

DEVELOPMENT OF PISA 2015 BASED CHEMICAL LITERACY ASSESSMENT INSTRUMENT FOR HIGH SCHOOL STUDENTS

Fika Aning Tiara, Oktavia Sulistina*

Department of Chemistry, Faculty of Mathematics and Natural Science, Universitas Negeri Malang. Jl. Semarang No. 5 Malang, Indonesia

Abstract – This study aims to develop valid and reliable chemical literacy assessment instruments based on PISA 2015. The development procedures carried out were 1) research and information collecting, 2) planning, 3) development preliminary form of product, 4) preliminary field testing, and 5) main product revision. Instrument of development result was validated (content validity and empirical validity). Content validity assessment data was obtained from the validity test results from two chemistry lecturers. Empirical validity test data were acquired from 68 grade XI students as test subjects who came from five high schools in Malang. An empirical validity test was used to obtain the level of validity, reliability, discrimination index, difficulty level, and effectiveness of distractors of the items developed in the instrument. The instrument of development results consisted of 20 multiple choice items and 4 attitude questionnaires. The results of the content validity test indicated a valid instrument (the average score for the aspects of substance, construction, and language was 83.9). The results of the empirical validity test showed that multiple-choice items had a correlation value of 0.37-0.77, categorised as valid, and the reliability value was 0.86, classified as highly reliable. The discrimination index obtained was five items ranked as sufficiently good and 15 items categorised as good, while five items classified as easy item, 14 moderate items, and one difficult item, all distractors were functioning. The empirical validity test results in the form of an attitude questionnaire showed a correlation value of 0.65-0.69, so they were valid, and the reliability value was 0.59, classified as quite high criteria. Instrument development results proved to be valid and reliable, so it is feasible to be used to measure students' chemical literacy skills.

Keywords: *Assessment instruments; Chemical literacy; PISA 2015*

INTRODUCTION

The 21st century or globalisation era carries complex demands and challenges (Mudiono, 2016). In this era, people are expected to possess 21st-century skills, which according to The World Economic Forum (2016), is consisted of fundamental literacy, competencies, and qualified characters. Those skills are required to overcome the challenges offered by the 21st century, such as the use of technology and energy, the management of climate change and

*Corresponding author: Program Studi Pendidikan Kimia, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Malang, Jawa Timur, Indonesia. Email: oktavia.sulistina.fmipa@um.ac.id

diseases, as well as water and food availability for the society that concentrate mainly on science and technology (United Nations Environment Programme, 2012). Consequently, individuals with understanding and ability to face challenges and solve life issues should be prepared (Thummathong & Thathong, 2016). That preparation emphasises the need for scientific literacy (Bond, 1989). The complex society advancement has demanded each individual to enhance their scientific literacy (Vogelzang et al., 2020).

Most educational initiative mentions that scientific literacy is crucial for social prosperity and individual ability to function in the scientific and technological dominated world (Shwartz, Ben-Zvi, Hofstein, 2005). Scientific literacy allows each individual to make a rational decision for every science and technology-related issue (Thummathong & Thathong, 2016). The term scientific literacy represents students' ability to comprehend, utilise, and implement science (Norris & Phillips, 2003). It covers the ability to scientifically explain phenomena, evaluate and design scientific inquiry, as well as scientifically interpret data and evidence (Organisation for Economic Co-operation and Development, 2016). Shwartz et al. (2005) and DeBoer (2000) in Celik (2014), explain that currently, there has not been a consensus on the scientific literacy definition. However, almost every definition of scientific literacy accentuates the ability to distinctively comprehend and explain phenomena, read and write to evaluate information, communicate the idea to other people, and implement scientific knowledge and reasoning in the daily life and decision-making process (Cigdemoglu et al., 2017).

The realisation of a scientifically literate society is the primary objective of science education (Norris & Philips, 2003; Vogelzang et al., 2020). Scientific literacy is the target of science learning reformation and the primary objective of science education (American Association for the Advancement of Science (AAAS), 1993). Science education reformation, standards, and curriculum across countries emphasise developing students' scientific literacy to function in the current technological society (Mumba & Hunter, 2009). In the last decades, scientific literacy has been measured using various assessment instruments (Naganuma, 2017). One of the international level programs that investigate students' scientific literacy is PISA (Programme of International Student Assessment). It examines students' ability to use scientific knowledge and skills (Naganuma, 2017). Besides, it also explores the level of essential skills and knowledge obtained by students to attain their success in modern society and the economic system (She et al., 2018). The scientific literacy scoring in PISA highlights the implementation of scientific knowledge in a real-life context (Organisation for Economic Co-operation and Development, 2016). Consequently, PISA has gained attention from researchers and public

