Jurnal Pijar MIPA
https://jurnalfkip.unram.ac.id/index.php/JPM
ISSN 1907-1744 (Print)
ISSN 2460-1500 (Online)

Development of Science Literacy Test Instruments on Temperature and Heat Materials by Integrating Local Wisdom in the Production of *Karang Batik*, a Speciality of Tuban Regency

Novy' Ainul Muizzah*, Mukhayyarotin Niswati Rodliyatul Jauhariyah

Department of Physics, State University of Surabaya, Surabaya, Indonesia
*e-mail: novymuizzah@gmail.com

Received: July 24, 2025. Accepted: September 10, 2025. Published: September 27, 2025

Abstract: The local wisdom of *Karang Batik* from Tuban can provide a contextual basis for developing scientific literacy test instruments that are more meaningful. This study aims to develop a science literacy test instrument integrated with the local wisdom of *Karang Batik* making, which is unique to Tuban, and to identify students' science literacy profiles on the subject of temperature and heat. The development was carried out using the ADDIE model with data collection through interviews, validation questionnaires, and tests. Data analysis included content validity (Aiken's V and Percentage of Agreement), empirical validity (item–total correlation), reliability (Cronbach's Alpha), difficulty level, and discriminating power. The instrument was tested on 30 university students and implemented for 70 students at Tuban 1 State Senior High School. The results showed that content validity was in the very good category (PoA ≥ 97%), 67% of the items were empirically valid, and reliability was 0.690, which is classified as high. The level of difficulty ranged from easy to moderate, with sufficient discrimination. Students' science literacy profiles varied from very low to very high, with the majority of students in the high category. These findings indicate that the developed instrument is suitable for measuring science literacy and contributes to contextual learning efforts based on local wisdom.

Keywords: *Karang Batik*; Scientific Literacy; Temperature And Heat; Test Instrument.

Introduction

In the face of global dynamics marked by rapid advances in science and technology, every individual is required to have the ability to play an active and meaningful role. This requires the education system to adjust its approach, especially in the transition to 21st-century learning that is centred on the learner. This era demands mastery of various essential skills needed to navigate the ever-evolving global landscape. One of the fundamental skills that is crucial to have is scientific literacy, which is not only relevant in an academic context but also in everyday decision-making based on logic and scientific evidence. The Ministry of Education, Culture, Research, and Technology has established six basic literacies as the foundation of 21st-century competencies, one of which is scientific literacy, which includes understanding scientific concepts and their application in real-life situations.

Science literacy has broad and complex dimensions. According to the OECD [1], science literacy encompasses three main competencies: explaining phenomena scientifically, designing and evaluating scientific investigations, and interpreting and using scientific information in decision-making. These competencies are the main benchmarks in international assessments such as PISA (Programme for International Student Assessment). However, PISA results show that Indonesian students' science literacy achievements are still low and tend to fluctuate. Indonesia's science literacy scores from 2012 to 2022 ranged from 384 to 403, far below the average for OECD member countries, which ranged from 489 to 494. Most recently, in PISA 2022, Indonesia scored 383 and ranked 69th out of 80 countries [2]. This fluctuation in scores indicates that there is no systematic and sustainable strategy to improve science literacy in Indonesia.

Various factors are believed to be the cause of low science literacy among students, ranging from a curriculum that is not yet fully contextual, limited facilities and infrastructure, to teaching methods that do not sufficiently encourage higher-order thinking skills [3], [4], [5]. The assessments used in schools also do not comprehensively measure science literacy skills. The questions created by teachers still tend to focus on memorisation and formula calculations, and do not make sufficient use of data in the form of graphs or tables [6], [7]. A more contextual learning approach that integrates local culture is believed to increase the relevance of the material and help students connect scientific concepts with the reality around them [8], [9].

In the context of educational evaluation, instruments are important tools. Good instruments must be valid, reliable, vary in difficulty, and be able to distinguish between the abilities of learners [10], [11]. A number of studies have developed contextual and locally-based science literacy instruments [7] [12], [13], but none have specifically integrated temperature and heat material with the local wisdom of *Karang Batik* in Tuban Regency. In fact, in the batik-making process, particularly dyeing, waxing, and drying, there are physical phenomena related to temperature, heat, and changes in the form of substances. Based on interviews with physics teachers at SMAN 1 Tuban, it was

How to Cite:
N. A. Muizzah and M. N. R. Jauhariyah, "Development of Science Literacy Test Instruments on Temperature and Heat Materials by Integrating Local Wisdom in the Production of Karang Batik, a Speciality of Tuban Regency", *J. Pijar.MIPA*, vol. 20, no. 6, pp. 1072–1079, Sep. 2025. https://doi.org/10.29303/jpm.v20i6.9817

found that even though students had visited the *Karang Batik* centre and participated directly in the process, they still found it difficult to connect the experience with the material studied in class. This indicates a learning gap that has not been optimally addressed in integrating scientific knowledge with the local context familiar to students.

