knowledge.

Student Response Questionnaire Results

Student response questionnaires are used to determine student responses to the developed test instrument. The results of the student response questionnaire can also be seen to determine the practicality of the developed instrument. Practicality can be seen from the student response questionnaire because it shows the student's assessment of the attractiveness of the test instrument, the ease of use of the test instrument and the clarity of instructions for working on the test instrument. Table 9 shows the percentages and categories based on the results of the student response questionnaire analysis.

Table 9. Results of Practical Analysis of Test Instruments

Student Code	Final Score	Percentage	Criteria
1A	29	73%	Practical
2A	27	68%	Practical
3A	36	90%	Very Practical
4A	32	80%	Practical
5A	29	73%	Practical
6A	30	75%	Practical
7A	33	83%	Practical
8A	29	73%	Practical
9A	35	88%	Very Practical
10A	33	83%	Practical
11A	36	90%	Very Practical
12A	29	73%	Practical
Average percentage		79%	Practical

Based on Table 9, the average percentage of the practicality of the test instruments obtained is 79% with practical criteria. The results of the analysis of student response questionnaires have practical criteria which indicate that the test instrument developed is interesting and easy to use. In a study conducted by, the results of the analysis of the student response questionnaire sheets that were in the practical category showed that the learning tools developed met the practicality criteria. The test instrument developed is considered practical and interesting to use to measure the level of science literacy of students.

Discussion

Assessment carried out during the learning process is important to do. The assessment can be through the use of test instruments, the results of which are then used as a benchmark for the learning achievement of students. Previous study stated that five conditions must comply to be a good instrument; validity, reliability, objectivity, practicality and economics (Arikunto & Suharsimi, 2009; Laliyo et al., 2019). To comply with the requirements of a good test instrument, the design of this test instrument will meet three requirements; validity, reliability and practicality. Validation of Test Instrument can be said to be good and feasible to use if it is declared valid. Validity is the level of accuracy of the data used to test the validity of the instrument used. Content and construction validity is the most important part of the process of designing test instruments (Junika et al., 2020; Sari et al., 2019). Content validity requires a match between the test equipment used as a measuring instrument and the ability to be measured (Priantini, 2020; Suciati et al., 2020). 16 question items are in the low criteria and 3 question items are in the very low criteria. The question items in the low and very criteria are considered invalid. There are 5 question items with very high criteria, 3 question items with high criteria and 12 question items with moderate criteria which are included in the category of valid. A test instrument can be said to be valid if the development of the test can accurately measure something to be measured. Therefore, it is necessary to determine the quality of question items that have low validity. Previous research states that the questions that can be used as a measuring tool for a study and are appropriate to be used for research data collection are the only questions that fall into the valid category (Csima et al., 2018; Elkordy, 2016).
The measurement results obtained in this study are the level of students' chemical literacy. Previous research states that a tool can measure certain abilities if the questions given are valid (Hairida, 2017; Rajendra & Sudana, 2017; Soeharto et al., 2019). Based on this statement, it can be concluded that the result of this study related to the questions included in the reliability analysis are only questions that include valid criteria. The students' chemical literacy will show how far the understanding of science and

the ability of students to apply this knowledge in every aspect of everyday life. Based on this statement, it can be concluded that if the achievement of the literacy level of students is at a low level, the understanding of knowledge and the ability of students to apply understanding to aspects of daily life is still very low. Based on the results of the test instruments that have been done, many students did not answer the questions, the answers given were wrong or the answers were misconceptions. Based on these results, the researchers concluded that the highest level achieved by students was the nominal level of scientific literacy. At this level, students are able to recognize a scientific concept and can answer questions briefly and correctly but their understanding is still limited and experience misconceptions.