policymakers in the educational field (She et al., 2018). Indonesia has reformed its national curriculum into Curriculum 2013 (Rahayu, 2017), which aims to produce a scientifically literate society (Rahayu, 2014). In the last ten years, this reformation is completed since Indonesia has gained the second rank out of the four lowest positions in its PISA participation, confirming its very low scientific literacy (Rahayu, 2014). The latest scientific literacy assessment conducted by PISA in 2018 still places Indonesia in the sixth-lowest rank (Organisation for Economic Co-operation and Development, 2018). Similar to Indonesia, many other countries have also reformed their education standards and curriculums. For instance, the USA, England, China, and Zambia have accentuated students' scientific literacy improvement (Mumba & Hunter, 2009).

USA's educational standards and world chemistry teachers have emphasised the importance of students' scientific literacy development, primarily on chemical literacy (Kohen et al., 2019). Chemical literacy is the ability to use chemistry in various relevant contexts (Shwartz, 2005). It covers knowledge and skills on chemistry required in the comprehension of chemistry-based socio-scientific issues (Kohen et al., 2020). A person with great chemical literacy understand the primary chemical idea, recognise the significance of chemistry in explaining daily phenomena, understand the connection between chemistry and socio-culture, demonstrate an interest in chemical issues, use chemical understanding in their daily life as customers, making a decision, and participate in social debate (Shwartz, 2005). Students and general society need to attain chemical literacy since it affects their social and personal decision-making process (Avargil et al., 2013; Dori et al., 2018 as cited in Kohen et al., 2019). Basic chemical comprehension is expected to contribute to scientific literacy that is perceived as the main objective of science education (Cigdemoglu & Geban, 2015).

The development of a chemical literacy assessment instrument is an effort to enhance students' chemical literacy. Students' competence (including chemical literacy) can be improved through an assessment since it assesses students' learning (assessment of learning), while also enhances students' competence (assessment for learning and assessment as learning) (Direktorat Pembinaan SMA, 2017). Students' chemical literacy measurement can be carried out using frameworks similar to PISA (Cigdemoglu, 2017; Rahayu, 2017). PISA 2015 framework has four aspects that are assessed in the scientific literacy assessment, consisting of context (personal, local, global), knowledge (content, procedural, epistemic), competence (explain a phenomenon scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically), and attitude (interest in science, valuing a scientific approaches to enquiry, and environmental awareness) (Organisation for Economic Co-operation and Development, 2016). PISA 2015

investigates students' essential knowledge and skill level, as well as students' attitude toward science (She et al., 2018). The science literacy framework facilitates the development of chemical literacy instruments. Thus, it contributes to the students' chemical literacy improvement, specifically, and their scientific literacy, generally. This study aims to generate valid and reliable chemical literacy assessment instruments based on PISA 2015.

METHOD

This research and development used the Borg and Gall model, with stages of (1) research and data collection, (2) planning, (3) initial product draft development, (4) initial field trial, and (5) revision based on the trial process. In the research and data collection phase, various literatures related to PISA, scientific literacy, chemical literacy, and relevant previous studies were reviewed. Besides, daily life issues relevant to chemistry materials were also identified. In the planning stage, the research objectives, validator, trial subjects, and product design were determined. Meanwhile, in the initial product development stage, a chemical literacy instrument was established. The instrument consisted of a translated summary of the PISA 2015 framework, the instrument outline, the test manuscript subsisting of 20 multiple-choice items and four attitude questionnaires, test instruction, answer key, and discussion, along with the scoring guide. The developed initial product was later validated by two validators who are chemistry lecturers. The obtained data were analysed using percentage calculation. The product propriety was determined by validity criteria from the percentage analysis (Riduwan et al., 2013). The draft's parts verified to be invalid were revised through descriptive analysis on the validators' comments and suggestions. The valid draft was tried out to students' online, using Google Form media, for duration of 90 minutes. The sample selection was completed using the random sampling technique. The try-out involved 68 11th grade senior high school students in Malang. The obtained data were empirically validated to attain the instrument items' validity, reliability, difficulty, discrimination, and distractor effectiveness. The items' validity was concluded using product-moment correlation criteria. Meanwhile, the reliability, item discrimination, item difficulty, and distractors effectiveness were determined using criteria from Arikunto (2012). In the revision based on the try-out results, the items with low validity, reliability, difficulty, discrimination, and distractor effectiveness were improved based on the selected criteria.

