Primary: Jurnal Pendidikan Guru Sekolah Dasar
journal homepage: https://primary.ejournal.unri.ac.id/
contents lists available at SINTA and DOAJ

Vol. 14 No. 3 (2025): June
http://dx.doi.org/10.33578/jpfkip.v14i3.p355-373

Development of cognitive diagnostic assessment instruments and a cognitive differentiation student worksheet to measure and improve scientific literacy capability

Joko Supriono [1*], Pramudiyanti [1], Fatkhur Rohman [1], Undang Rosidin [1]

[1] Master of Primary School Teacher Training, Faculty of Teacher Training and Education Universitas Lampung, Bandar Lampung, Indonesia, 35141

Article info	Abstract
Keywords: cognitive diagnostic assessment, cognitive differentiated student worksheet, and scientific literacy ability.	This research develops cognitive diagnostic assessment instruments and cognitively differentiated worksheets designed to enhance students' scientific literacy in science learning. Using a Research and Development (R&D) approach based on the Plomp model, the study involved 28 students, each from Class IV A and IV B at MIN 1 North Lampung. Data was collected through documentation, questionnaires, interviews, and observation, and analyzed using t-tests, practicality tests, and product validation. Expert validation showed the instruments and worksheets were highly valid, with average scores of 92% for cognitive diagnostic assessment questions, 94% for post-test questions, and 94% for material, linguistic, and media validation. The practicality test showed educators' responses at 84% and students' at 87%, both rated as very practical. A t-test revealed a significant difference ($p = 0.001$, $p < 0.05$) between the experimental and control classes in post-test scores. The results indicate that the developed cognitive diagnostic assessment instruments and differentiated worksheets are valid, practical, and effective in enhancing students' scientific literacy.

* Corresponding Author.
E-mail address: joko.supriono10@gmail.com (Joko Supriono)
DOI: http://dx.doi.org/10.33578/jpfkip.v14i3.p355-373
Received 29 April 2024; Received in revised form 28 May 2025; Accepted 03 June 2025
Available online 29 June 2025
e-ISSN 2598-5949 | p-ISSN 2303-1514

1. Introduction

The concept of Society 5.0, proclaimed by Japan, represents a new society that integrates technology with human-centered values. This concept envisions solving social problems through the fusion of technologies such as big data, the Internet of Things (IoT), artificial intelligence (AI), and community services to enhance both digital and physical infrastructure (Narvaez Rojas et al., 2021). The Society 5.0 era promotes sustainable community development (Sułkowski et al., 2021), emphasizing the evolving relationships between technology and society, as well as the role of technology in mediating interactions among individuals (Deguchi et al., 2020). With these changes, traditional knowledge and skills are becoming less relevant, requiring new competencies, such as creativity, entrepreneurship, and global competence (Zhao & Watterston, 2021). These shifts demand significant curriculum changes in education to equip students with these new skills.

Zhao & Watterston (2021) suggest that curricula must enable students to develop new competencies and allow for personalized learning, providing students the freedom to choose their learning paths. Teachers must also adjust their teaching methods to accommodate these new educational needs, focusing on differentiated instruction (Tomlinson, 2014) and personalized learning (Kallio & Halverson, 2020). Traditional direct teaching methods should be reconsidered, as they often lead to long-term adverse outcomes (Bonawitz et al., 2011; Buchsbaum et al., 2011; Kapur, 2016). Instead, education must shift toward supporting inquiry-based, learner-centered approaches that address authentic problems and develop students' ability to navigate uncertainty (Zhao & Watterston, 2021).

The evolving education system also necessitates changes in assessment practices. Traditional low-level assessments should be replaced with those that assess higher-order thinking skills (HOTS), fostering critical and creative thinking (Akib & Muhsin, 2019). Such assessments align with the demands of the 21st-century workforce, which requires the ability to analyze, synthesize, and apply knowledge in solving real-world problems (Pellegrino, 2014). Assessment should also emphasize collaboration and communication skills, as well as critical thinking (Basilio & Bueno, 2021).

The integration of 21st-century skills into education is crucial. According to Trilling & Fadel (2009), these skills encompass life and career skills, learning and innovation skills, and information, media, and technology skills. Burkhardt et al. (2003) identify essential skills for future generations, including digital literacy, innovative thinking, effective communication, and high productivity. Digital literacy, in particular, encompasses a range of skills, including scientific, economic, technological, visual, information, and multicultural literacy (Burkhardt et al., 2003). Aoun (2017) expands on this by adding data literacy, technological literacy, and human literacy as critical components of digital literacy. These literacies enable individuals to navigate both the digital world and their interactions with others.

Internationally, institutions like PISA emphasize the importance of 21st-century skills in shaping curricula. The OECD's 2030 framework emphasizes the need for curricula that equip students for future industry demands by fostering critical, creative, and reflective thinking (OECD, 2019). Scientific literacy is critical, as it plays a central role in addressing societal challenges (Briseño-Garzón et al., 2014; BouJaoude, 2002). In Indonesia, the National Literacy Movement (GLN) has prioritized literacy, with a primary focus on reading literacy (Nugraha & Octavianah, 2020). The performance of Indonesian students in PISA, however, has been consistently low, particularly in scientific literacy. In 2018, Indonesia ranked 72nd out of 78 countries, with a scientific literacy score significantly below the OECD average (Pratiwi, 2019).