It is in line with previous study which describes the development and evaluation of a chemistry critical thinking test, set in a chemistry context, and designed to be administered to undergraduate chemistry students at any level of study (Danczak et al., 2020). The studies suggest that the final version of the DOT test has good internal reliability, strong test–retest reliability, moderate convergent validity relative to a commercially available test and is independent of previous academic achievement and university of study. This article reports the development and validation of an instrument designed to evaluate students' metacognitive skillfulness in solving chemistry problems: the Metacognitive Activities Inventory (MCAI) (Cooper & Sandi-Urena, 2009). Reliability of the MCAI was measured in terms of internal consistency, and validity was examined in two dimensions: face validity, and construct validity. Evidence reported in this study indicates that the MCAI is a robust, reliable, and valid instrument.The implication of this research is to provide information related to the application of test instruments to measure students' level of Chemical Literacy. This is very important especially for educators, especially in the field of Chemistry because it can be used to measure the level of Chemical Literacy of students. However, this study still has several limitations, one of which is on the subject of the study because it only involves one school. Therefore, further research can be carried out by carrying out a trial phase in several high schools to determine the validity and reliability of the test instrument.

4. CONCLUSION

The assessment can be through the use of test instruments, the results of which are then used as a benchmark for the learning achievement of students. Validation of Test Instrument can be said to be good and feasible to use if it is declared valid. Based on the results, it can be concluded that the design of the test instrument to measure the level of chemical literacy of students on macromolecular material is feasible to use with the results of expert validation and limited trials. The design of the test instrument also shows that it is can be used as an assessment of learning and can increase students' initial knowledge.

5. REFERENCES

Akbar, S. (2013). *Instrument Perangkat Pembelajaran*. PT Remaja Rosdakarya.

Anandari, C. L., & Iswandari, Y. A. (2019). Extensive reading in Indonesian schools: a successful story. *TEFLIN Journal*, *30*(2), 137–152. https://doi.org/10.15639/teflinjournal.v30i2/137-152.

Arikunto, & Suharsimi. (2009). *Dasar-dasar evaluasi pendidikan*. PT Bumi Aksara.

Azwar, S. (2013). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Camel, V., Maillard, M. N., Piard, J., Dumas, C., Cladière, M., Fitoussi, G., Brun, E., Billault, I., & Sicard-Roselli, C. (2020). Chimactiv: An open access website for student centered learning in analytical chemistry. *Journal of Chemical Education*, *97*(8), 2319–2326. https://doi.org/10.1021/ACS.JCHEMED.0C00023.

Cooper, M. M., & Sandi-Urena, S. (2009). Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving. *Journal of Chemical Education*, *86*(2), 240. https://doi.org/10.1021/ed086p240.

Csima, M., Fináncz, J., Nyitrai, Á., & Podráczky, J. (2018). Research on the health literacy of professionals working in early childhood education. *Kontakt*, *20*(4), e356–e362. https://doi.org/10.1016/j.kontakt.2018.10.002.

Danczak, S. M., Thompson, C. D., & Overton, T. L. (2020). Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills. *Chemistry Education Research and Practice*, *21*(1), 62–78. https://doi.org/10.1039/C8RP00130H.

Duit, R., Harald, G., Kattmann, U., Komorek, M., & Lka, P. (2012). The Model of Educational Reconstruction-A Framework for Improving Teaching and Learning Science. *In Science Education Research and Practice in Europe*, 13–37. https://doi.org/10.1007/978-94-6091-900-8_2.

El Islami, R. A. Z., & Nuangchalerm, P. (2020). Comparative study of scientific literacy: Indonesian and thai pre-service science teachers report. *International Journal of Evaluation and Research in Education*, *9*(2), 261–268. https://doi.org/10.11591/ijere.v9i2.20355.

Elkordy, A. (2016). Development and implementation of digital badges for learning science, technologly, engineering and math (STEM) practices in secondary contexts: A pedagogical approach with empirical evidence. In *Foundation of Digital Badges and Micro-Credentials: Demonstrating and Recognizing Knowledge and Competencies* (pp. 483–508). Springer International Publishing. https://doi.org/10.1007/978-3-319-15425-1_27.