RESULTS AND DISCUSSION

The product generated in this study is a PISA 2015 based chemical literacy assessment instrument for high school students. The product consists of (1) cover, (2) preface, (3)

instruction for use, (4) table of content, (5) PISA 2015 scientific literacy framework, (6) question outline, (7) instrument manuscript, (8) answer key, (9) scoring guide, and (10) references. The developed instrument refers to the scientific literacy aspects in the PISA 2015 scientific literacy framework. Those aspects include context, competence, knowledge, and attitude. The context, competence, and knowledge aspect can be measured through multiple-choice items. Meanwhile, the attitude aspect is measured using a attitude questionnaire. Thus, the instrument consists of 20 multiple choice items and four attitude questionnaire items. The chemistry materials adopted in this study are chemical bonds, electrolyte and non-electrolyte solutions and thermochemistry.

In PISA 2015, scientific literacy is broken down into four aspects: context, competence, knowledge, and attitude. In this framework, students are confronted in personal context situations, as well as the local or national and global situation. These contexts are presented in various fields of health and diseases, natural resources, environmental quality, hazard, as well as science and technology limitations. Those fields aid individuals to attain particular values from scientific literacy in improving their life and public policy quality (OECD, 2016). This instrument presents different phenomena in the environmental quality, health, and natural resources, in the personal and local context. In the ecological quality field, two topics are adopted as the primary issues, namely oil spills at sea and stun fish catch. In the health sector, the problems selected are related to fat, carbohydrates, and energy. Meanwhile, in the energy and resources, the topic chosen in this instrument includes the transition of kerosene fuel to liquid petroleum gas (LPG). All of those selected issues and phenomena come from daily problems that students frequently encountered.

Someone with good scientific literacy is inclined to discuss technology and science that requires some competence to scientifically describe phenomena, evaluate and design scientific inquiry, and scientifically interpret data and evidence (OECD, 2016). The deployment of those competence aspects within the instrument is presented in Table 1.

Table 1. Competence aspects deployment in the developed instrument

Competence	Indicator	Item number
Explaining a phenomena scientifically	Identify and implement the proper scientific knowledge	2.3, 3.1, 3.5
	Identify, use, and generate an explanation	2.5
	Describe the effect of scientific knowledge on the social environment	3.4, 4.3
	Create and enhance a prediction	4.2
Evaluate and design scientific inquiry	Identify the investigated question in the provided scientific study	1.3
	Propose methodology to analyse the scientific question	2.2
	Describe and evaluate the way a researcher ensure reliable data and generalise the explanation	2.4, 4.1
	Differentiate questions that can be scientifically investigated	2.6
	Evaluate a method used in investigating the scientific question	4.4
Interpret data and evidence scientifically	Identify assumption, evidence, and reasoning related to science	1.1
	Analyse, interpret data and make the correct conclusion	1.2, 1.4
	Differentiate argumentation with scientific theory and evidence from argumentation made based on other considerations	1.5
	Evaluate scientific argument and facts from different sources	2.1
	Modify data to other forms	3.2, 3.3

The scientifically explaining a phenomena competence requires students to recall the relevant knowledge to explain phenomena. In this instrument, students have to use their knowledge on the cause of solution ability to deliver electricity, energy transformation from food processing in the body, and calculation of heat combustion from fuel. The students can use these pieces of information to describe the relevant phenomena in their surroundings, such as the reason why stun fish catching is prohibited, how food produces energy, and the effect of energy knowledge on the fuel subsidy saving. The causal effect of phenomena can be established from an experiment with a valid procedure to obtain new knowledge (OECD, 2016). The evaluation and criticism of scientific findings and scientific investigation require the competence of evaluate and design scientific inquiry. Knowledge of scientific investigation is expected to show this competence. In this instrument, students need to know the objectives of an experiment design, involved variables in a scientific investigation, attempt to minimise the uncertainty of the data (reliability) in a scientific investigation, differentiate questions that can be investigated scientifically, the involvement of assumptions on an experiment, and the ways to evaluate an experiment design. Scientific data and evidence that supports a claim and conclusion should be analysed and interpreted. Therefore, the competence to scientifically interpret data and evidence is also required. Besides, this competence is also demanded in evaluating arguments and

conclusions based on scientific evidence (Osborne, 2010). This instrument expects students to provide the proper reasoning based on the supplied assumption and evidence, determine the number of lone pair in one triglyceride molecule after they analyse the molecular structure, analyse someone's argumentation on the oil spill at sea and the bacterial degradation, while also conclude if that argument is made based on scientific theory and evidence. Besides, they are also expected to evaluate someone's argument related to the destruction of the aquatic biota due to the stunning fishing and transform the argument on palmitic acid and sucrose combustion into energy level diagram and thermochemical equations.