Based on the above description, this study was conducted to develop a science literacy test instrument on the subject of temperature and heat by integrating the local wisdom of *Karang Batik*, which is unique to Tuban Regency. This study aims to describe the theoretical and empirical feasibility of the instrument, as well as to reveal the profile of science literacy abilities of high school students based on the developed instrument. The results of this study are expected to contribute to the development of science literacy assessments that are more contextual and relevant to students' surrounding environments, as well as to promote the improvement of science education quality in Indonesia that is adaptive to the demands of the 21st century.

Research Methods

This study is classified as Research and Development (R&D) research, using the ADDIE development model developed by [14], which consists of five stages: Analyze, Design, Development, Implementation, and Evaluate.

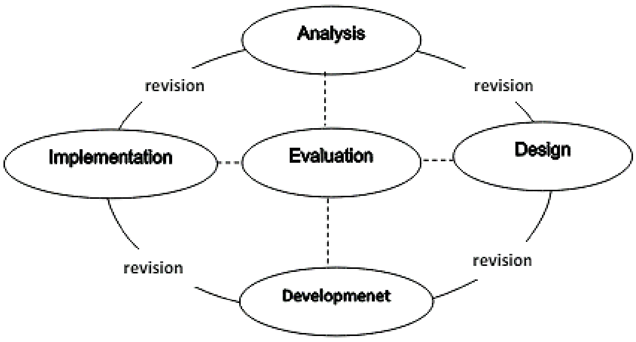

Figure 1. ADDIE Model Development Design (Branch, 2009)

In this study, each stage in the ADDIE development model is interrelated, forming a continuous and sustainable process. The Analyse stage was conducted through interviews with one of the Year 11 Physics teachers and the principal of SMAN 1 Tuban to analyse the needs and curriculum used as a reference in determining the material and assessment techniques. In the Design stage, a science literacy test instrument was developed that integrated the local wisdom of *Karang Batik*. The instruments developed included a question matrix, a question item specification table, and a question assessment rubric. In the Development stage, the instruments created were tested for theoretical validity by expert validators. In addition, limited trials were conducted with first-year students of the Physics Education Study Programme at Unesa to determine the empirical validity of the items. In the Implementation stage, the items that were assessed as valid and usable were applied to students to determine their science literacy profiles. Meanwhile, the Evaluate stage was conducted at each stage of development.

The participants in this study were selected using purposive sampling. A limited trial was first conducted on 30 first-year students of the Physics Education Study Programme at Unesa to obtain empirical validity data for the questions. Subsequently, the instruments that had been declared valid were implemented for 70 students in classes XI-G and XI-F at SMAN 1 Tuban. The selection of participants was based on availability and suitability with the material on temperature and heat being studied, while also considering the students' previous experience with the *Karang Batik*-making process.

Ethical clearance for this study was obtained from the school, and informed consent was secured from all participants. The research procedures adhered to ethical standards in educational research, ensuring confidentiality and the use of data solely for academic purposes.

Data were analyzed using quantitative descriptive techniques. The theoretical validity of the test instrument was assessed by three expert validators based on content, construction, and language. A Likert scale (Table 1) was used, and validators also provided suggestions for improvement.

Table 1. Likert Scale

Criteria	Score
Very valid	5
Valid	4
Quite Valid	3
Less Valid	2
Not Valid	1

Referring to Table 1, an instrument is considered valid if it obtains a minimum score that falls within the “valid” category. The scores given by the three validators are then used to calculate the percentage of agreement for each validator using the formula:

$$\text{Percentage of Agreement} = 100\% \times \left(1 - \frac{A-B}{A+B}\right)$$

In addition, construct validation was also carried out for each validator analysed using Aiken's Validity Theory (Aiken's V) with the following formula:

$$V = \frac{\Sigma S}{n(C-1)}$$

Explanation:
V : Aiken’s Validity
S : r - lo
r : value given by the validator
lo : lowest validation value
n : number of validators
C : highest validity value

The Aiken validity score obtained was compared with the Aiken validity standard on a scale of 0–1, with categories as shown in Table 2.

Table 2. Category Value Validity Aiken

Aiken's value	Criteria
$0 < V \leq 0.4$	Less Valid
$0.4 < V \leq 0.8$	Quite Valid
$0.8 < V < 1$	Very Valid

After being declared theoretically feasible, the instrument was piloted with 30 first-year Physics Education students at Unesa. Pilot data were analyzed in Microsoft Excel to assess item quality, including empirical validity (using product-moment correlation), reliability, difficulty level, and discriminating power.

$$r_{xy} = \frac{N\,\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}}$$

Explanation:
r_{xy} : Correlation coefficient for each item
N :Number of test respondents
X :Number of item scores
Y :Total number of scores

The results obtained are then compared with the rtable value to test the validity of the test items. According to the rule, if the rhitung value is greater than the ttable value, then the test item is considered significant [15].

Furthermore, the reliability of the essay-type test is measured using Cronbach's Alpha coefficient formula, as shown in the formula.

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\Sigma\sigma i^2}{\sigma t^2}\right)$$

Explanation:
$r11$: Reliability value
n : Number of items
$\Sigma\sigma i2$: Total variance of item scores
$\sigma t2$: Total variance

To determine the reliability category of the instrument in Table 3.

Table 3. Reliability Value Classification

Correlation Coefficient	Criteria
0.000 – 0.199	Very Low
0.200 – 0.399	Low
0.400 – 0.599	Moderate
0.600 – 0.799	High
0.800 – 1.000	Very High

The difficulty level of a question can be determined by the sample's ability to answer the question. The difficulty level of each question item can be calculated using a formula.

$$TK = \frac{\frac{\sum X}{N}}{\text{maximum score for each question}}$$

Explanation:
TK : Difficulty level
ΣX : Total score of students for each question item
N : Number of students

The difficulty level criteria for each question item are categorized in Table 4.

Table 4. Difficulty Level Criteria

Difficulty Level	Criteria
0.00 ≤ DL ≤ 0.30	Difficult
0.31 ≤ DL ≤ 0.70	Moderate
0.71 ≤ DL ≤ 1.00	Easy

The discriminating power of an item is the extent to which an instrument can distinguish between students with high and low abilities [16]. The discriminating power index coefficient can be calculated using a formula, with the criteria for the discriminating power of items in Table 5.

$$DP = \frac{\bar{X}\,K_A - \bar{X}\,K_B}{\text{maximum score for eaach question}}$$

Explanation:
DP : Discrimination index for question n
$\bar{X}K_A$: Average score of students in the top group for question n
$\bar{X}K_B$: Average score of students in the bottom group for question n

Table 5. Criteria for Item Discrimination

Discrimination Index	Criteria
0.00 ≤ DP	Very poor
0.00 ≤ DP ≤ 0.20	Poor
0.21 ≤ DP ≤ 0.40	Fair
0.41 ≤ DP ≤ 0.70	Good
0.71 ≤ DL ≤ 1.00	Very Good

In addition to the analysis mentioned earlier, the items deemed suitable for use were then administered to students to determine their science literacy profiles. Science literacy profiles can be calculated for each student by determining the percentage of critical thinking skills using a formula.

$$\text{Science Literacy Level} = \frac{\text{Score obtained}}{\text{Maximum score}}$$

The percentage of science literacy obtained is then categorized to analyze the extent of students' science literacy abilities, as presented in Table 6.

Table 6. Science Literacy Ability Categories

Value Range (%)	Category
0 – 39	Very Low
40 – 55	Low
56 – 65	Moderate
66 – 79	High
80 – 100	Very High

Results and Discussion

Analysis

The analysis stage began with pre-research interviews with 11th-grade physics teachers to determine the availability of science literacy test instruments. The results showed that the previously used instruments only covered scientific concept understanding and did not encompass all aspects of science literacy. The analysis also included the curriculum system implemented at SMA Negeri 1 Tuban, namely the Merdeka Curriculum. The curriculum review was conducted to examine the learning outcomes of grade XI as a basis for selecting materials and assessment techniques. The material used focused on temperature and heat, with the learning outcomes indicating that students mastered the concepts of heat and thermodynamics, as well as scientific process skills [17].

The interviews also revealed that the students had participated in a traditional Tuban batik-making activity. This experience is relevant for integrating into the learning of temperature and heat, so that students not only understand the concepts theoretically but also relate them to local wisdom [18].

Design

This stage involves designing test instruments, including blueprints, item specifications, and scoring rubrics, focusing on temperature and heat integrated with the local wisdom of *Karang Batik*.

The instruments in this study were developed comprehensively to cover all aspects of science literacy competencies; however, the findings indicate a dominance of the second indicator, namely the ability to interpret data and scientific evidence, consistent with the results of [19], which show low scores among students in this aspect. The instrument consists of multiple-choice and essay questions; multiple-choice questions assess conceptual understanding based on experimental stimuli, while essay questions measure critical thinking, scientific reasoning, and investigative design skills.

Each question begins with a contextual stimulus that contains implicit information to stimulate in-depth analysis [20]. The development of stimuli considers linguistic aspects, relevance to instructional content, and connection to real-life contexts [21]. It is presented in various formats, including visual and textual elements such as graphs, tables, images, and infographics. In this study, the stimulus was developed from the context of local wisdom, specifically the process of making *Karang Batik*, which includes scientific phenomena, structured data, and information to support evaluation and science-based decision-making.

Development

After the planning stage, the next stage involved developing the science literacy test instrument. The instrument that had been designed was then reviewed by three validators to obtain input and suggestions for improvement. The validation process was carried out by two physics lecturers from Surabaya State University and one physics teacher from SMA Negeri 1 Tuban. The review was conducted based on three domains: language, content, and construction. The validation results were analyzed using the Percentage of Agreement (PoA) method among validators to obtain theoretical validity, which was then presented in Table 7.

Table 7. Test Instrument Validity Results

Assessment Aspect	Modus	PoA (%)	Criteria
Content Domain	5	97.4	Highly Valid
Construction Domain	5	97.3	Highly Valid
Language Domain	5	100	Highly Valid

The results of theoretical validation indicate that the integrated science literacy test instrument, incorporating local wisdom, obtained a mode of 5 in the content, construction, and language domains, all of which fall under the highly valid category. Content validity refers to the extent to which the items, questions, or tasks in a test instrument accurately represent the behavior of the sample being measured in a comprehensive and balanced manner [22].

The content validity of the research indicates that the instrument designed is in line with the material according to the curriculum used. Meanwhile, construct validity refers to the suitability of the grid for the developed test items [23]. In this study, this refers to the instructions for completion and the assessment rubric. Language validity refers to the clarity of sentences, adherence to language rules, and the use of language that is easy to understand.

Additionally, the results of the validators' assessments were analyzed using the Percentage of Agreement (PoA) formula to assess the consistency of assessments between validators. The PoA percentage obtained was 97.4% for content, 97.3% for construction, and 100% for language aspects, all of which met the criteria for high validity. In addition, each item developed was also analyzed using Aiken's validity, so that the validity value of each item was obtained as shown in Table 8.

Table 8. Results of Aiken's Validity

Item	V Aiken	Description
1	0,92	Very Valid
2	0,83	Very Valid
3	0,92	Very Valid
4	0,92	Very Valid
5	0,83	Very Valid
6	1,00	Very Valid
7	1,00	Very Valid
8	0,92	Very Valid
9	0,92	Very Valid
10	1,00	Very Valid
11	0,75	Quite Valid
12	0,92	Very Valid
13	1,00	Very Valid
14	0,83	Very Valid
15	0,92	Very Valid

The results of Aiken's validity analysis demonstrate that the developed science literacy test instrument has strong validity. Most items achieved the category of very valid ($V \geq 0.83$), and four items reached the maximum score of 1.00, reflecting a high level of agreement among expert validators in terms of content, construct, and language. These findings show that the integration of Karang Batik local wisdom can be effectively incorporated into assessment items, making them both pedagogically sound and contextually relevant.

Although one item, namely item 11 with a V Aiken value of 0.75, was rated as quite valid and requires minor refinement for better clarity and alignment with the targeted indicator, the overall instrument is considered highly appropriate for classroom implementation. The results are in line with previous studies emphasizing the importance of contextualized assessments to improve student engagement and comprehension. Thus, by embedding elements of local culture such as Karang Batik, the instrument not only assesses science literacy but also enhances cultural relevance in science education.

After the theoretical review and validation by the three validators, the test instrument, consisting of 15 questions, was then pilot-tested on 30 first-year students of the Physics Education Program at Unesa to obtain empirical validity, including validity, reliability, difficulty level, and item discrimination, which were developed using Microsoft Excel. The details of the pilot test results are presented in Table 9.

Table 9. Results of Empirical Validity Testing of Valid Test Items

Item	Item Validity	Reliability	Item Difficulty Level	Item Discrimination Power	Description
1	0.446		0.827	0.240	can be used
2	0.762		0.760	0.213	can be used
3	0.464		0.944	0.233	can be used
4	0.518		0.294	0.222	can be used
5	0.487	0.690	0.694	0.289	can be used
6	0.586		0.606	0.222	can be used
7	0.468		0.781	0.229	can be used
8	0.387		0.838	0.343	can be used
9	0.591		0.720	0.293	can be used
10	0.420		0.693	0.253	can be used

Empirical validity was obtained from the results of a limited trial involving 30 first-year students in the Physics Education Study Program at Unesa. The analysis was conducted using the product-moment correlation formula to identify the r_{count} value. This analysis aimed to determine the extent to which each item consistently and accurately measured the intended construct or competency.

Based on the validity test results using the product-moment correlation formula, out of the 15 developed items, 10 were deemed valid because they had a correlation value (r_{count}) greater than the rtable value at the 5% significance level. Conversely, five items were invalid and therefore could not be used in the instrument. The 10 valid items include two items assessing the competency of explaining phenomena scientifically, six items assessing the competency of designing and evaluating scientific investigation designs and critically interpreting data and evidence, and two items assessing the competency of researching, evaluating, and using scientific information for decision-making and action.

This empirical validity also reinforces the theoretical validity that had previously been established through expert assessment. Thus, the developed instrument has been proven to be suitable for measuring the intended competencies, particularly in the context of science literacy, which includes the ability to explain scientific phenomena, design investigations, and critically interpret data.

The reliability test of the instrument was conducted using 10 questions that had been validated to assess the accuracy, consistency, and stability of the results obtained from the instrument in measuring what it was intended to measure [24]. The reliability test results indicated an r11 value or alpha value of 0.690. The instrument is considered reliable if r11 is greater than the r_{table} value [15]. The p-value for a sample size of 30 with a significance level of 5% is 0.361, indicating that it is greater than the r_{table} value (0.690 > 0.361). The test instrument is declared reliable with a high category according to [24] reliability value classification.

The results of the difficulty level analysis show that the 15 questions developed have easy, medium, and difficult levels. The details are as follows: 10 questions are classified as easy, question number 4 is classified as medium, and 1 question is classified as difficult, with a difficulty level ranging from 0.294 to 0.973. According to[15], questions in the moderate category have a difficulty level ranging from 0.31 to 0.70. The designed question instrument has varying levels of difficulty, ranging from easy to difficult.

Analysis of the discriminative power reveals that the instrument's discriminative power ranges from 0.105 to 0.293, with 67% of the items categorized as adequate and 33% as poor. Thus, the discriminative power of the developed instrument is classified as good. This means that the developed instrument is effective in distinguishing between students with high and low abilities. Overall, the analysis results indicate that the developed instrument meets the criteria for validity and reliability and possesses good item quality. Therefore, this instrument is deemed suitable for measuring students' science literacy profiles on the topic of temperature and heat, integrated with the local wisdom of *Karang Batik*-making.

Implementation

At this stage, 10 questions that were deemed valid and feasible were administered directly to 70 students in grades XI-G and XI-F at SMA Negeri 1 Tuban. The purpose of administering the instrument was to identify the students' science literacy profiles. The results of the students' science literacy profiles are presented in Figure 2.

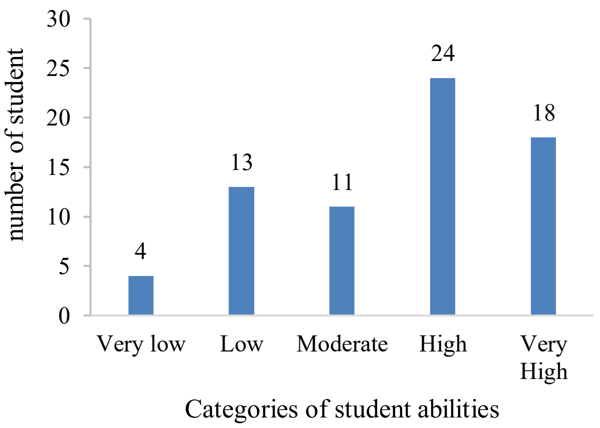

Figure 2. Graph of Critical Thinking Skills Profile of Students

Based on Figure 2. the results of the science literacy profile percentage of students measured using an integrated science literacy test instrument incorporating local wisdom in the making of Karang Batik, a specialty of Tuban, on the subject of temperature and heat show a varied distribution of students' science literacy abilities, with four students classified as having very low abilities, 13 students in the low category, 11 students with adequate ability, 24 students demonstrating high ability, and 18 students in the very high category.

The results indicate that some students have very low literacy skills. However, the majority of students have high and very high science literacy skills. Students with low literacy skills often struggle to comprehend basic science concepts, which hinders their ability to solve problems and apply knowledge in practice [25]. According to [26], the phenomenon of low student abilities is a result of the lack of science literacy-based learning processes in schools. The results of the students' science literacy profiles were further analyzed based on the average science literacy competency scores for each category of students, presented in a diagram in Figure 3.

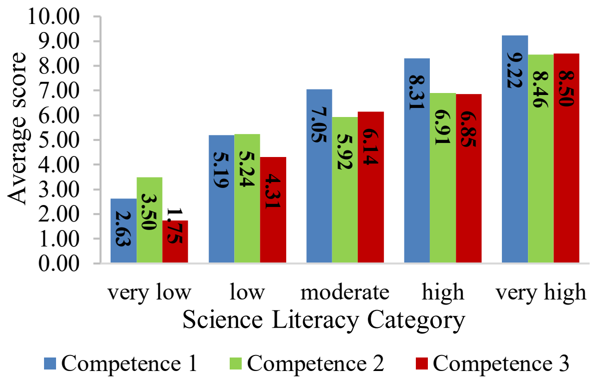

Figure 3. Average Science Literacy Competency Scores for Each Category of Students

Figure 3 presents the average results for each category of science literacy skills among students, as measured using an integrated science literacy test instrument based on local wisdom in the making of *Karang Batik*, a speciality of Tuban, specifically in the context of temperature and heat within each science literacy competency. Students in the very low category obtained low average scores in every competency, particularly in the competencies of researching, evaluating, and using scientific information for decision-making and action, with an average score of 1.75 out of a maximum score of 10. Students in the low category showed improvement in every competency. However, the third competency still had the lowest average score, which was 4.31.

Students in the adequate category showed more balanced average scores, with competency 1, explaining phenomena scientifically, still dominating with the highest average score of 7.05. Conversely, competency 2, designing and evaluating scientific investigations and critically interpreting data and evidence, had the lowest average score of 5.92. This suggests that students continue to face challenges in designing scientific investigations and analyzing data. Furthermore, in the high category, students obtained high and relatively balanced average scores. This suggests that students have a solid understanding of concepts and scientific thinking skills. However, when observed in competency 3, the average score obtained was the lowest compared to the others, specifically 6.85, indicating that students' ability to use scientific information for decision-making or action remains lacking. Students in the very high category achieved average scores close to the maximum across all competencies, ranging from 8.46 to 9.22, with the lowest average score in competency 2, which involves designing and evaluating scientific investigation designs and critically interpreting data and evidence. This may be due to limited experience in developing scientific investigations.

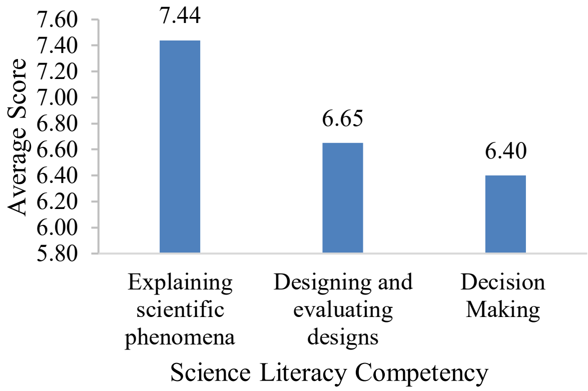

Figure 4. Overall Average Score Based on Science Literacy Competency

Based on Figure 4, the highest average score of students was in the competency of explaining scientific phenomena (7.44), followed by organizing and evaluating research designs and interpreting data (6.65), and the lowest was in the competency of researching, evaluating, and using scientific information for decision making (6.40).

The high score in the first competency is attributed to lower cognitive demands, which are limited to memorizing concepts [4]. Lower scores in the second and third competencies indicate that students have weak skills in problem-solving, understanding data presented in graphs or tables [27], and critically evaluating scientific information and relating it to real-world contexts [4]. This phenomenon is related to the weak ability of students to critically evaluate information sources, as well as their inability to relate scientific information to real-life situations in order to make decisions.

These findings underscore the importance of enhancing students' abilities to evaluate and apply scientific information in informed decision-making. Therefore, the development of instruments or learning models focused on these skills is essential to pursue in further research. In addition, the limitation of this study lies in the number of questions that are not balanced between competencies, particularly in the third competency, which is researching, evaluating, and using scientific information for decision-making and action. The limited number of questions means that the participants' ability to use scientific information in decision-making and action is not optimally represented.

Evaluation

Evaluation is carried out continuously at every stage of instrument development. During the analysis and design stage, evaluation focuses on the suitability of learning needs and expert input regarding the clarity of instructions, the readability of stimuli, and suitability to the context of science literacy. The development stage involves theoretical testing by experts and empirical testing with 30 students to assess validity, reliability, difficulty level, and discriminating power.

The results of the item analysis show five invalid items (numbers 2, 6, 9, 11, and 14), with validity values below the r-table (0.361) or negatively correlated with the total score. For example, item number 2: 'The tool used to melt wax in batik making is a small pan, usually made of aluminium or steel. Based on the concept of heat transfer, why are these materials chosen? Give your reasons!' has a calculated r value of -0.062, which indicates incompatibility with the measured competency. In addition, most of the invalid items were too easy and had low discriminative power (<0.20), making them ineffective in distinguishing students based on their level of science literacy.

Translated with DeepL.com (free version)The overall reliability of the instrument was 0.690, indicating high reliability, but invalid items can reduce the consistency of the instrument and therefore need to be eliminated. Additionally, the implementation stage was evaluated based on the clarity of instructions, the duration of work, and student comfort. The evaluation results were used to refine the instrument before it was widely used.

Conclusion

This study concludes that the science literacy instrument on temperature and heat, enriched with *Karang Batik* local wisdom, shows high validity in content (97.44%), construct (97.30%), and language (100%), with a reliability score of 0.690. Empirical testing indicates that most items are valid and fall within easy-to-moderate difficulty levels, while students' literacy profiles vary widely from very low to very high. Beyond the statistical findings, the integration of *Karang Batik* adds a significant contextual dimension to science education. By connecting abstract scientific principles with familiar cultural practices, the instrument increases the relevance of learning, fosters student engagement, and supports a more meaningful grasp of scientific concepts. For teachers, the instrument serves not only as a valid evaluation tool but also as a practical resource to strengthen science literacy while simultaneously preserving and valuing local cultural identity. Future research is recommended to extend this approach to other scientific domains and cultural settings, thereby enhancing the contribution of local wisdom in shaping contextualized and impactful science education.

Author's Contribution

Novy' Ainul Muizzah: Responsible for research conceptualization, data collection, and data analysis; Mukhayyarotin Niswati Rodliyatul Jauhariyah: Academic supervision, methodological review, analysis result validation, and manuscript quality assurance.

Acknowledgements

Thank you to the teachers who granted permission for the research to be conducted at the school, as well as to all the students who participated in this activity. The author also extends their most profound appreciation to the validators who provided valuable input to improve the research instruments, as well as to fellow students who assisted in the data collection process.

References

[1] OECD, "Kerangka Kerja Sains Pisa 2025 (Draf)," vol. 2025, pp. 9–84, 2023.

[2] OECD, *Pisa 2022 Results*, vol. I. 2023.

[3] F. Yusmar and R. E. Fadilah, "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, 2023, doi: 10.24929/lensa.v13i1.283.

[4] L. Zuhrotul, M. Niswati, and R. Jauhariyah, "Analisis Profil Kompetensi Literasi Sains Siswa Sma Pada Materi," *Inov. Pendidik. Fis.*, vol. 12, no. 2, pp. 74–80, 2023, [Online]. Available: https://ejournal.unesa.ac.id/index.php/inovasi-pendidikan-fisika/article/view/53009

[5] Y. Rosa Lina *et al.*, "Pengembangan Instrumen Asesmen Berbasis Literasi Sains tentang Materi Sistem Pencernaan Makanan, Zat Aditif, dan Zat Adiktif untuk SMP," *Bioeducation J.*, vol. 2, no. 2, p. 145, 2018.

[6] N. Novita, M. Mellyzar, and H. Herizal, "Asesmen Nasional (AN): Pengetahuan dan Persepsi Calon Guru," *JISIP (Jurnal Ilmu Sos. dan Pendidikan)*, vol. 5, no. 1, 2021, doi: 10.58258/jisip.v5i1.1568.

[7] W. W. Murti and T. Sunarti, "Pengembangan Instrumen Tes Literasi Sains Berbasis Kearifan Lokal Di Trenggalek," *ORBITA J. Kajian, Inov. dan Apl. Pendidik. Fis.*, vol. 7, no. 1, p. 33, 2021, doi: 10.31764/orbita.v7i1.4386.

[8] N. D. Mashitoh and M. N. R. Jauhariyah, "Analysis of Students' Science Literacy Competency Profiles Measured Using Class Based Instrument on Renewable Energy Materials." *J. IPF: Inovasi Pendidikan Fisika,* vol 12, no. 3, pp 184-190, 2024, doi:10.26740/ipf.v13n3.p184-190

[9] S. R. Cahyani, D. A. Pratama, A. Z. Mawaddah, N. D. Mashitoh, and U. A. Deta, "Analisis Konsep Fisika pada Budaya Kesenian Kearifan Lokal Permainan Gulat Okol Surabaya," *J. Ilmu Pendidik. dan Pembelajaran*, vol. 3, no. 1, pp. 9–15, 2024, doi: 10.58706/jipp.v3n1.p9-15.

[10] I. A. F. Faruq and A. R. Afiah, "Instrumen Penilaian Berbagai Kompetensi atau Indikator Dalam Pencapaian Hasil Belajar SD/MI," *Univ. Muhammadiyah Sidoarjo*, no. 152071200010, p. 7, 2018.

[11] L. Z. Wardi and M. N. R. Jauhariyah, "Validity and reliability scientific literacy-based physics assessment instrument on atomic nucleus and radioactivity material," ... *Phys. Educ. J.*, vol. 9, no. 1, 2025, doi: 10.21067/mpej.v9i1.10450.

[12] M. Asniati, "Pengembangan Instrumen Soal Literasi Sains Berbasis Google Form Untuk Siswa Smp Pada Materi Kalor," *Skripsi*, vol. 53, no. 9, pp. 1689–1699, 2019, [Online]. Available: file:///C:/Users/User/Downloads/fvm939e.pdf

[13] Maya Ningsetyo and Titin Sunarti, "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, 2024, doi: 10.61132/arjuna.v2i2.612.

[14] R. M. Branch, *Approach, Instructional Design: The ADDIE*, vol. 53, no. 9. 2009.

[15] S. Arikunto, *Dasar-Dasar Evaluasi Pendidikan Edisi 2.* 2012.

[16] S. Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik Penelitian : Suatu Pendekatan Praktik*. 2013.

[17] Kemendikbudristek, *Peraturan Menteri Pendidikan, Kebudayan, Riset, dan Teknologi Nomor 032/H/KR/2024*, no. 021. 2024.

[18] J. Murdana, "Pengembangan LKPD Fisika Materi Suhu dan Kalor Bercirikan STEM Terintegrasi Etnosains pada Batik Bakaran," *Skripsi,* vol. 8 no. 5, pp. 2003–2005, 2022.

[19] M. S. Rohmania, "Analisis Kemampuan Literasi Sains Pada Domain Pengetahuan dan Kompetensi Materi Suhu dan Kalor Kelas 5 SDN Sumput," *Jpgsd*, vol. 12, no. 08, pp. 1592–1602, 2024.

[20] F. O. Widarta and W. Artika, "Analisis Bentuk Stimulus, Dimensi Kognitif, dan Karakteristik HOTS pada Instrumen Evaluasi Mata Pelajaran IPA Karya Guru," *J. IPA Pembelajaran IPA*, vol. 5, no. 3, pp. 197–208, 2021, doi: 10.24815/jipi.v5i3.21429.

[21] F. M. Elita, M. B. Zakawali, Z. Putri, S. Safitri, and R. I. Indra Putri, "Pengujian Soal Berbasis HOTS pada Pembelajaran Sejarah di Sekolah IT Raudhatul Ulum Sakatiga Indralaya," *Danadyaksa Hist.*, vol. 2, no. 2, p. 158, 2023, doi: 10.32502/jdh.v2i2.5672.

[22] M. F. Ramadhan, R. A. Siroj, and M. W. Afgani, "Validitas and Reliabilitas," *J. Educ.*, vol. 6, no. 2, pp. 10967–10975, 2024, doi: 10.31004/joe.v6i2.4885.

[23] S. S. Rohmad, *Pengambangan Istrumen Angket*, vol. 7, no. 2. 2021.

[24] Sugiyono, *Metodologi Penelitian Kuantitatif, Kualitatif dan R & D*. 2020.

[25] A. Rahman, "Asesmen Kemampuan Literasi Sains Siswa pada Mata Pelajaran IPA di SMP Kota Bandung" *Tesis,* 2024.

[26] D. H. Arofi and P. Dwijananti, "Kemampuan Literasi Sains Siswa SMAN 3 Pekalongan pada Aspek Kompetensi Materi Radioaktivitas," *Unnes Phys. Educ. J.*, vol. 13, no. 1, pp. 75–86, 2024, doi: 10.15294/upej.v13i1.8634.

[27] T. S. J. Waini, A. P. Yani, R. W. Wardana, M. Uliyandari, and S. Sutarno, "Analisis Kemampuan Literasi Sains Siswa Kelas Viii Di Smpn 9 Kota Bengkulu Pada Isu Pencemaran Lingkungan Di Kota Bengkulu," *DIKSAINS J. Ilm. Pendidik. Sains*, vol. 5, no. 1, pp. 62–69, 2024, doi: 10.33369/diksains.5.1.62-69.