To improve scientific literacy, Indonesia must adopt strategies to enhance teachers' competencies and implement meaningful science learning that connects to real-life contexts. The use of inquiry-based learning strategies is one approach that has been shown to improve students' scientific thinking abilities (Kang, 2022; Oliver et al., 2021). Information technology also offers opportunities to integrate science content with technology, enhancing students' scientific literacy (Chen et al., 2021). Teachers should design assessments that measure scientific literacy and use them to improve the learning process (Suhandi & Robi'ah, 2022). In line with the OECD's recommendations, Indonesia has introduced the National Assessment (AN), which emphasizes literacy and character surveys over cognitive ability (Sari, 2020).

One significant step taken by the Indonesian government is the introduction of the Indonesian Madrasah Competency Assessment (AKMI), which aims to assess students' literacy abilities in reading, numeracy, scientific literacy, and socio-cultural literacy. The results of these assessments provide valuable insights into students' strengths and weaknesses, helping teachers tailor their teaching to meet individual student needs (Lessy, 2022). However, teachers at MIN 1 North Lampung have reported difficulties in preparing and implementing scientific literacy assessments based on the AKMI recommendations. There is also a lack of communication regarding the purpose of the assessments and how to implement them effectively.

To address these challenges, this research proposes the development of a cognitive diagnostic assessment instrument and differentiated student worksheets to improve scientific literacy. The proposed solution involves aligning assessments with the students' cognitive levels and using problem-based learning (PBL) strategies. Research has shown that diagnostic assessments can enhance learning outcomes (Firmanzah & Sudibyo, 2021; Hikmasari et al., 2018) and that PBL can effectively improve scientific literacy (Izzatunnisa et al., 2019; Ain & Mitarlis, 2020). The aim is to create a more personalized learning experience that enhances students' scientific literacy abilities and prepares them for future challenges.

In conclusion, transforming education systems to meet the demands of Society 5.0 requires changes in curriculum, teaching methods, assessments, and literacy development. These changes must be supported by practical teacher training and the development of appropriate assessment tools. By implementing diagnostic assessments, differentiated learning strategies, and inquiry-based methods, Indonesia can improve its students' scientific literacy and better prepare them for the challenges of the 21st century.

2. Method

This research was conducted through a development (R&D) approach. This research aims to develop cognitive diagnostic assessment instruments and cognitively differentiated student worksheets in science and science lessons, Phase B, that are valid, practical, and effective for measuring and improving students' scientific literacy abilities. This research uses a Plomp model, which is divided into three phases, namely, 1) Preliminary research that can be carried out at this stage is literature study, needs and context analysis, curriculum analysis, and conceptual framework development. 2) Designing and developing prototypes. At this stage, product design is carried out, followed by formative evaluation to measure the product interactively. 3) Evaluation, which is a measurement stage where the intervention carried out meets predetermined criteria (Plomp & Nieveen, 2013).

2.1 Preliminary research stage

The first stage of product development involves three key steps: literature analysis, needs and context assessment, and the development of a conceptual framework. The literature analysis focuses on reviewing relevant articles to establish the theoretical foundation and identify the novelty of the research, specifically regarding cognitive diagnostic assessment instruments and student worksheets for improving scientific literacy. The needs and context assessment involves collecting empirical data from school settings through questionnaires and structured interviews with Class IV teachers and students at MIN 1 North Lampung, examining factors such as curriculum implementation, learning outcomes, 21st-century learning, and teachers' understanding of scientific literacy. This step also includes documenting the results of the 2021 and 2023 AKMI assessments. Finally, a conceptual framework is developed, outlining the components for measuring scientific literacy, including content, competencies, material depth, and PBL learning syntax, tailored to the cognitive levels of students. This framework guides the product development to ensure alignment with the identified needs and goals.

2.2 Design and develop the prototyping stage

This stage involves three key research activities: designing a product prototype, conducting formative evaluation, and revising the product. First, the design of the product prototype begins after the research design proposal is made. The prototype, comprising a diagnostic assessment instrument and student worksheets, was developed based on the theoretical foundation established during the preliminary research stage. The second activity, formative evaluation, follows Plomp's model (Plomp & Nieveen, 2013), which includes several steps. Initially, the researchers conduct self-evaluation to identify errors in the prototype, such as typos, layout issues, or font choices, resulting in prototype 2. An expert review is then conducted by three experts in relevant fields (language, content, media, and assessment), resulting in prototype 3. The third activity, the one-to-one evaluation, involves testing prototype 3 with one educator and students from Class IV A, representing a range of performance levels. This evaluation assesses the product's practicality and identifies any remaining issues, ultimately leading to the development of prototype 4. Finally, a small group evaluation is conducted with one educator and nine students from Class IV B to assess the practicality of prototype 4 further, culminating in the development of prototype 5.

2.3 Evaluation stage

This stage will test the effectiveness of the final prototype, after which a field test evaluation will be conducted. This stage aims to obtain an assessment of the product's effectiveness with a larger number of students, specifically 28 class IV C students as the control group and 28 class IV A students as the experimental group. The final result of the prototype is a cognitive diagnostic assessment instrument and a Scientific Literacy student worksheet that are valid, practical, and effective in measuring Scientific Literacy abilities and improving them in the phase B class of water cycle material and efforts to maintain water availability.

3. Results

Product development using the Borg & Gall model can be completed well. The results of the initial research, conducted through preliminary studies in the form of field and literature reviews, indicate a problem: the low level of students' scientific literacy abilities. The 2022 PISA survey indicates that the scientific literacy abilities of Indonesian students fall into the low category, ranking 67th out of 81 countries. The low literacy of Indonesian students is caused by school culture and environmental factors that have not enabled activities to help students analyze, think critically, and apply scientific knowledge in everyday life. Apart from that, the teacher-centered learning factor causes students' interest in learning science to be relatively low. Therefore, educators are required to be able to use cognitive diagnostic assessments and cognitively differentiated student worksheets so that students can learn according to their needs (Roslina et al, 2025).

3.1 Preliminary research stage

This stage involves three key steps: literature analysis, needs assessment, and the development of a conceptual framework. The literature analysis emphasizes the significance of cognitive diagnostic assessment instruments and student worksheets in evaluating and enhancing students' scientific literacy. Cognitive diagnostic assessments help identify students' mastery of science concepts and learning difficulties, while student worksheets strengthen scientific thinking through structured activities. The needs assessment involves collecting data from Class IV teachers and students at MIN 1 North Lampung through questionnaires and structured interviews, revealing challenges such as students' struggles with literacy-based questions and teachers' difficulties in preparing appropriate assessment tools. Despite positive results from the AKMI, many teachers are unable to implement government recommendations due to a lack of knowledge and resources. The conceptual framework developed from this data outlines components for measuring scientific literacy, including content, competencies, and PBL learning strategies tailored to students' cognitive levels, guiding product development to ensure alignment with identified needs and objectives.

3.2 Design and develop the prototyping stage

This stage involves three key research activities: designing a product prototype, conducting formative evaluation, and revising the product. The product prototype, consisting of a diagnostic assessment instrument and student worksheets, was developed based on the theoretical foundation established during the preliminary research. The formative evaluation process, as outlined by Plomp and Nieveen (2013), includes self-evaluation, where researchers identify and correct errors such as typos, layout issues, and font choices, resulting in prototype 2. An expert review follows, with experts in language, content, media, and assessment validating the prototype, which leads to the development of prototype 3. Material expert validation is then conducted, with three experts providing feedback. The validation results showed that the prototype scored an average of 94% in validity, categorized as "Very valid, very complete, can be used." Based on these evaluations, revisions were made to improve the product's quality.

Table 1. Material expert validation assessment results

Validator	Score obtained	Score Max	%
Validator 1	63	72	87%
Validator 2	69	72	96%
Validator 3	69	72	96%
Total	201	216	94%
Criteria	**Very valid**		

The linguistic validation process yielded valuable suggestions and input, which were used to refine the product. Three linguists conducted the validation, with the results showing scores of 37/40, 38/40, and 37/40, respectively. These scores were analyzed using descriptive percentage analysis, resulting in an average linguistic validation score of 93%. Based on the expert validation assessment criteria, the linguistic aspect was categorized as "Very valid, very complete, can be used." These results highlight the high validity of the linguistic elements in the product.

Table 2. Linguist expert validation assessment results

Validator	Score obtained	Score Max	%
Validator 1	37	40	92%
Validator 2	38	40	95%
Validator 3	37	40	92%
Total	112	120	93%
Criteria	**Very valid**		

The media expert validation process provided valuable suggestions and input, which were used to improve the product. Three media experts conducted the validation, with scores of 64/72, 69/72, and 68/72, respectively. These scores were analyzed using descriptive percentage analysis, resulting in an average media expert validation score of 93%. Based on the expert validation assessment criteria, the media aspect was categorized as "Very valid, very complete, can be used." These results indicate a high level of validity in the media aspect of the product.

Table 3. Media expert validation assessment results

Validator	Score obtained	Score Max	%
Validator 1	64	72	89%
Validator 2	69	72	96%
Validator 3	68	72	94%
Total	201	216	94%
Criteria	**Very valid**		

The validation of the Cognitive Diagnostic Assessment Instrument provided valuable suggestions and input, which were incorporated into the product. Three evaluation experts conducted the validation, with the following results: Expert 1 scored 84% in construction, 89% in relevance, and 73% in clarity; Expert 2 scored 100% in construction, relevance, and clarity; Expert 3 scored 82% in construction, 100% in relevance, and 98% in clarity. These scores were analyzed using descriptive percentage analysis, yielding an average evaluation expert validation score of 92%. Based on the expert validation assessment criteria, the instrument's validity was categorized as "Very valid, very complete, can be used." These results confirm the high level of validity for the Cognitive Diagnostic Assessment Instrument.

The validation of the post-test questions involved expert evaluation, which provided suggestions and input for improvements. Three evaluation experts conducted the validation, with the following results: Expert 1 scored 83% in construction, 86% in relevance, and 81% in clarity; Expert 2 scored 100% in all areas (construction, relevance, and clarity); Expert 3 scored 98% in construction, 97% in relevance, and 100% in clarity. These results were analyzed using descriptive percentage analysis, yielding an average evaluation expert validation score of 94%. Based on the expert validation assessment criteria, the post-test questions were categorized as

"Very valid, very complete, can be used," confirming the high level of validity of the post-test questions.

Table 4. Results of expert validation assessment evaluation of cognitive diagnostic assessment questions

Information	Score obtained	Score Max	%
Construction	105	125	84%
Relevance	111	125	89%
Clarity	92	125	73%
Construction	125	125	100%
Relevance	125	125	100%
Clarity	125	125	100%
Construction	103	125	82%
Relevance	125	125	100%
Clarity	122	125	98%
Total	1.033	1.125	92%
Criteria	**Very valid**		

Table 5. Expert validation assessment results post test question evaluation

Information	Score obtained	Score Max	%
Construction	104	125	83%
Relevance	108	125	86%
Clarity	101	125	81%
Construction	125	125	100%
Relevance	125	125	100%
Clarity	125	125	100%
Construction	123	125	98%
Relevance	121	125	97%
Clarity	125	125	100%
Total	1,057	1.125	94%
Criteria	**Very valid**		

In the one-to-one evaluation stage, a practicality test was conducted on prototype 3 to identify visible errors and assess the product's practicality through feedback from students and educators. The evaluation involved students from Class IV A, selected based on their academic ranking (high, medium, low), and one educator. The evaluation results from the educator showed a score of 30 out of 32, which, when analyzed using descriptive percentage analysis, yielded an average score of 94%. Based on the practicality assessment criteria, the product was categorized as "Very practical, very complete, can be used," confirming its high level of practicality. These results are presented in Table 6.

Table 6. Results of one-to-one educator evaluation

Response	Score obtained	Score Max	%
Validator 1	30	32	94%
Total	30	32	94%
Criteria	**Very valid, very complete, usable**		

The results of the one-to-one evaluation with students yielded a score, with the maximum possible score. These results were then analyzed using descriptive percentage analysis to obtain an average score, reflecting the material expert validation score of %. Based on the practicality assessment criteria table, the product was categorized as "Very practical, very complete, can be used," indicating its high level of practicality. The results of the expert validation assessment are presented in Table 7.

Table 7. Results of the student one-to-one evaluation assessment

Response	Score obtained	Score Max	%
High Level	12	32	38%
Medium	11	32	34%
Low Level	6	32	19%
Total	29	32	91%
Criteria	**Very valid, very complete, usable**		

In the small group evaluation stage, a second practicality test was conducted on prototype 4 to assess its practicality from both educators and students in a small group setting. This evaluation involved one educator and nine students from Class IV B, selected based on specific criteria, resulting in prototype 5. The educator assessment yielded a score of 27 out of 32, which, after descriptive percentage analysis, resulted in an average score of 84%. Based on the practicality assessment criteria table, the product was categorized as "Very practical, very complete, can be used," indicating its high level of practicality. The results of the expert validation assessment are presented in Table 9.

Table 8. Educator assessment results

Response	Score obtained	Score Max	%
Teacher	27	32	84%
Total	27	32	84%
Criteria	**Very valid, very complete, usable**		

The results of the student assessment yielded an average score of 87%. Using the practicality assessment criteria table, it was determined that the level of practicality based on the students' responses falls into the category "Very practical, very complete, can be used." The results of the expert validation assessment are presented in Table 9.

Table 9. Student assessment results

Response	Score obtained	Score Max	%
Student 1	28	32	87%
Student 2	30	32	94%
Student 3	31	32	97%
Student 4	28	32	87%
Student 5	27	32	84%
Student 6	26	32	81%
Student 7	25	32	78%
Student 8	28	32	87%
Student 9	30	32	94%
Total	**251**	**288**	**87%**
Criteria	**Very valid, very complete, usable**		

[e-ISSN 2598-5949 | p-ISSN 2303-1514]

3.3 Assessment phase

This stage involves testing the effectiveness of the final prototype through a field test evaluation. The goal is to assess the product's effectiveness with a larger group of students, including 28 students from Class IV C as the control group and 28 students from Class IV A as the experimental group. The final prototype yields a valid, practical, and effective diagnostic cognitive assessment instrument and student worksheet for Scientific Literacy, designed to measure and improve scientific literacy skills in phase B, covering topics such as the air cycle and efforts to maintain air quality. Prior to data analysis, the researcher conducted an instrument revitalization test as follows.

Table 10. Validity test results of the cognitive diagnostic assessment

Question Number	r_{count}	r_{table}	Information
1	0,476	0.374	Valid
2	0,429	0.374	Valid
3	0,642	0.374	Valid
4	0, 642	0.374	Valid
5	0, 642	0.374	Valid
6	0,639	0.374	Valid
7	0,335	0.374	Valid
8	0,500	0.374	Valid
9	0,500	0.374	Valid
10	0,468	0.374	Valid
11	0,409	0.374	Valid
12	0,476	0.374	Valid
13	0,404	0.374	Valid
14	0,525	0.374	Valid
15	0,438	0.374	Valid
16	0,517	0.374	Valid
17	0,694	0.374	Valid
18	0,724	0.374	Valid
19	0,639	0.374	Valid
20	0,639	0.374	Valid
21	0,388	0.374	Valid
22	0,396	0.374	Valid
23	0,456	0.374	Valid
24	0,461	0.374	Valid
25	0,461	0.374	Valid

Based on the validity calculation of the cognitive diagnostic assessment questions in Table 10, using the Microsoft Office Excel 2013 program, with the test criteria that rcount > rtable at a significance level of 1% (α = 0.01), the rtable value is 0.374. In calculating the validity of the questions, it can be concluded that 25 out of 25 questions are valid and can be used in research.

Table 11. Post-test question validity test results

Question Number	r_{count}	r_{table}	Information
1	0,429	0.374	Valid
2	0,400	0.374	Valid
3	0,499	0.374	Valid
4	0,499	0.374	Valid

Question Number	r_{count}	r_{table}	Information
5	0,642	0.374	Valid
6	0,401	0.374	Valid
7	0,447	0.374	Valid
8	0,500	0.374	Valid
9	0,500	0.374	Valid
10	0,468	0.374	Valid
11	0,409	0.374	Valid
12	0,476	0.374	Valid
13	0,404	0.374	Valid
14	0,499	0.374	Valid
15	0,404	0.374	Valid
16	0,476	0.374	Valid
17	0,694	0.374	Valid
18	0,724	0.374	Valid
19	0,482	0.374	Valid
20	0,482	0.374	Valid
21	0,429	0.374	Valid
22	0,391	0.374	Valid
23	0,417	0.374	Valid
24	0,500	0.374	Valid
25	0,417	0.374	Valid

Based on the calculation of the validity of the post-test questions in Table 11 using the Microsoft Office Excel 2013 program, with the test criteria that rcount > r_{table} at a significance level of 1% (α = 0.01), the rtable is 0.374. In calculating the validity of the questions, it can be concluded that 25 out of 25 questions are valid and can be used in research.

Reliability testing is carried out to determine the level of accuracy or consistency of a cognitive diagnostic assessment question. This test was taken from 28 respondents with 25 questions and was carried out using Cronbach's Alpha assisted by the Microsoft Office Excel 2013 program which can be seen in tables 12 and 13.

Table 12. Reliability test results for cognitive diagnostic assessment questions

Reliability	N	Category
0,871	28	Very strong

Based on Table 12, it is known that the reliability test results for cognitive diagnostic assessment questions were obtained at 0.871. This indicates that the reliability of the categorized test items is very strong and can be relied upon.

Table 13. Reliability test results for post-test questions

Reliability	N	Category
0, 863	28	Very strong

Based on Table 13, it is known that the reliability test results for the post-test questions were obtained at 0.863. This indicates that the reliability of the categorized test items is very strong and can be used. The differentiating power of a question item refers to a question used to categorize students into two groups: the upper group, comprising students with high ability, and the lower group, comprising students with low ability. This differential power test was conducted using the Microsoft Office Excel 2013 program, as shown in Tables 14 and 15.

Table 14. Differential power test results on cognitive diagnostic assessment questions

Question Number	Discriminating Power Index	Information
1	0,500	Very Good
2	0,429	Very Good
3	0,643	Very Good
4	0, 643	Very Good
5	0, 643	Very Good
6	0,429	Very Good
7	0,286	Pretty Good
8	0, 286	Pretty Good
9	0, 286	Pretty Good
10	0,357	Good
11	0,214	Pretty Good
12	0,357	Good
13	0,286	Pretty Good
14	0,429	Very Good
15	0,357	Good
16	0,429	Very Good
17	0,571	Very Good
18	0,571	Very Good
19	0,500	Very Good
20	0,500	Very Good
21	0,214	Pretty Good
22	0,214	Pretty Good
23	0,214	Pretty Good
24	0,071	Not Good
25	0,143	Not Good

Based on Table 14, it can be seen that there are 14 cognitive diagnostic assessment questions in the "Very Good" category, three cognitive diagnostic assessment questions in the "Good" category, eight cognitive diagnostic assessment questions in the "Fairly Good" category, and two cognitive diagnostic assessment questions in the "Not Good" category.

Table 15. Different power test results for post-test questions

Question Number	Discriminating Power Index	Information
1	0,513	Very Good
2	0,503	Very Good
3	0,446	Very Good
4	0,446	Very Good
5	0,600	Very Good
6	0,313	Good
7	0,446	Very Good
8	0,267	Pretty Good
9	0,267	Pretty Good
10	0,313	Good
11	0,159	Not Good

Question Number	Discriminating Power Index	Information
12	0,426	Very Good
13	0,349	Good
14	0,272	Pretty Good
15	0,205	Pretty Good
16	0,282	Pretty Good
17	0,636	Very Good
18	0,503	Very Good
19	0,338	Good
20	0,338	Good
21	0,226	Pretty Good
22	0,205	Pretty Good
23	0,379	Good
24	0,267	Pretty Good
25	0,379	Good

Based on Table 15, it can be observed that nine post-test items fall into the "Very Good" category, seven items are categorized as "Good," eight items are in the "Quite Good" category, and one cognitive diagnostic assessment item is categorized as "Not Good." The difficulty level test in this study aims to categorize the questions as easy, medium, or difficult. The difficulty level analysis was conducted using Microsoft Office Excel 2013. The results of the difficulty level analysis are presented in Tables 16 and 17.

Table 16. Difficulty level test results for cognitive diagnostic assessment questions

Question Number	Difficulty Level Value	Criteria
1	0,536	Currently
2	0,571	Currently
3	0,679	Currently
4	0,679	Currently
5	0,679	Currently
6	0,786	Mudah
7	0,786	Easy
8	0,857	Easy
9	0,857	Easy
10	0,679	Currently
11	0,607	Currently
12	0,464	Currently
13	0,429	Currently
14	0,500	Currently
15	0,536	Currently
16	0,571	Currently
17	0,429	Currently
18	0,500	Currently
19	0,464	Currently
20	0,464	Currently
21	0,679	Currently
22	0,607	Currently
23	0,821	Easy
24	0,893	Easy

[e-ISSN 2598-5949 | p-ISSN 2303-1514]

Question Number	Difficulty Level Value	Criteria
25	0,929	Easy

Based on Table 16, it can be seen that the level of difficulty is represented by seven cognitive diagnostic assessment questions in the "Easy" category and 18 cognitive diagnostic assessment questions in the "Medium" category.

Table 17. Post-test question difficulty level test results

Question Number	Difficulty Level Value	Criteria
1	0,571	Currently
2	0,500	Currently
3	0,607	Currently
4	0,607	Currently
5	0,679	Currently
6	0,679	Currently
7	0,607	Currently
8	0,857	Easy
9	0,857	Easy
10	0,679	Currently
11	0,607	Currently
12	0,464	Currently
13	0,429	Currently
14	0,393	Currently
15	0,429	Currently
16	0,464	Currently
17	0,429	Currently
18	0,500	Currently
19	0,357	Currently
20	0,357	Currently
21	0,571	Currently
22	0,429	Currently
23	0,643	Currently
24	0,857	Easy
25	0,643	Currently

Based on Table 17, the difficulty level analysis shows that three post-test questions are categorized as "Easy," while 22 questions fall into the "Medium" category. The researchers continued the learning process by using student worksheets as a tool to enhance students' scientific literacy. The student worksheets were designed based on the analysis of students' cognitive abilities, with the learning material and activities tailored to the needs and mastery levels of each group. This approach allows the student worksheet to function not only as a learning aid but also as a strategy that encourages students to actively explore scientific concepts, connect them to real-life situations, and develop critical and analytical thinking skills. The goal is to provide a more focused, meaningful, and contextual learning experience, leading to gradual and continuous improvement in students' scientific literacy. To measure the impact, the researchers conducted a post-test to evaluate the extent to which students' understanding and skills in scientific literacy had improved after using the student worksheet.

The normality test was carried out to determine whether the data obtained came from a normally distributed population or not, so a normality test was carried out using the SPSS version 26 program. The normality test in this study used the one-sample Kolmogorov-Smirnov formula.

Table 18. Normality Test Results

Treatment Group	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Experimental Cognitive Diagnostics	0,152	28	0,094
Post Test Experiment	0,156	28	0,077
Control Cognitive Diagnostics	0,144	28	0,140
Post Test Control	0,140	28	0,172

Based on the presentation in Table 18, the significance values (p) for the experimental cognitive diagnostic and post-test are 0.094 and 0.077, respectively. For the control group, the significance values for the cognitive diagnostic and post-test are 0.140 and 0.172, respectively. Since the significance values are greater than α = 0.05, it can be concluded that the data distribution is normally distributed.

The homogeneity test, conducted to determine if two or more groups come from populations with the same variance, was performed using the SPSS 26 program. The homogeneity test in this study employed one-way ANOVA, and the results are presented in Table 19.

Table 19. Test results of the homogeneity of variances

Treatment	Mean	Information
Post Test Kelas Experiment _ Post Test Kelas Control	0,679	Homogen

Based on Table 19, it can be seen that the significance value (p) in the experimental class post-test and control class post-test obtained a value of 0.679. The significance value obtained is greater than 0.05, meaning the data is homogeneously distributed. The t-test aims to determine whether the influence of one independent variable is statistically significant on the dependent variable, assuming that the other independent variables are held constant. The homogeneity test results are presented in Tables 20 and 21.

Table 20. T Test Results

Treatment	Sig (2-tailed)	Information
Post Test Kelas Experiment _ Post Test Kelas Control	0,001	H1 Accepted

Table 21. Differences in T Test Results

Treatment	N	Mean
Post Test Kelas Experiment	28	90,75
Post Test Kelas Control	28	88,86

Based on the t-test data presented in Table 20, the results show that the t-test value for the experimental and control post-test scores is 0.001, which is less than the significance level of 0.05 (0.001 < 0.05). This indicates a significant difference between the two groups. Additionally, Table 21 shows that the average post-test score for the experimental class is 90.75, while the

[e-ISSN 2598-5949 | p-ISSN 2303-1514]

average score for the control class is 88.86. Therefore, it can be concluded that the null hypothesis (Ha) is rejected, and the alternative hypothesis (H1) is accepted, meaning there is a significant difference in the scientific literacy abilities between students in the experimental class and those in the control class.

4. Discussion

The results of this research include the development of cognitive diagnostic assessment instruments and cognitively differentiated student worksheets (LKPD) for Phase B science lessons, designed to measure and improve students' scientific literacy. The cognitive diagnostic assessment was designed to evaluate students' scientific literacy abilities. Elementary school teachers commonly use such assessments to understand students' needs within the independent curriculum framework. Assessments are integral to learning objectives, not only facilitating the learning process but also providing feedback to students. The results of these assessments allow teachers to reflect on and measure students' scientific literacy abilities, as well as evaluate the quality of future learning. Through diagnostic assessments, teachers can identify students' strengths and weaknesses in understanding science material, allowing them to design more effective and targeted learning strategies. Therefore, this assessment is a crucial step in building a solid foundation of scientific literacy for students (Nissa et al., 2024).

Cognitively differentiated student worksheets are practical teaching tools for enhancing students' scientific literacy because they are designed to tailor learning materials and activities to each student's cognitive level. This approach provides appropriate challenges for students of varying abilities, allowing them to learn more effectively. The student worksheets not only present information but also encourage students to observe, analyze, conclude, and apply scientific concepts to real-life contexts. By providing space for students to think critically and reflectively, differentiated student worksheets help build a strong conceptual understanding and scientific skills, which are central to scientific literacy (Roslina et al., 2025). According to research by Permata et al. (2017), diagnostic assessments have proven effective in enhancing problem-solving abilities (Hikmasari et al., 2018). The implementation of student worksheet media has also led to improvements in scientific literacy (Izzatunnisa et al., 2019). Studies by Ain & Mitarlis (2020) have confirmed its effectiveness in boosting scientific literacy. Additionally, Wong et al. (2021) suggest integrating literacy practices within problem-based learning (PBL), a method also recommended by Kasuga (2022) for science instruction.

The practicality test, which focuses on the ease of use of the diagnostic assessment instruments and LKPD, was conducted to ensure that these materials are user-friendly for educators and students. This includes clarity in understanding the guidelines, instructions, and the steps involved in using the LKPD in daily learning contexts. The practicality test results were analyzed using descriptive data analysis. The educator's practicality test yielded an average score of 84%, categorized as "very practical," while the student practicality test resulted in an average score of 87%, also categorized as "very practical."

Subsequently, data analysis was conducted to evaluate the effectiveness of the developed diagnostic assessment instruments and LKPD in measuring and improving students' scientific literacy. The researchers used a t-test for data analysis. Prior to this, normality and homogeneity tests were performed. The data normality test, conducted using the Kolmogorov-Smirnov formula with SPSS 26, indicated that the data were normally distributed, as the significance value was greater than α = 0.05. Following this, a homogeneity test was performed using one-way ANOVA with SPSS 26, revealing that the data were homogeneously distributed, as evidenced by the significance value (p) of 0.679, which was greater than 0.05. These results confirmed that the data met the assumptions for parametric testing, leading to the application of the t-test.

The t-test was conducted to determine whether there was a significant difference between the post-test scores of the experimental and control groups. The results of the t-test, performed using SPSS 26, showed a mean value of 0.001, which is less than the significance level of 0.05 (0.001 < 0.05). Additionally, the difference in the average post-test scores was 90.75 for the experimental class and 88.86 for the control class. These findings indicate that the diagnostic assessment instruments and LKPD were highly effective in improving student learning outcomes.

5. Conclusion and Implications

Based on the research results and discussions, the following conclusions can be drawn: 1) Cognitive diagnostic assessments and LKPD are valid for measuring and improving Scientific Literacy abilities in science and science learning Phase B. This is supported by the validation test results from material experts, which yielded an average percentage of 94% (very valid criteria); media experts, with an average of 94% (very valid criteria); language experts, with an average of 93% (very valid criteria); cognitive assessment question evaluation experts, with an average of 92% (very valid criteria); and post-test question experts, with an average of 94% (very valid criteria). 2) Cognitive diagnostic assessments and LKPD are practical for measuring and improving Scientific Literacy abilities in science and science learning, Phase B. This is confirmed by the practicality test results, where student responses had an average percentage score of 84% (considered very practical) and teacher responses had an average of 97% (also considered very practical). 3) Cognitive diagnostic assessments and LKPD are practical for measuring and improving Scientific Literacy abilities in science and science learning, Phase B. This is evident from the effectiveness test results, which used the t-test and showed a value of 0.001 < 0.05, indicating a significant difference between the post-test scores of the experimental and control classes.

Acknowledgement (if applicable)

Confession: The author would like to thank Dr. Pramudianti, M.Pd., Dr. Fatkhur Rohman, M.Pd.., and Prof. Dr. Invite Rosidin, M.Pd., who guided the author throughout the research process. Apart from that, we must also not forget our families and fellow strugglers who continue to fight, support, and encourage us.

References

Ain, Q., & Mitarlis, M. (2020). Pengembangan LKPD Berorientasi Inkuiri Terbimbing untuk Meningkatkan Literasi Sains pada Materi Faktorfaktor yang Mempengaruhi Laju Reaksi. *UNESA Journal of Chemical Education*, *9*(3), 397–406.
Ain, Q., & Mitarlis, M. (2020). Pengembangan LKPD Berorientasi Inkuiri Terbimbing untuk Meningkatkan Literasi Sains pada Materi Faktorfaktor yang Mempengaruhi Laju Reaksi. *UNESA Journal of Chemical Education*, *9*(3), 397–406.

Akib, E., & Muhsin, M. A. (2019). Assessment of teaching in the 21st century. *Journal of Physics: Conference Series*, *1179*(1), 12065.

Archer-Bradshaw, R. E. (2017). Teaching for Scientific Literacy? An Examination of Instructional Practices in Secondary Schools in Barbados. *Research in Science Education*, *47*(1).

Baird, J., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T., & Daugherty, R. (2011). *Policy effects of PISA*.

Basilio, M. B., & Bueno, D. C. (2021). Instructional Supervision and Assessment in the 21st Century and Beyond. *Online Submission*, *4*, 1–8.

Betari, M. E. (2020). Peningkatan kemampuan literasi sains siswa melalui penerapan model pembelajaran berbasis masalah pada pembelajaran IPA di SD. *Jurnal Pendidikan Dasar Flobamorata*, *1*(2), 1–17.

Bonawitz, E., Shafto, P., Gweon, STUDENT WORKSHEET., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330.

BouJaoude, S. (2002). Balance of Scientific Literacy Themes in Science Curricula: The Case of Lebanon. *International Journal of Science Education*, *24*(2).

Briseño-Garzón, A., Perry, K. STUDENT WORKSHEET., & Purcell-Gates, V. (2014). "To Learn About Science": Real Life Scientific Literacy Across Multicultural Communities. *Community Literacy Journal*, *8*(2), 81–107.

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*(3), 331–340.

Burkhardt, G., Monsour, M., Valdez, G., Gunn, C., Dawson, M., Lemke, C., Coughlin, E., Thadani, V., & Martin, C. (2003). Engage 21st Century Skills: Literacy in the digital age. In *North Central Regional Educational Laboratory and the Metiri Group* (Vol. 37).

Chen, J., Zhang, STUDENT WORKSHEET., Wei, STUDENT WORKSHEET., & Hu, J. (2021). Discrimination of the contextual features of top performers in scientific literacy using a machine learning approach. *Research in Science Education*, *51*, 129–158

Deguchi, A., Hirai, C., Matsuoka, STUDENT WORKSHEET., Nakano, T., Oshima, K., Tai, M., & Tani, S. (2020). Society 5.0: A people-centric super-smart society. *Society 5.0: A People-Centric Super-Smart Society*, 1–177.

Diallo, M., Tonn, B., Alvarez, P., Bardet, P., Chong, K., Feldman, D., Mahajan, R., Scott, N., Urban, R. G., & Yablonovitch, E. (2013). Implications: convergence of knowledge and technology for a sustainable society. *Convergence of Knowledge, Technology and Society: Beyond Convergence of Nano-Bio-Info-Cognitive Technologies*, 371–431.

Firmanzah, D., & Sudibyo, E. (2021). Implementasi Asesmen Diagnostik Dalam Pembelajaran Ipa Pada Masa Pandemi Covid-19 Di Smp/Mts Wilayah Menganti, Gresik. *Pensa E-Jurnal : Pendidikan Sains*, *9*(2).

Fukuda, K. (2020). Science, technology and innovation ecosystem transformation toward society 5.0. *International Journal of Production Economics*, *220*, 107460.

Hikmasari, P., Kartono, K., & Mariani, S. (2018). Analisis Hasil Asesmen Diagnostik dan Pengajaran Remedial pada Pencapaian Kemampuan Pemecahan Masalah Matematika melalui Model Problem Based Learning. *Prisma, Prosiding Seminar Nasional Matematika*, *1*.

Hikmasari, P., Kartono, K., & Mariani, S. (2018). Analisis Hasil Asesmen Diagnostik dan Pengajaran Remedial pada Pencapaian Kemampuan Pemecahan Masalah Matematika

[e-ISSN 2598-5949 | p-ISSN 2303-1514]

melalui Model Problem Based Learning. *Prisma, Prosiding Seminar Nasional Matematika*, *1*.

Hwang, J. (2019). Relationships among Locus of control, learned helplessness, and mathematical literacy in PISA 2012: focus on Korea and Finland. *Large-Scale Assessments in Education*, *7*(1). https://doi.org/10.1186/s40536-019-0072-7

Izzatunnisa, I., Andayani, STUDENT WORKSHEET., & Hakim, A. (2019). Pengembangan LKPD berbasis pembelajaran penemuan untuk meningkatkan kemampuan literasi sains peserta didik pada materi kimia SMA. *Jurnal Pijar Mipa*, *14*(2), 49–54.

Izzatunnisa, I., Andayani, STUDENT WORKSHEET., & Hakim, A. (2019). Pengembangan LKPD berbasis pembelajaran penemuan untuk meningkatkan kemampuan literasi sains peserta didik pada materi kimia SMA. *Jurnal Pijar Mipa*, *14*(2), 49–54.

Kallio, J. M., & