

DEVELOPMENT OF STUDENTS' SCIENCE LITERACY ASSESSMENT BASED ON MULTIPLE REPRESENTATIONS ON THERMOCHEMISTRY MATERIAL

Lu'lu'atul Munawaroh* and Ifah Silfianah

Program Studi Tadris Kimia, Fakultas Tarbiyah dan Ilmu Keguruan, Universitas Islam Negeri Sayyid

Ali Rahmatullah Tulungagung

e-mail: luluatul749@gmail.com

Abstract

One of the challenges in chemistry learning is students' low ability to solve science literacy problems based on multiple representations, particularly in thermochemistry, which is still dominated by symbolic evaluation. This condition indicates the need for more comprehensive evaluation instruments. Therefore, this study was conducted to develop a test instrument using the 4D model from Sivasailam Thiagarajan, modified to 3D (Define, Design, Development). The definition stage includes analyzing needs, students, tasks, concepts, scientific literacy, and multiple representations; the design stage involves determining the instrument form, grid, and question development; and the development stage consists of validation testing, empirical testing, and student profile testing. The research results show that the instrument is valid (90.9%), highly practical (76.66%), with a variety of difficulty levels and an average discrimination index of 0.30, which was categorized as fair. From 30 items, the validity test resulted in 18 valid items with an initial high reliability (0.815). However, students' science literacy profile is still low (41.15%) because they are not yet accustomed to tackling multiple representation-based literacy. It is recommended that teachers train students to work on science literacy problems based on multiple representations.

Keywords: instrument, science literacy, multiple representation, thermochemistry

INTRODUCTION

Education in Indonesia aims to develop students' academic potential while also shaping their personalities. One strategy implemented to achieve this goal is even tho the implementation of the Merdeka Curriculum, which provides students with the freedom to learn so they can develop their interests and talents according to their needs. As for the implementation of this curriculum, the emphasis is on literacy in various aspects, particularly on literacy that utilizes information, technology, and phenomena around us, such as science literacy [1].

Scientific literacy is becoming an important competency that students must possess to face the challenges of the 21st century. Even to this ability, students can be encouraged to think critically, connect scientific concepts with everyday phenomena, and participate in scientific issues in society. The importance of this ability is in stark

contrast to the results of the PISA analysis, which show that Indonesian students' science literacy achievement is still low. PISA 2018 data placed Indonesia at 70th out of 78 countries with a score of 396 [2], while in PISA 2022 the ranking rose to 69th out of 80 countries but the score decreased to 383 [3]. These results are still considered low when viewed against the international average score set by PISA at 500 [4]. This low achievement is caused by several factors, including students not being trained to solve science literacy questions due to a lack of developed science literacy question facilities [5], low student interest and reading habits [6], teachers' poor understanding of science literacy concepts leading to evaluation tools that are not sufficiently referenced [7], and teacher-centered teaching methods where students only memorize material concepts [8]. In this study, the development of instruments refers to three dimensions of science literacy, namely context, knowledge, and competence. The context

dimension includes personal, local/national, and global contexts; the knowledge dimension includes content, procedural, and epistemic knowledge; while the competence dimension includes the ability to explain phenomena scientifically, interpret data and evidence scientifically, and evaluate and design scientific investigations. These three dimensions serve as the basis for formulating measurable indicators and assessment items. Chemistry, particularly thermochemistry, is a suitable subject for the development of the instrument because it is closely related to everyday phenomena, such as the production and use of salt, the use of hand warmers, and photosynthesis. Thus, the instrument can provide students with the opportunity to understand concepts, analyze data, and apply scientific knowledge in problem-solving.

Thermochemistry is one of the branches of chemistry whose concepts are abstract and encompass three aspects of multiple representations. Among them are the first symbolic aspect, for example, in the process of salt formation, there is a release of heat to the environment amounting to 410.9 kJ/mol, symbolized by ΔH , which indicates the occurrence of a thermochemical reaction equation. Second, the macroscopic aspect, one of which is the use of salt in the process of preserving salted fish. Then finally, the sub-microscopic aspects such as salt which is formed from the elements sodium and chlorine gas which produces a solid compound sodium chloride [9]. In the learning process, thermochemistry material emphasizes problem-solving skills, but it often only uses symbolic representations, making it difficult for students to fully understand the concepts of thermochemistry [10]. Therefore, a learning medium or instrument that encompasses all three aspects of multiple representations is also needed to encourage students to fully understand thermochemistry concepts [11]. Additionally, the presence of these multiple representations can also improve students' scientific literacy skills [12].

Based on the interviews conducted with chemistry teacher at MAN 1 Blitar, it was found that thermochemistry is one of the topics taught to eleventh-grade students. Additionally, it shows that the learning process has not yet become

accustomed to applying scientific literacy skills and rarely integrates the three levels of representation simultaneously, leading to students' understanding being rote and students not being accustomed to solving scientific literacy questions based on multiple representations. This is due to the limited evaluation tools for learning, specifically scientific literacy instruments based on multiple representations. This needs analysis even to interviews indicates the need for a test instrument that measures student learning outcomes, including scientific literacy indicators, by integrating multiple representation aspects so that students' scientific literacy abilities can be measured more comprehensively.

Based on these conditions, the urgency of this research lies in the need for evaluation instruments that are valid, reliable, and aligned with the demands of the Merdeka Curriculum. This multiple representation-based science literacy test instrument on thermochemistry is expected to serve as a measurement tool and training resource that supports the improvement of students' science literacy. The objectives of this study are to develop a multiple representation-based science literacy test instrument on thermochemistry that meets validity and reliability criteria and can be used to comprehensively measure students' learning outcomes in the cognitive aspect. Additionally, this study also aims to measure the science literacy skills profile of 11th-grade students at MAN 1 Blitar, which will be analyzed based on the results of the developed test instrument.

METHOD

This research uses the Research and Development (R&D) method with the 4D model (Define, Design, Develop, Disseminate) from S. Thiagarajan, but was only carried out up to the 3D stage (Define, Design, Development) due to time constraints. According to Sugiyono, Research and Development (R&D) is a research method used to produce a specific product and test its effectiveness [13]. The selection of this model by S. Thiagarajan is due to its relatively simple stages, which do not require a long time [14].

The study population consisted of all science students at MAN 1 Blitar, with the research sample

being one class of XI IPA A for the limited trial (empirical test) and two classes of XI IPA, namely XI IPA B and XI IPA C, for the science literacy ability test based on multiple representations. This sample determination uses the purposive sampling technique, which means the selected sample consists of students who have received thermochemistry material in class.

In data collection, this research uses interview techniques with chemistry teachers and students as a needs analysis (preliminary study) to determine student characteristics and development, expert validation questionnaires by lecturers and chemistry teacher to assess instrument feasibility, student response questionnaires to determine the practicality of the developed instrument, and a science literacy ability test in the form of multiple-representation-based essay questions on thermochemistry material for analyzing the characteristics of the developed questions and analyzing students' science literacy ability profiles on thermochemistry material. The test items are developed based on science literacy indicators that encompass aspects of context, knowledge, and competence. Each item integrates macroscopic, submicroscopic, and symbolic representations thru the presentation of phenomena, molecular or particle representations, as well as chemical equations or symbols. The final exam items are developed at the C3 and C4 cognitive levels. The relationship between science literacy competencies, multiple representation modes, cognitive levels, and the corresponding item numbers is summarized in the construct mapping presented in Table 1.

Table 1. Construct Mapping of the Science Literacy Assessment Instrument

Science Literacy Competence	Multiple Representation	Cognitive Level & Item Numbers
Explaining phenomena scientifically	Macroscopic, submicroscopic, and symbolic	C3: 7, 26; C4: 1, 3, 6, 9, 12, 15, 21, 24, 27
Interpreting data and evidence scientifically	Macroscopic, submicroscopic, and symbolic	C3: 2, 4, 5, 8, 10, 13, 14, 16, 17, 18, 22, 23, 25, 28, 30;

Science Literacy Competence	Multiple Representation	Cognitive Level & Item Numbers
Evaluating and designing scientific inquiry	Macroscopic, submicroscopic, and symbolic	C4: 19; C3: 11; C4: 20, 29

The mapping shows the alignment between science literacy competencies, multiple representations, cognitive levels, and corresponding test items.

The collected data was then analyzed in several stages, including the interview results, which were qualitatively analyzed to describe development needs. Meanwhile, the analysis of expert validation results is calculated using the percentage of Likert scale scores as per the following formula:

$$NP = \frac{\sum x}{\sum xi} \times 100\%$$

Explanation:

NP = Percentage value

$\sum x$ = Sum of validator responses

$\sum xi$ = Sum of total scores

After obtaining the percentage results, they were then categorized based on the following validation criteria to determine the instrument's suitability and will be used as a reference for improving the research instrument in Table 2 [15].

Table 2. Eligibility Criteria

Interval Skor (%)	Kriteria
0 – 20	Very unsuitable
21 – 40	Unsuitable
41 – 60	Moderately suitable
61 – 80	Suitable
81 – 100	Very suitable

The analysis of student response questionnaire results was conducted quantitatively based on calculating the percentage of the total Likert scale score to assess practicality I Table 3. The scores for each answer option and their percentage calculation formula are as follows.

Table 3. Likert Scale

Answer Choices	Score
Strongly Agree	5
Agree	4
Somewhat Agree	3
Disagree	2
Strongly disagree	1

$$NP = \frac{\sum x}{\sum xi} \times 100\%$$

Explanation:

NP = Percentage value

$\sum x$ = Sum of validator responses

$\sum xi$ = Sum of total scores

After obtaining the percentage results, they were then categorized based on the following practicality percentage criteria in Table 4 [15].

Table 4. Percentage of Practicality of Questions

Interval Skor (%)	Criteria
0 – 20	Very unsuitable
21 – 40	Unsuitable
41 – 60	Moderately suitable
61 – 80	Suitable
81 – 100	Very suitable

The results of the instrument trial (empirical test) are analyzed to determine the characteristics of the questions, which include validity, reliability (Cronbach's Alpha), difficulty level, and discrimination power. Before conducting empirical analysis, students' answers to each essay question were scored by the researcher as the sole evaluator using the scoring guidelines that had been prepared for each item. The scoring guidelines contain score criteria for each step or component of the answer based on the accuracy and completeness of the students' responses. Student responses that are not entirely correct are still scored according to the criteria set in the rubric, so the awarding of partial credit is not based on the subjective judgment of the evaluator. For example, in a step with a maximum score of 3, an answer that meets the criteria receives a score of 3, the use of the correct formula but resulting in an incorrect final calculation receives a score of 2, an incorrect answer that still shows effort receives a score of 1, while an unanswered question receives a score of 0. The scores for each step or component are then summed to obtain the total score for each item. The

total score obtained thru rubric-based scoring is subsequently used in the analysis of validity, reliability, difficulty level, and discriminative power. Because all student answers were scored by a single rater, namely the researcher, inter-rater agreement was not calculated. The reliability of the instrument was analyzed using Cronbach's Alpha during the empirical testing phase. The validity and reliability tests were conducted using the SPSS 26 application. The validity criteria of the questions can be reviewed based on the r_{table} and $r_{calculated}$ criteria or by comparing them with Sig. (2-tailed), with a probability of 0.05, as follows in Table 5 and Table 6 [13].