As presented in Table 1, the measured competencies involved the content, procedural, and epistemic knowledge. Students should comprehend the universe, along with the facts, concepts, ideas, and theories that become the fundamental of science (OECD, 2016). Therefore, students can use this knowledge to explain phenomena. To attain valid and reliable data, scientists use standard procedures supported by procedural knowledge (OECD, 2016). Procedural knowledge is also used to review evidence that supports a particular claim. Scientific investigation requires knowledge to guide the investigation process. The investigation procedure and practice need fundamental, while scientific claim needs a foundation of trust. All of those requirements are covered in epistemic knowledge (OECD, 2016). The deployment of knowledge aspect in the developed instrument is shown in Table 2.

Table 2. Knowledge aspects deployment in the developed instrument

Knowledge	Indicator	Item number
Content	Material structure (bond) [chemical bond]	1.1, 1.2
	Materials property (electric current) [Electrolyte and non-electrolyte solutions]	2.3, 2.5
	Chemical reaction (energy transfer)[Thermochemistry]	3.1, 3.3, 4.2
Procedural	Concept on independent, dependent, and control variables	1.4
	Strategy to control variable	2.2
	Ways to evaluate and minimise unpredictability, such as recalculate measurement and the average point	2.4
	Presenting data in the form of graphic/table/ diagram and properly use them	3.2
	Mechanism to ensure the data accuracy	4.1
	Design that follows the provided scientific question	4.4
Epistemic	The empirical investigation's function in developing the selected knowledge, purpose, and design	1.3
	Scientific observation, facts, hypotheses, models, and theories	1.5, 3.5
	The role of collaboration and criticism, along with peer assessment to generate trust in the scientific claim	2.1
	The purpose and objective of science, technology, and elements related to the scientific and technological question	2.6, 3.4
	The scientific claim that is supported by scientific data and reasoning	4.3

The developed multiple-choice items are equipped with the confidence level scale for their answer. This scale is provided to improve students' answer accuracy. The example of an item that discussed chemical bond material is illustrated in Figure 1.

1.1 Setelah membaca wacana di atas, seorang siswa berasumsi bahwa minyak merupakan senyawa kovalen nonpolar. Asumsi tersebut didasarkan atas bukti bahwa minyak tidak dapat larut dalam air dan struktur molekul minyak yang menunjukkan ikatan kovalen. Penalaran yang tepat yang Anda berikan terhadap asumsi dan bukti tersebut ditinjau dari jenis ikatan dan kepolaran adalah

	Jenis ikatan	Kepolaran
A.	Ikatan C dengan H, C dengan O, C dengan C terbentuk karena perpindahan elektron antaratom	Kepolaran minyak berbeda dengan kepolaran air
B.	Ikatan C dengan H, C dengan O, C dengan C terbentuk karena perpindahan elektron antaratom	Minyak didominasi oleh ikatan C-H, perbedaan keelektronegatifannya kecil (kurang dari 0,5)
C.	Ikatan C dengan H, C dengan O, C dengan C terbentuk karena penggunaan elektron bersama dari kedua atom yang berikatan maupun dari salah satu atom saja	Kepolaran minyak berbeda dengan kepolaran air
D.	Ikatan C dengan H, C dengan O, C dengan C terbentuk karena penggunaan elektron bersama dari kedua atom yang berikatan	Minyak didominasi oleh ikatan C-H, perbedaan keelektronegatifannya kecil (kurang dari 0,5)

A. menebak
B. kurang yakin
C. yakin
D. sangat yakin

Figure 1. Multiple-choice items on chemical bond material

The example of item on electrolyte and non-electrolyte solutions material is illustrated in Figure 2.

2.1 Seorang siswa bernama Kiki tidak yakin penyebab rusaknya biota air adalah setrum. Kiki berargumen bahwa setrum yang tegangannya tidak tinggi akan terbatas alirannya di dalam air sungai sehingga hanya ikan yang tubuhnya terkena alat setrum secara langsung yang akan mati, setrum tidak akan mengalir sampai ke biota air disekitarnya. Pendapat anda terkait pernyataan Kiki adalah

A. setuju. Setrum dengan tegangan kecil memang bisa melumpuhkan ikan, tapi tidak sampai merusak biota air

B. setuju. Rusaknya biota air bukan dari setrum melainkan faktor lain seperti kuat arus aliran air sungai, lagipula air tidak dapat menghantarkan aliran listrik sehingga hanya ikan yang terkena alat setrum secara langsung yang akan mati

C. tidak setuju. Setrum dalam air sungai dapat melumpuhkan dan merusak biota air di sekitarnya karena air sungai dapat menghantarkan listrik

D. tidak setuju. Berapapun tegangan setrum yang digunakan tetap dapat merusak seluruh biota air di sepanjang sungai karena air sungai dapat menghantarkan listrik.

Figure 2. The multiple-choice item on electrolyte and non-electrolyte solutions

The example of multiple-choice items that discussed the thermochemical material is shown in Figure 3.

4.4 Andi selanjutnya ingin mengetahui dengan pasti dan membandingkan kalor yang dihasilkan dari pembakaran bahan bakar dengan melakukan percobaan. Namun, bahan bakar yang Andi temukan adalah minyak tanah dan solar. Berikut ini rancangan prosedur kerja dari percobaan yang akan dilakukan oleh Andi

Prosedur kerja

1. Timbanglah 100 mL air, lalu masukkan ke dalam gelas beker. Ukur dan catat suhu air
2. Timbang pembakar spiritus kosong
3. Isilah pembakar spiritus dengan 50 mL minyak tanah
4. Timbang dan catat massa pembakar spiritus dan minyak tanah
5. Nyalakan pembakar spiritus yang berisi minyak tanah
6. Masukkan termometer ke dalam air
7. Panaskan air di atas spiritus sambil diaduk menggunakan batang pengaduk
8. Hentikan pembakaran dengan mematikan pembakar tepat sebelum air mendidih atau suhu mendekati 100°C. Ukur dan catat suhu air
9. Ulangi percobaan dengan mengganti minyak tanah dengan solar

Hal yang perlu dievaluasi dari rancangan percobaan Andi adalah

A. langkah kedua tidak perlu dilakukan karena yang diperlukan adalah massa minyak tanah bukan massa spiritus kosong

B. pengadukan air pada langkah ketujuh tidak perlu dilakukan karena tidak akan memberi pengaruh terhadap pemanasan air

C. sebelum mengulangi percobaan untuk solar, seharusnya dilakukan penimbangan pembakar spiritus dan sisa minyak tanah

D. pencatatan suhu air pada langkah pertama tidak perlu dilakukan karena suhu mulai diukur ketika air dipanaskan dan seharusnya langkah pertama dilakukan setelah langkah kelima

A. menebak
B. kurang yakin
C. yakin
D. sangat yakin

Figure 3. The multiple-choice item on thermochemical material

In addition to those items, four attitude questionnaires were also developed. The attitude aspects measured in the questionnaire include interest in chemistry, rate chemical approaches for inquiry, and environmental concern. The questionnaire that measures students' interest in chemistry, rating chemical approaches for inquiry, and environmental concern are shown in Figures 4, 5, and 6, respectively.

4.5 Beri tanda centang (√) pada kolom yang disediakan!
 SS = sangat setuju R = ragu-ragu STS = sangat tidak setuju
 S = setuju TS = tidak setuju

Pernyataan	SS	S	R	TS	STS
Saya tertarik melakukan perhitungan energi berbagai macam bahan bakar dengan menerapkan konsep termokimia yang saya miliki					

Figure 4. Attitude questionnaire on interest toward chemistry

3.6 Beri tanda centang (√) pada kolom yang disediakan!
 SS = sangat setuju R = ragu-ragu STS = sangat tidak setuju
 S = setuju TS = tidak setuju

Pernyataan	SS	S	R	TS	STS
Penentuan besarnya energi yang dihasilkan dari pembakaran bahan makanan dapat diketahui dari percobaan menggunakan bom kalorimeter dengan serangkaian metode ilmiah					

Figure 5. Attitude questionnaire on students' rating on chemical approaches for inquiry

2.7 Beri tanda centang (√) pada kolom yang disediakan!
 SS = sangat setuju R = ragu-ragu STS = sangat tidak setuju
 S = setuju TS = tidak setuju

Pernyataan	SS	S	R	TS	STS
Jika saya menemui orang yang menangkap ikan dengan setrum, saya akan menegur/melaporkannya					

Figure 6. Attitude questionnaire on students' environmental concern

The developed initial product has been validated by two validators. The validity test was carried out by completing the product's feasibility questionnaire divided into three parts of (1) general instrument validity test, (2) validity test on each multiple-choice item, and (3) validity test on the attitude questionnaire item. The elements assessed in the general instrument validity test involve the instruction for use, display, layout, and a scoring guide. Meanwhile, the aspects evaluated in the multiple choices and attitude questionnaire items validity test include their content, construct, and language. Each of those aspects has indicators. The validators assessed the product using a five scale score, referring to the Likert Scale explained by Riduwan (2011).

Table 3. Likert scale explained by Riduwan

Scale	Description
5	Strongly feasible/ very suitable/very good/ strongly agree
4	feasible/ suitable / good / agree
3	Sufficiently feasible / sufficiently suitable / sufficiently agree /neither agree nor disagree
2	Less feasible / less suitable / less good /agree
1	Highly not feasible/very unsuitable/poor/strongly disagree

The scores assigned by the validator have been analysed using percentage calculation by dividing the total score with the highest score and multiply by 100. The product feasibility is determined

by the validity criteria of the percentage analysis from Riduwan, et al. (2013). The product classification is divided into very feasible (81-100 score), feasible (61-80), sufficiently feasible (41-60), less feasible (21-40), and not feasible (0-20) (Riduwan, 2013). The results of the instrument's general validity test are shown in Table 4.

Table 4. Results of instrument's general validity test

Aspects	Validity score		Average score	Description
	Validator 1	Validator 2		
Instruction for use	80	80	80	Feasible
Display and layout	80	95	87.5	Very Feasible
Scoring guide	80	90	85	Very Feasible
The total average			84.2	Very Feasible

The instrument's general validity assessment result shows that the product's instruction of use, display, layout, and scoring guide are very feasible to be implemented. The validity test results on the multiple-choice and attitude questionnaire items are presented in Table 5.

Table 5. Validity test results on the multiple-choice and attitude questionnaire items

Aspect	Validity Score		Average score	Description
	Validator 1	Validator 2		
Multiple choice				
Content	74.5	86.2	80.4	Feasible
Construct	75.6	80.6	78.1	Feasible
Language	75	89.5	82.2	Very Feasible
Attitude questionnaire				
Content	80	90	85	Very Feasible
Construct	80	95	87.5	Very Feasible
Language	85	95	90	Very Feasible
The total average			83.9	Very Feasible

The instrument validity test results from the content, construct, and language aspects indicate that the multiple-choice and attitude questionnaire items are very feasible to be implemented. Some of the items were revised in accordance with the validators' suggestions and comments.

After some items have been revised, the instrument was empirically validated by involving 68 11th grade students from five state senior high schools in Malang. The obtained data is in the form of students' answers to the multiple-choice questions and responses to the attitude questionnaire. The students' answers were scored. The students' correct multiple choice answers were scored 1, while the wrong answers were scored 0. At the same time, students' responses to

the questionnaire were scored 1-5. Those data were analysed to discover the instrument items' validity, reliability, difficulty, discrimination, and distractor effectiveness.

Validity

The item validity has been analysed using Pearson product-moment correlation using Microsoft excel to calculate the r_{count} . The selected significance level is 0.05 with 68 students so that the r_{table} is 0.24. The items are categorised as valid if $r_{\text{count}} > r_{\text{table}}$. The items validity data are divided into multiple-choice items data and attitude questionnaire items data, as presented in Table 6.

Table 6. Results of multiple-choice and attitude questionnaire items' validity test

Item Number	Value of r_{xy}	Description	Item No.	Value of r_{xy}	Description
Multiple Choice					
1.1	0.37	Valid	2.6	0.42	Valid
1.2	0.45	Valid	3.1	0.42	Valid
1.3	0.44	Valid	3.2	0.62	Valid
1.4	0.61	Valid	3.3	0.77	Valid
1.5	0.45	Valid	3.4	0.43	Valid
2.1	0.55	Valid	3.5	0.60	Valid
2.2	0.52	Valid	4.1	0.57	Valid
2.3	0.55	Valid	4.2	0.48	Valid
2.4	0.46	Valid	4.3	0.50	Valid
2.5	0.52	Valid	4.4	0.64	Valid
Attitude Questionnaire					
1.6	0.67	Valid	3.6	0.65	Valid
2.7	0.67	Valid	4.5	0.69	Valid

According to the analysis results, the developed multiple-choice and attitude questionnaire item scores range from 0.37 to 0.77, classified as valid. Therefore, the developed instrument can be used to measure students' chemical literacy.

Item Discrimination

Item discrimination analysis has been completed on the multiple-choice items by calculating the discriminating index of each item (D). The discriminatory power of an item is divided into very good (D 0.71-1.00), good (D 0.41-0.70), sufficiently good (D 0.21-0.40), poor (D 0.00-0.20), and eliminated (D < 0.00) (Arikunto, 2012). The results of item discrimination analysis are presented in Table 7.

Table 7. Results of item discrimination analysis on the multiple-choice item

Item	D(Discriminating Index)	Description	Item	D(Discriminating Index)	Description
1.1	0.26	Sufficient	2.6	0.29	Sufficient
1.2	0.41	Good	3.1	0.38	Sufficient
1.3	0.41	Good	3.2	0.56	Good
1.4	0.50	Good	3.3	0.71	Good
1.5	0.44	Good	3.4	0.35	Sufficient
2.1	0.56	Good	3.5	0.56	Good
2.2	0.41	Good	4.1	0.53	Good
2.3	0.44	Good	4.2	0.41	Good
2.4	0.35	Sufficient	4.3	0.53	Good
2.5	0.53	Good	4.4	0.71	Good

The results of item discrimination analysis confirm that the developed multiple-choice items have sufficiently good and good discrimination power. Five items are observed to have sufficiently good discriminating power, while the other 15 items have good discriminatory power. Consequently, all developed items have the ability to differentiate high achieving and low achieving students.

Item Difficulty

Item difficulty analysis discovers the difficulty level of the multiple-choice items. Therefore, this analysis was not carried out on the attitude questionnaire items since they are presented in the form of statements. This analysis was carried out by calculating the items' difficulty index (P). The results are classified into difficult items (P 0.00-0.30), moderate items (P 0.31-0.70), and easy items (P 0.71-1.00) (Arikunto, 2012). The analysis results show that five items (25%) are categorised as easy, 14 items (70%) are categorised as moderate, and 1 item (5%) are categorised as difficult. The suggested difficulty index ranges from 0.31-0.70, classified as moderate, but the easy and difficult items still can be adopted (Arikunto, 2012). The graphic of the instrument's difficulty index distribution is presented in Figure 7.

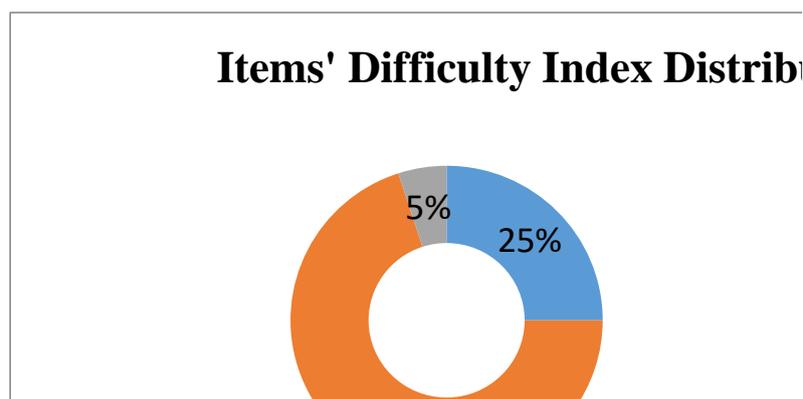


Figure 7. Distribution of Item Difficulty Index

Distractor Effectiveness

The distractor effectiveness analysis was also completed on the multiple-choice items. The distractor with proper function should have been selected by at least 5% of the test participants (Arikunto, 2012). The analysis results indicate that all distractors have properly functioned since they have been chosen by a minimum of 5% of test participants.

Reliability

The instrument's reliability analysis was completed on both multiple-choice and attitude questionnaire items. The analysis formula used for those two types of items is different. The reliability analysis on the multiple-choice items was carried out using the KR-20 formula. Meanwhile, the analysis on the questionnaire items was conducted using Cronbach's alpha. The instrument's reliability has been determined using the test reliability criteria from Arikunto (2012).

Table 8. Test's reliability Criteria

Reliability Coefficient	Category
0.800 – 1.00	Very High
0.600 – 0.800	High
0.400 – 0.600	Sufficient
0.200 – 0.400	Low
0.00 – 0.200	Very Low

The multiple-choice reliability analysis result shows a 0.86 reliability coefficient so that the developed multiple-choice items have very high reliability. On the other hand, the reliability of the attitude questionnaire is 0.59. Therefore, the developed instrument produces reliable results if it is tested on the same subjects at different times. Good instruments should provide reliable data that reflects the actual situation (Arikunto, 2012).

CONCLUSION

The chemical literacy assessment instrument is categorised as valid (average validity score of 83.9), reviewed from the instrument's content, construct, and language usage. Meanwhile, the empirical validity analysis, the multiple-choice items, is observed to have a 0.37 to 0.77 correlation score, so that the items are classified as valid, with high reliability (0.86 reliability coefficient). Besides, five of the items have sufficiently good discriminating power, while 15 items have good discriminatory power. The item difficulty analysis results show that five, 14, and one item are classified as easy, moderate, and difficult, respectively. All of those items also have high functioning distractors. Additionally, the empirical validity test on the attitude questionnaire shows a 0.65-0.69 correlation score, classified as valid, with a sufficiently high reliability (0.59 reliability coefficient). Therefore, the developed instrument is valid and reliable so that it can measure students' chemical literacy.

REFERENCES

- American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy: a project 2061 report*. New York: Oxford University Press.
- Arikunto, S. (1993). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Bond, D. (1989). In Pursuit of Chemical Literacy: A Place for Chemical Reactions. *Journal of Chemical Education*, 66(2), 157.
- Celik, S. (2014). Chemical Literacy Levels of Science And Mathematics Teacher Candidates. *Australian Journal of Teacher Education*, 39(1), 1 – 15
- Cigdemoglu, C., & Geban, O. (2015). Improving Students' Chemical Literacy Level on Thermochemical and Thermodynamics Concepts through Context-Based Approach. *Chemistry Education Research and Practice*, 16, 302 – 317.
- Cigdemoglu, C., Arslan, H. O., & Cam, A. (2017). Argumentation to Foster Pre-Service Science Teachers' Knowledge, Competency, And Attitude on The Domains of Chemical Literacy of Acids And Bases. *Chemistry Education Research and Practice*, 18(2), 288 – 303.
- Direktorat Pembinaan SMA. (2017). *Panduan Penilaian oleh Pendidik dan Satuan Pendidikan Sekolah Menengah Atas*. Jakarta: Kementerian Pendidikan dan Kebudayaan RI.
- Kohen, Z., Herscovitz, O., & Dori, Y. J. (2020). How to Promote Chemical Literacy? Online Question Posing and Communicating With Scientists. *Chemistry Education Research and Practice*, 21(1), 250 – 266
- Mudiono, A. (2016). *Keprofesionalan Guru dalam Menghadapi Pendidikan di Era Global*. Makalah disajikan dalam Seminar Nasional, Jurusan KSDP FIP UM, Malang 25 September.
- Mumba, F., & Hunter, W. J. F. (2009). Representative Nature of Scientific Literacy Themes in A High School Chemistry Course: The Case of Zambia. *Chemistry Education Research and Practice*, 10(3), 219 – 226.
- Naganuma, S. (2017). An Assessment of Civic Scientific Literacy in Japan: Development of A More Authentic Assessment Task And Scoring Rubric. *International Journal of Science Education, Part B*, 7(4), 301 – 322

- Norris, S. P., & Philip, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87(2), 224 – 240.
- Organisation for Economic Co-operation and Development (OECD). (2016). *PISA 2015 Assessment And Analytical Framework: Science, Reading, Mathematic And Financial Literacy*. Paris: OECD Publishing
- Organisation for Economic Co-operation and Development (OECD). (2018). *PISA 2018 Result Combined Executive Summaries Volume I, II, & III*. Paris: Organisation for Economic Co-operation and Development.
- Osborne, J. F. (2010). Arguing to Learn in Science: The Role of Collaborative, Critical Discourse. *Science*, 328(5977), 463 – 466
- Rahayu, S. (2014). *Menuju Masyarakat Berliterasi Sains: Harapan dan Tantangan Kurikulum 2013*. Makalah disajikan dalam Seminar Nasional Kimia dan Pembelajarannya, Jurusan Kimia FMIPA UM, Malang 6 September.
- Rahayu, S. (2017). *Mengoptimalkan Aspek Literasi dalam Pembelajaran Kimia Abad 21*. Makalah disajikan dalam Seminar Nasional Kimia, Jurusan Pendidikan Kimia FMIPA UNY, Yogyakarta, 14 Oktober.
- Riduwan. (2011). *Belajar Mudah Penelitian: untuk Guru-Karyawan, dan Peneliti Pemula*. Bandung: Alfabeta
- Riduwan. (2013). *Dasar-Dasar Statistika*. Bandung: Alfabeta
- She, H. C., Stacey, K., & Schmidt, W. H. (2018). Science And Mathematics Literacy: PISA for Better School Education. *International Journal of Science and Mathematics Education*, 16(1), 1 – 5
- Shwartz, Y., Ben-Zvi, R., & Hofstein, A. (2005). The Importance of Involving High-School Chemistry Teachers in The Process of Defining the Operational Meaning of Chemical Literacy. *International Journal of Science Education*, 27(3), 323 – 344.
- Thummahong, R., & Thathong, K. (2016). Construction of A Chemical Literacy Test for Engineering Students. *Journal of Turkish Science Education*, 13(3), 185 – 198.
- United Nations Environment Programme (UNEP). (2012). *21 Issues for the 21st Century: Result of the UNEP Foresight Process on Emerging Environmental Issues*. Nairobi, Kenya: United Nations Environment Programme.
- Vogelzang, J., Admiraal, W. F., & van Driel, J. H. (2020). Effects of Scrum Methodology on Students' Critical Scientific Literacy: The Case of Green Chemistry. *Chemistry Education Research And Practice*, 21(3), 940 – 952.
- World Economic Forum (WEF). (2016). *New Vision for Education: Fostering Social And Emotional Learning through Technology*.