Fahmina, S., Indriyanti, N., Setyowati, W., Masyukuri, M., & Yamtinah, S. (2019). Dimension of Chemical Literacy and its Influence in Chemistry Learning. *Series: Journal of Physics*, 1–8. https://doi.org/10.1088/1742-6596/1233/1/012026.

Giri, I. M. A., & Ardiawan, I. K. N. (2018). The integration of character education in local tradition: an ethnographic study. *In SHS Web of Conferences*. https://doi.org/10.1051/shsconf/20184200082

Hairida, H. (2017). Pengembangan Instrumen Untuk Mengukur Self Efficacy Siswa Dalam Pembelajaran Kimia. *Edusains*, *9*(1), 53–59. https://doi.org/10.15408/es.v9i1.4000.

Hidayah, I. R., Kusmayadi, T. A., & Fitriana, L. (2021). Minimum Competency Assessment (AKM): An Effort To Photograph Numeracy. *Journal of Mathematics and Mathematics Education*, *11*(1), 14–20. https://doi.org/10.20961/jmme.v11i1.52742.

Hong, L. Y., & Talib, C. A. (2018). Scientific Argumentation in Chemistry Education: Implications and Suggestions. *Asian Social Science*, *14*(11), 16. https://doi.org/10.5539/ass.v14n11p16.

Hottecke, D., & Allchin, D. (2020). Reconceptualizing nature-of-science education in the age of social media. *Science Education*, *104*(4), 641–666. https://doi.org/10.1002/SCE.21575.

Junika, N., Izzati, N., & Tambunan, L. R. (2020). Pengembangan soal statistika model PISA untuk melatih kemampuan literasi statistika siswa. *Mosharafa: Jurnal Pendidikan Matematika*, *9*(3), 499–510. https://doi.org/10.31980/mosharafa.v9i3.615.

Kholis, N., Kartowagiran, B., & Mardapi, D. (2020). Development and validation of an instrument to measure a performance of vocational high school. *European Journal of Educational Research*, *9*(3), 955–966. https://doi.org/10.12973/EU-JER.9.3.955.

Laliyo, L. A. R., Botutihe, D. N., & Panigoro, C. (2019). The Development of Two-Tier Instrument Based On Distractor to Assess Conceptual Understanding Level and Student Misconceptions in Explaining Redox Reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12.

Mozeika, D., & Bilbokaite, R. (2010). Teaching and learning method for enhancing 15–16 years old students' knowledge as one of scientific literacy aspect in chemistry: Results based on research and approbation. *The International Journal of Educational Researchers*, *3*(2), 1–16. https://dergipark.org.tr/en/pub/ijers/issue/8489/105626.

Muntholib, Ibnu, S., Rahayu, S., Fajaroh, F., Kusairi, S., & Kuswandi, B. (2020). Chemical Literacy: Performance of First Year Chemistry Students on Chemical Kinetics. *Indones. J. Chem*, *20*(2), 468–482. https://doi.org/10.22146/ijc.43651.

Mutakinati, L., Anwari, I., & Yoshisuke, K. (2018). Analysis of students' critical thinking skill of middle school through stem education project-based learning. *Jurnal Pendidikan IPA Indonesia*, *7*(1), 54–65. https://doi.org/10.15294/jpii.v7i1.10495.

Nugrahanto, S., & Zuchdi, D. (2019). *Indonesia PISA Result and Impact on The Reading Learning Program in Indonesia*. *297*(Icille 2018), 373–377. https://doi.org/10.2991/icille-18.2019.77.

Nursa'adah, E., Liliasari, L., Mudzakir, A., & Barke, H. (2018). The Model of Educational Reconstruction: Students' Conceptual Knowledge on Solid State Chemistry Domain. *Jurnal Pendidikan IPA Indonesia*, *7*(2), 193–203. https://doi.org/10.15294/jpii.v7i2.14297.

Oktariani, & Febliza, A. (2019). Analysis of The Alignment Between Chemistry Content on TIMSS And Science-Chemistry Textbooks of Junior High School. *Jurnal Pendidikan Indonesia (JPI)*, *8*(1), 123–132. https://doi.org/10.23887/jpi-undiksha.v8i1.16101.

Pratiwi, S., Cari, C., & Aminah, N. (2019). Pembelajaran IPA Abad 21 dengan Literasi Sains Siswa. *Jurnal Materi Dan Pembelajaran Fisika (JMPF)*, *9*(1), 34–42. https://doi.org/10.20961/jmpf.v9i1.31612.

Priantini, D. A. M. M. O. (2020). The Development of Teaching Video Media Based on Tri Kaya Parisudha in Educational Psychology Courses. *Dewa Ayu Made Manu Okta Priantini*, *4*(4), 448–455. https://doi.org/10.23887/jet.v4i4.29608.

Rajendra, I. M., & Sudana, I. M. (2017). The Influence of Interactive Multimedia Technology to Enhance Achievement Students on Practice Skills in Mechanical Technology. *Journal of Physics*, *953*, 1–5. https://doi.org/10.1088/1742-6596/953/1/012104.

Rasmawan, R. (2018). Development of Chemistry Module for Junior High School Based on Inquiry Accompanied by Performance-Based Assessment. *Jurnal Pendidikan Indonesia*, *7*(2), 111–119. https://doi.org/10.23887/jpi-undiksha.v7i2.10617.

Rosmani, & Halim, A. (2017). Analisis Perbandingan Hasil Belajar Kimia Siswa Terhadap Penerapan Ktsp Dan Kurikulum 2013 Di Beberapa Sekolah Favorit Banda Aceh. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, *05*(01), 94–101. http://www.e-repository.unsyiah.ac.id/JPSI/article/view/8417.

Rostikawato, D. (2016). Rekonstruksi Bahan Ajar dengan Konteks Socio-Scientific Issues pada Materi Zat Aditif Makanan untuk Mengingkatkan Literasi Sains Siswa. *Jurnal Inovasi Pendidikan IPA*, *2*(2), 156–164. https://doi.org/10.21831/jipi.v2i2.8814.

Sari, D. A., Ellizar, E., & Azhar, M. (2019). Development of problem-based learning module on electrolyte and nonelectrolyte solution to improve critical thinking ability. *Journal of Physics: Conference Series*, *1185*(1). https://doi.org/10.1088/1742-6596/1185/1/012146.

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://pubs.rsc.org/en/content/articlehtml/2006/rp/b6rp90011a.

Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649.

Suciati, Munadi, S., Sugiman, & Febriyanti, W. D. R. (2020). Design and validation of mathematical literacy instruments for assessment for learning in Indonesia. *European Journal of Educational Research*, *9*(2), 865–875. https://doi.org/10.12973/eu-jer.9.2.865.

Tesi Muskania, R., & Wilujeng, I. (2017). Pengembangan Perangkat Pembelajaran Project-Based Learning Untuk Membekali Foundational Knowledge Dan Meningkatkanscientificliteracy. *Jurnal Cakrawala Pendidikan*, *36*(1), 34–43. https://doi.org/10.21831/cp.v36i1.8830.

Thummathong, R., & Thathong, K. (2018). Chemical literacy levels of engineering students in Northeastern Thailand. *Kasetsart Journal of Social Sciences*, *39*(3), 478–487. https://doi.org/10.1016/j.kjss.2018.06.009.

Yusmaita, E., & Nasra, E. (2018). Design of Chemical Literacy Assessment by Using Model of Educational Reconstruction (MER) on Solubility Topic. *In IOP Conference Series: Materials Science and Engineering*, 335. https://iopscience.iop.org/article/10.1088/1757-899X/335/1/012106/meta.