Table 5. Criteria for r_{table} and r_{count}

Rxy	Significance Level	Description
$r_{calculated} > r_{table}$	< 0.05	Valid
$r_{calculated} < r_{table}$	> 0.05	Not valid

Explanation:

Rxy = Correlation coefficient with the correlated variable

Table 6. Criteria for Sig. 2 (2-tailed) Values

Description	Criteria
Sig. (tailed) < 0.05 (+)	Valid
Sig. (2-tailed) < 0.05 (-)	Not Valid
Sig. (2-tailed) > 0.05	Not Valid

After analyzing the results. Then, interpreting the correlation coefficient according to the following criteria in Table 7 [16].

Table 7. Correlation Criteria

Internal Coefficient	Relationship Level
0.80 – 1.00	Very Strong
0.60 – 0.79	Strong
0.40 – 0.59	Moderate
0.20 – 0.39	Low
0.00 – 0.19	Very Low

As for the results of the reliability test analysis using the Cronbach Alpha technique in SPSS 26, they were then categorized according to the reliability index criteria as follows in Table 8 [17].

Table 8. Grouping of Reliability Coefficient Criteria

Interval	Criteria
0.80 – 1.00	Very High
0.60 – 0.80	High
0.40 – 0.60	Sufficient
0.20 – 0.40	Low
0.00 – 0.20	Very Low

The difficulty level of the test items was analyzed by calculating the percentage of student scores, and then the results were categorized according to the item difficulty index in Table 9 [18].

$$\text{Difficulty level} = \frac{\text{Average student score}}{\text{Maximum score for then item}}$$

Table 9. Item Difficulty Index

Interval (%)	Category
0.00 – 0.30	Difficult
0.30 – 0.70	Moderate
0.70 – 1.00	Easy

Meanwhile, the item discrimination power was analyzed by calculating the percentage of student scores, and the results were categorized according to the discrimination index classification in Table 10 [19].

$$DB = \frac{\Sigma TB}{\Sigma T} - \frac{\Sigma RB}{\Sigma R}$$

Explanation:

DB = item discrimination index

ΣTB = number of high-achieving students who answered correctly

ΣT = number of high-achieving students

ΣRB = number of low-achieving students who answered correctly

ΣR = number of low-achieving students

Tabel 10. Eligibility Criteria

Discrimination Index Number	Classification	Interpretation
Negative	-	Very poor
0.00 – 0.19	Poor	Poor
0.20 – 0.39	Satisfactory	Fair
0.40 – 0.69	Good	Good
0.70 – 1.00	Excellent	Very good

After conducting an empirical test analysis, a quantitative analysis was then performed by calculating the results of student answers based on

the scoring guidelines that had been created. The calculation is as follows:

$$\bar{x} = \left(\frac{\Sigma x}{n} \right) \times 100\%$$

Explanation:

$\bar{x}$ = Average test score

Σx = Sum of scores for all subjects

n = Number of subjects

After obtaining the percentage results, the students' science literacy proficiency levels were then categorized as follows in Table 11 [20].

Table 11. Science Literacy Categories

Value Interval (%)	Criteria
86 – 100	Very high
71 – 85	High
56 – 70	Sufficient
41 – 55	Low
≤ 40	Very low

RESULTS AND DISCUSSION

Based on the research conducted, the following results were obtained.

Process of Developing a Science Literacy Ability Test Instrument Based on Multiple Representations.

The development of a science literacy ability test instrument based on multiple representations for thermochemistry material was carried out using the 4D model, limited to 3 stages: define, design, and develop.

Define Stage

At this stage, a needs analysis was conducted even to interviews, which showed that chemistry teacher at MAN 1 Blitar believe that science literacy skills are very important, especially in training students to think critically in the era of the Merdeka Curriculum. However, although the chemistry teacher has begun to prepare simple practice test items, the still dominant written test-based evaluation has not integrated scientific literacy or multiple representations. Therefore, the teacher supports the development of this multiple representation-based scientific literacy test instrument, particularly for thermochemistry material. This is reinforced by the results of student interviews, which indicate that students are not yet accustomed to solving science literacy questions,

especially those based on multiple representations that are limited to symbolic representation only.

Next, a task analysis was conducted, which involved determining the learning outcomes and learning objectives for thermochemistry material, aligned with the Merdeka Curriculum at MAN 1 Blitar, Phase F. Subsequently, a concept analysis was performed by creating a concept map for thermochemistry material, as well as analyzing multiple representations and scientific literacy in thermochemistry material. An analysis of multiple representations was conducted following an analysis of phenomena related to thermochemistry topics that lend themselves to such representations. Subsequently, the results of this analysis were contextualized and linked to thermochemistry concepts in everyday life, taking into account scientific literacy aspects specifically context, knowledge, and competence though the scientific identity aspect was excluded as the study utilized the PISA 2022 framework.

Design Phase

In this phase, the product design process is carried out, including the framework for the test instrument, with several steps such as determining the instrument's form, which is descriptive. Then, learning objectives and test item indicators were designed, considering science literacy indicators such as context, knowledge, and competence aspects. A test instrument outline for science literacy ability based on multiple representations for thermochemistry material was also developed, taking into account question levels ranging from C3 to C6. During the instrument design phase, the interview guide sheet was adapted to the researcher's needs for needs analysis, item creation, and scoring guidelines aligned with learning objectives and science literacy indicators. Expert validation questionnaires and student response questionnaires were also created to measure the feasibility of the item.

Figure 1 is a multiple-representation-based scientific literacy question on the topic of thermochemistry.

Bacaan untuk nomor 12-14



(Gambar 9. Proses pembuatan ikan asin)

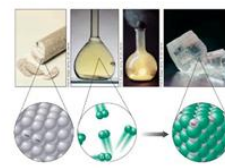
<https://images.app.goo.gl/P1PWrmz8KGohzcpw8>

Di kehidupan sehari-hari khususnya di tempat belanja lokal tidak asing dengan yang namanya ikan asin. Bagaimana tingkat ketahanan penyimpanan ikan asin tersebut? Ikan asin memiliki tingkat keawetan yang lebih lama daripada ikan segar. Hal ini dikarenakan ikan asin diberikan garam 10% dari berat daging ikan sebagai bahan pengawet dan juga dikeringkan sehingga kandungan air yang semakin rendah dapat mencegah pembusukan pada ikan. Garam yang menyerap air dari dalam tubuh ikan melalui proses osmosis, sehingga kandungan air yang menjadi media hidup bakteri berkurang. Garam yang digunakan merupakan salah satu bahan kimia yang memiliki rumus senyawa NaCl yang tersusun dari logam natrium yang bereaksi dengan gas klorin. Ion logam natrium (yang sangat lunak sehingga dapat dipotong dengan pisau) bereaksi dengan ion gas klorin yang mengandung molekul Cl_2 untuk membentuk padatan natrium klorida yang melepaskan kalor ke lingkungan sebesar $410,9 \text{ kJ/mol}$. Selain sebagai pengawet ikan asin, garam juga dimanfaatkan sebagai bahan tambahan pengolahan makanan, sebagai penyerap dan pengangkut nutrisi, menjaga keseimbangan cairan dalam tubuh. Namun apabila mengonsumsi garam yang terlalu sering juga tidak disarankan, karena dapat memicu tekanan darah tinggi, gagal ginjal, dan kanker.

12. Berdasarkan informasi yang disajikan, jenis entalpi apakah yang terjadi?
Jelaskan!

Jawab:.....

13. Perhatikan gambar submikroskopis di bawah ini!



Keterangan: NaCl = Na = Cl_2
Gambar 10. Mikroskopis pembentukan garam (NaCl)
(Zumdahl & Zumdahl, 2014)

Apabila suatu garam tersusun seperti pada gambar, bagaimana persamaan termokimia yang terjadi?

Jawab:.....

Figure 1. Multiple-Representation-Based Scientific Literacy Question

Development Stage

The development stage is the final stage carried out in this research. At this stage, the instruments that have been designed and developed are then validated by two experts, namely a chemistry education lecturer and a chemistry teacher from MAN 1 Blitar. The results of the content validity test on 30 items showed an average feasibility level of 90.9%, which falls into the very feasible category. Improvements were made primarily to the language editing, the clarity of the submicroscopic representations, and the addition of reaction equations to strengthen the connections between the representations. Among the improvements are the following Table 12.

Table 12. Comparison of Instruments Before and After Revision

No.	Before Revision	After Revision
1.	If the water in the dam is evaporated, it will require 241.5 kJ of heat from the environment...	If the water in the dam is evaporated, it will require 241.5 kJ of heat from the environment...
2.	 Figure of carbon dioxide combustion (Karen C and Timberlake)	 Description: $\text{C} = \text{C}$ $\text{O}_2 = \text{O}_2$ $\text{CO}_2 = \text{CO}_2$ Figure of carbon dioxide gas formation (Timberlake, 2016)

After the repairs were made, a product development test (empirical test) was conducted involving class XI A with 34 students to obtain data on the characteristics of the developed questions. This empirical test is a trial in a limited class during the instrument development stage. Additionally, an analysis of the practicality of the questions was conducted after the test instrument was administered, where students provided responses, comments, and suggestions on the prepared student response questionnaire. The results are as shown in Figure 2.

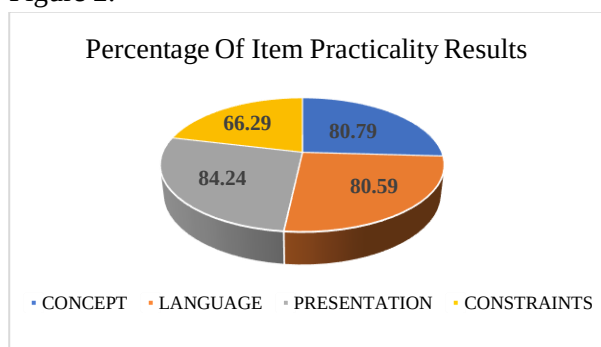


Figure 2. Percentage of Item Practicality Test Results

Based on the image, the average response questionnaire results for practicality testing were 76.66% with a high category, with excellent concepts, language, and presentation, with average

percentages of 80.79%, 80.59%, and 84.24% respectively. However, the percentage of obstacles experienced by students while working was still considered high at 66.92%. This was because students had never worked on science literacy questions based on multiple representations, and there were time constraints while working. Additionally, this is also caused by students' low reading comprehension, making it difficult for them to find information presented in literature. This is consistent with the research by [21], students' low literacy skills are a constraint due to students' lack of interest in reading. However, presenting questions in multiple representations can help in answering the given test items. This is consistent with Astuti's research in [22], which states that using materials with multiple representations is effective in developing students' conceptual understanding and problem-solving skills. Student responses indicate reactions to the presentation of science literacy-based questions and multiple representations showed in Figure 3.

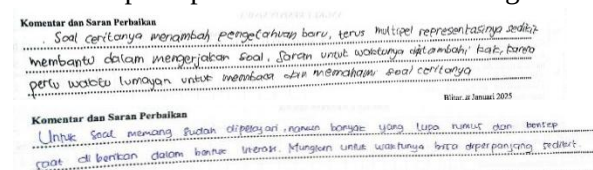


Figure 3. Student Comments and Suggestions

Based on Figure 3, student responses indicate reactions to the presentation of science literacy-based questions and multiple representations. From the aspect of science literacy, one student stated that questions in the form of stories can add new knowledge, while another student mentioned that the material tested had actually been studied, but they still had difficulty recalling formulas and concepts when presented in the form of literacy questions. That response aligns with the characteristics of scientific literacy, which not only requires students to remember scientific knowledge but also to use the knowledge and concepts they have learned to understand and solve problems in specific contexts. Thus, presenting questions in the form of stories provides a problem context that allows the chemistry knowledge that has been learned to be used in a more contextual situation.

From the aspect of multiple representations, students stated that the use of multiple representations was somewhat helpful in solving problems. In chemistry learning, multiple representations are used to present a concept thru macroscopic, submicroscopic, and symbolic representations so that information about a phenomenon can be understood from complementary perspectives. The students' responses indicate that the presence of multiple forms of representation can provide additional information in the process of understanding and solving problems. This is in line with Ainsworth's functional theory, which states that multiple representations have a complementary function, meaning that each form of representation can provide information that is not fully conveyed by other representations [23]. Thus, the combination of several representations can help students obtain more complete information in understanding a concept or problem.

In line with this, use of multiple representations serves to obtain additional information, support existing cognitive processes so that they can complement each other in understanding problem situations, thereby enabling the interpretation of information and problem-solving, as well as building a deeper understanding [24]. However, the students' responses indicating that multiple representations were "somewhat helpful" also suggest that the use of several representations may not be easily understood by all students. Additionally, both students suggested extending the time for completion because it takes time to read and understand the story problems. This shows that questions combining the context of science literacy and multiple representations require a process of reading, understanding information, and connecting the presented information before students determine their answers. Thus, the students' responses in Figure 4 not only indicate acceptance of the use of science literacy and multiple representations but also provide an overview of the students' experiences when faced with questions that demand the use of knowledge in context and understanding of the information presented thru various representations.

Analysis of Test Item Characteristics

After conducting empirical tests, the characteristics of the test items were produced, including the difficulty level of the questions, the discrimination power of the questions, the validity, and the reliability of the questions. The analysis of the test items shows a proportional distribution of difficulty levels, namely 8 difficult items, 15 medium items, and 7 easy items, with an average of 0.46 in the medium category. This is in line with the research by Maya & Titin, which found that the difficulty level ranges from 0.31 to 0.68, with a moderate criterion [25].

The results of the item discrimination power analysis using the 27% rule also indicate that the ability of the test items to differentiate between high and low ability students still varies. A total of 12 questions are classified as good, 6 questions as fair, 11 questions as poor, and 1 question as very poor, with an average discrimination index of 0.30, which falls into the fair category. Based on the analysis conducted, the large number of items classified as poor indicates that these items can be answered correctly by low-ability students. Meanwhile, items with good criteria show a better ability to differentiate between the two groups, especially when more items can be answered correctly by high-ability students compared to low-ability students. Based on these results, the overall quality of the instrument's discriminative power falls into the sufficient category, and there are still several items that require improvement or should not be retained in the final instrument.

The results of the validity test analysis using SPSS 26 at a significance level of 5% or $\alpha = 0.05$ with a sample (n) of 34 students yielded an r-table value of 0.339, indicating that 18 items were valid, namely items 1, 2, 5, 6, 7, 8, 11, 13, 14, 17, 18, 19, 22, 23, 24, 26, 27, and 28. The number of invalid items was 12, namely items 3, 4, 9, 10, 12, 15, 16, 20, 21, 25, 29, and 30. The presence of 12 invalid items is related to the low discrimination power of some items. Where the level of item validity is related to the item discrimination index, meaning that the higher the discrimination index, the better the item validity, while the difficulty level does not affect the item validity. This is in accordance with the research by Nurhalimah et al. regarding the

relationship between validity, discrimination index, and item difficulty, which shows that item validity and discrimination index have a direct correlation, while the difficulty index does not show a correlation with validity, thus the difficulty level does not affect item validity [26]. Therefore, even if the items are categorized as difficult or easy, they can still be considered valid if they have a sufficient to good level of discrimination. In this study, there are 11 items categorized as poor and 1 item categorized as very poor in the discrimination analysis, as well as 12 invalid items. The pattern is consistent with Maharani's research, which, in the development of the science literacy test instrument, found 9 invalid items and 9 items categorized as poor in the discrimination analysis [27]. The similarity of these findings indicates that in the development of science literacy instruments, empirical test results can reveal items that do not meet validity criteria along with low discrimination power. This condition could explain the discovery of 12 invalid items in this study. The findings also show that some items in the initial instrument do not yet have adequate empirical quality. Therefore, the results of the empirical test serve as the basis for selecting the initial instrument, namely by not using the 12 invalid items in the medium-scale test and retaining the 18 valid items for further testing. The 18 valid questions were tested on a moderate scale to determine the science literacy skills based on multiple representations of 11th-grade students at MAN 1 Blitar on the topic of thermochemistry, consisting of 4 easy questions, 10 moderate questions, and 4 difficult questions.

After conducting validity analysis, 18 items were identified as valid. Consequently, a reliability analysis using SPSS 26 was conducted on the 18 items, and a Cronbach's Alpha value of 0.815 was obtained, which falls into the very high category according to the criteria in Table 8. Thus, the final instrument consisting of 18 items has very high reliability, indicating good internal consistency in measuring the same construct showed in Table 13.

Reliability Statistics	
Cronbach's Alpha	N of Items
.815	18

The high reliability test results are consistent with the large number of valid questions. The more valid test items there are, the higher the instrument's reliability level, and conversely, if the test items are low in validity, the reliability level is also low [28].

Therefore, the relatively limited number of respondents is one of the limitations of this study in estimating the characteristics of the 30 items, so the results of the item characteristic analysis need to be interpreted with caution. Nevertheless, this empirical test provides initial information used to select items that meet the criteria before the instrument is applied to a larger group of students. Subsequent research is recommended to involve a larger number of respondents at the empirical testing stage to obtain more stable item characteristic estimates.

Thus, overall, the results of this study prove that the multiple representation-based science literacy test instrument developed for thermochemistry material is valid, reliable, practical, and has good item quality, making it suitable for use in medium-scale testing to determine students' science literacy levels in thermochemistry, and also for wider use in mapping students' science literacy abilities and as a more meaningful alternative for evaluating chemistry learning.

Testing the Science Literacy Ability Profile Based on Multiple Representations

The questions that were declared valid and reliable were then pilot tested on 68 students from classes XI IPA B and X IPA C with a completion time of 90 minutes. The purpose of this pilot test was to determine the science literacy ability based on multiple representations of class XI students at MAN 1 Blitar on the topic of thermochemistry showed in Figure 4.

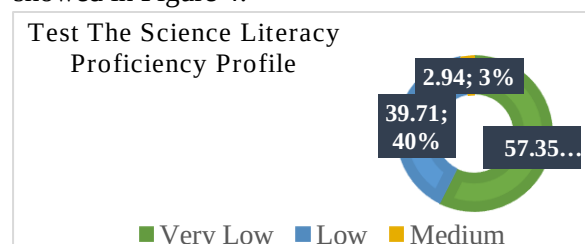


Figure 4. Results of the Multiple Representation-Based Science Literacy Ability Profile Test

The analysis of student responses on the medium-scale test results showed that the science literacy ability based on multiple representations of class XI students at MAN 1 Blitar on thermochemistry material was still low, with an average percentage of 41.15%. This low result can be seen in Figure 4, where out of 68 students, 39 had very low ability, 27 had low ability, and only 2 had moderate ability. The results of this study are consistent with Bungawati's research, which found that science literacy skills have low criteria with a student score of 56% in the low category [29]. Additionally, this aligns with Malik's research, which showed that students' science literacy skills were in the low category with a score of 43% [30].

The condition related to the low science literacy skills at MAN 1 Blitar, as conveyed by the chemistry teacher during the interview, is that during the learning process, the teacher has not been optimal in applying science literacy and multiple representations due to time constraints, even tho they have prepared a science literacy question design as an initial step to introduce science literacy to the students. In the students' responses, it was found that the students are not yet accustomed to solving science literacy and multiple representation questions. The research results of Zahro and Fauziah explain that the students' lack of familiarity with solving problems using scientific skills is one of the causes of the low science literacy skills among students [31].

In addition, the results of student comments or responses in the questionnaire also indicate that some students have difficulty recalling the chemistry formulas they have memorized and studied, and have a low interest in re-reading the material. This condition is related to the students' difficulty in solving the given problems. Sutrisna's research results explain that the low science literacy skills of students are influenced by several factors, namely the still low interest in reading and evaluation tools that do not yet focus on the development of science literacy [4]. Meanwhile, the research results of Aswita et al. also explain that students' difficulties in solving science literacy questions are caused by weak analytical skills, formula application, precision, understanding of algorithmic information, as well as low memory

retention and reading interest among students [32]. Both findings are consistent with the survey results in this study, particularly regarding the difficulty in recalling formulas and the low interest in reading, which serve as the basis for interpreting students' science literacy skills.

Here are the results of students with moderate science literacy skills, as shown in Figure 5.

5. Berdasarkan peristiwa tersebut, jika terjadinya pembakaran karbohidrat dapat menghasilkan kalor dalam tubuh yang diamati menggunakan termometer, di mana suhu tubuh Axel setelah berolahraga meningkat dari 34,2°C menjadi 37,3°C, berapa gram karbohidrat yang harus dikonsumsi oleh Axel untuk menjaga pola makan yang baik? (3)

Jawab: Diket: $m_{\text{karbohidrat}} = 89 \text{ kg} \rightarrow 89000 \text{ g}$
 Kalor $q_{\text{tubuh}} = 67 \text{ kJ}$
 $\Delta T = 37,3 - 34,2 = 3,1^\circ\text{C}$
 $q = m \times c \times \Delta T$
 $67000 = 89000 \times c \times 3,1$
 $c = \frac{67000}{89000 \times 3,1} = 1,655,4 \text{ J/g}^\circ\text{C}$
 $m_{\text{karbohidrat}} = \frac{q}{c \times \Delta T} = \frac{67000}{1,655,4 \times 3,1} = 109,4 \text{ g}$

Figure 5. Student Responses of Moderate Ability to Question Number 5

Based on Figure 5, it shows that students were able to solve question number 5 correctly even to the stages of data analysis and interpretation of the information presented, application of the heat formula to determine the heat value, and accurate calculation of carbohydrate mass. The results of this answer indicate that students can correctly interpret scientific data and evidence, placing their science literacy skills in the moderate category. Figure 6 shows that students with low abilities were able to solve question number 14, but their answers were not entirely accurate. This is because students are not careful enough when analyzing the information presented, resulting in inaccurate data for the mass values they write down. Consequently, their final answers are incorrect due to errors in the magnitude of the mass value, even tho the formula they wrote down is correct. Thus, the criteria for students' literacy skills are still considered low because students are not yet accurate in interpreting data and evidence scientifically. the low ability student responses to several question number showed in Figure 6 and Figure 7.

14. Perhatikan gambar submikroskopis berikut!

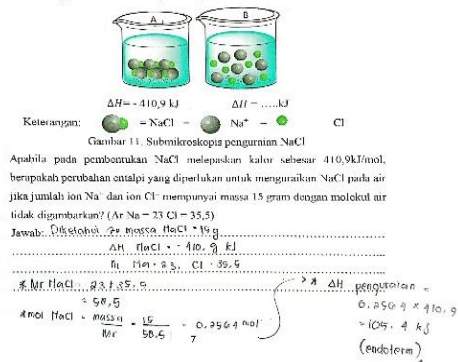


Figure 6. Low-Ability Student Responses to Question Number 14

6. Berdasarkan informasi yang disajikan, jenis entalpi apakah yang terjadi? Jelaskan!

Jawab: Karena kalor di lepaskan ke lingkungan (65.8 kJ). Jenis entalpi yang terjadi adalah entalpi negatif ($\Delta H < 0$)

Figure 7. Low-Ability Student Responses to Question Number 6

Figure 7 shows that students have not yet been able to answer question number 6 correctly. This is because students are not accurate in understanding the concept between the type of enthalpy and the type of reaction in thermochemistry, leading them to answer exothermic with the indicator of heat release and a negative ΔH for the dissolution of limestone. However, the correct answer is the enthalpy of dissolution (ΔH_{sol}). Thus, students' science literacy skills on this question are very low. Thus, it can be concluded overall that most eleventh-grade students at MAN 1 Blitar still have low science literacy skills based on multiple representations in thermochemistry material, making it highly necessary to improve students' science literacy skills even to science literacy-based learning and evaluation processes that incorporate all three aspects of multiple representations and are linked to daily life. In addition, it is also necessary to implement the Inquiry Based Learning model integrated with the Nature of Science (NOS) as an implicative effort to train students in connecting various representations and improving their science literacy skills.

CONCLUSION

Based on the description presented, it can be concluded that this research produced a highly feasible (expert validation 90.9%) and highly

practical (76.66%) multiple representation-based science literacy ability test instrument for thermochemistry material. The results of the item characteristic analysis showed that the developed items had an average percentage with moderate difficulty level (0.46), sufficient discrimination power (0.30), and validity testing that resulted in 18 out of 30 items being valid with high reliability (0.815) from 18 items. The test results show that the science literacy of 11th-grade students at MAN Blitar in thermochemistry material, based on multiple representations, is still low (41.15%). This research has reached the stage of analyzing science literacy skills, so further research is needed to analyze the factors affecting the low science literacy of students. Additionally, the influence of learning models or learning media on science literacy skills can also be examined.

REFERENCES

- Muliaman, A., Sakdiah, H., and Ginting, F. W. 2022. Analisis Employability Skill dan Literasi Sains Siswa melalui Authentic Self-Assessment pada Kurikulum Merdeka di SMA Aceh Utara. *JPF (Jurnal Pendidik. Fis. Univ. Islam Negeri Alauddin Makassar)*, Vol. 11, No. 1, pp. 24–32.
- Nurfadillah, T., Elvia, R., and Elvinawati, 2023. Pengembangan Instrumen Tes Kimia Berbasis Literasi Sains untuk Mengukur Literasi Sains Siswa. *Alotrop*, Vol. 7, No. 1, pp. 44–56.
- Efendi, S. 2024. Pengembangan Instrumen Literasi Sains pada Materi IPA SMP/MTS Kelas VIII. Riau: Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Sutrisna, N. 2021. Analisis Kemampuan Literasi Sains Peserta Didik SMA di Kota Sungai Penuh. *J. Inov. Penelit.*, Vol. 1, No. 12, pp. 2683–2694.
- Djarmiko, R., and Mahbubah, K. 2022. Identifikasi Keterampilan Literasi Sains Siswa Pada Materi Asam Basa. *SCIENING Sci. Learn. J.*, Vol. 3, No. 1, pp. 59–63.
- Fuadi, H., Robbia, A. Z., Jamaluddin, J., and Jufri, A. W. 2020. Analisis Faktor Penyebab Rendahnya Kemampuan Literasi Sains Peserta Didik. *J. Ilm. Profesi Pendidik.*, Vol. 5, No. 2, pp. 108–116.
- Juwita, E., and Rosidin, U. 2022. Analisis Kemampuan Literasi Sains Siswa Kelas IX

- MTs Negeri 1 Lampung Barat pada Materi Bioteknologi Berbasis Etnosains. *J. Edukasi Mat. dan Sains*, Vol. 10, No. 2, pp. 232–242.
8. Yusmar, F., and Fadilah, R. E. 2023. Analisis Rendahnya Literasi Sains Peserta Didik Indonesia: Hasil Pisa dan Faktor Penyebab. *LENSA (Lentera Sains) J. Pendidik. IPA*, Vol. 13, No. 1, pp. 11–19.
 9. Pikoli, M., Sukertini, K., and Isa, I. 2022. Analisis Model Mental Siswa dalam Mentransformasikan Konsep Laju Reaksi melalui Multipel Representasi. *Jumbara J. Educ. Chem.*, Vol. 4, No. 1, pp. 8–12.
 10. Sariati, Kadek, N., Suardana, Nyoman, I., Wiratini, and Made, N. 2020. Analisis Kesulitan Belajar Kimia Siswa Kelas XI pada Materi Larutan Penyangga. *J. Imiah Pendidik. dan Pembelajaran p-ISSN 1858-4543 e-ISSN 2615-6091*, Vol. 4, No. 1, pp. 86–97.
 11. Irfandi, Murwindra, R., and Musdansi, D. P. 2022. Analisis Penyebab Miskonsepsi Siswa pada Materi Termokimia di SMAN 1 Teluk Kuantan. *J. Pendidik. dan Konseling*, Vol. 4, No. 6, pp. 7809–7813.
 12. Yuliana, I. F., and Sholichah, N. 2021. Pengembangan Modul Termokimia Berbasis Multi Representasi untuk Melatih Literasi Kimia Mahasiswa pada Materi Termokimia. *Chem. Educ. Pract.*, Vol. 4, No. 2, pp. 179–185.
 13. Sugiyono. 2022. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
 14. Maydiantoro, A. 2021. Model-Model Penelitian Pengembangan (Research and Development). *Respository LPPM Unila*, No. 10, pp. 1–8.
 15. Arikunto, S. 2010. *Manajemen Penelitian*. Jakarta: Rineka Cipta.
 16. Sugiyono. 2016. *Statistika untuk Penelitian*. Bandung: Alfabeta.
 17. Arikunto. 2002. *Metodologi Penelitian Suatu Pendekatan Proposal*. Jakarta: PT Rineka Cipta.
 18. Iswan. 2018. *Manajemen Pendidikan Kewirausahaan*, 1st ed. Depok: PT Raja Grafindo Persada.
 19. Arikunto, S. 2016. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
 20. Arikunto, S. 2006. *Prosedur Penelitian Suatu Pendekatan Praktek*. Jakarta: PT. Rineka Cipta.
 21. Ramdhani, M., Wahidah, B., and Wahyuni, W. 2022. Problematika Budaya Literasi Membaca di SMAN 1 Aikmel. *J. Bastrindo*, Vol. 3, No. 2, pp. 137–145.
 22. Anggraini, M., Huda, I., Safrida, S., Rahmatan, H., and Djufri, D. 2022. Penerapan Model Multiple Representation Berbasis Somatis, Auditory, Visual, Intelektual terhadap Self Efficacy Peserta Didik. *J. Pendidik. Sains Indones.*, Vol. 10, No. 1, pp. 11–20.
 23. Ainsworth, S. 2006. DeFT: A Conceptual Framework for Considering Learning with Multiple Representations. *Learn. Instr.*, Vol. 16, pp. 183–198.
 24. Distrik, W., Supardi, Z. A. I., Jatmiko, B., and Yuberti. 2021. The Effects of Multiple Representations-Based Learning in Improving Concept Understanding and Problem-Solving Ability. *J. Phys. Conf. Ser.*, Vol. 1796, No. 1, pp. 1–10.
 25. Ningsetyo, M., and Sunarti, T. 2024. Pengembangan Instrumen Tes Literasi Sains Berbasis Kearifan Lokal di Probolinggo. *J. Arjuna Publ. Ilmu Pendidikan, Bhs. dan Mat.*, Vol. 2, No. 2, pp. 59–68.
 26. Nurhalimah, S., Hidayati, Y., Rosidi, I., and Hadi, W. P. 2022. Hubungan antara Validitas Item Dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda Pas. *Nat. Sci. Educ. Res.*, Vol. 4, No. 3, pp. 249–257.
 27. Maharani, I. 2021. Pengembangan Instrumen Tes Berbasis Literasi Sains Berbantuan Google Form pada Materi Hukum Newton. Universitas Negeri Medan.
 28. Ardhani, Y. 2020. Kualitas Butir Soal Penilaian Akhir Tahun Mata Pelajaran Teknologi Dasar Otomotif Kelas X Teknik Kendaraan Ringan Otomotif di SMK Muhammadiyah Gamping. *J. Pendidik. Vokasi Otomotif*, Vol. 3, No. 1, pp. 85–94.
 29. Bungawati, B. 2024. Profil Kemampuan Literasi Sains Siswa pada Materi Ekosistem. *J. Binomial*, Vol. 7, No. 1, pp. 43–50.
 30. Listiani, H., Satar, S., Nurbaya, and Akobiarek, M. 2024. Profil Kemampuan Literasi Sains (Scientific Literacy) Siswa Kelas X SMA YPK Diaspora Kotaraja, Papua. *Pendas: Jurnal*

- Ilmiah Pendidikan Dasar*, Vol. 09, No. 04, pp. 767–778.
31. Zahro, F., and Fauziah, A. N. M. 2025. Analisis Kemampuan Literasi Sains IPA pada Siswa SMP dalam Memecahkan Soal IPA. *Pensa E-Jurnal Pendidik. Sains*, Vol. 13, No. 1, pp. 30–34.
32. Aswita, Rusman, and Rahmayani, R. F. I. 2017. Identifikasi Kesulitan Siswa dalam Memahami Materi Termokimia dengan Menggunakan Three-Tier Multiple Choice Diagnostic Instrument di Kelas XI MIA 5 MAN MODEL Banda Aceh. *J. Ilm. Mhs. Pendidik. Kim.*, Vol. 2, No. 1, pp. 35–44.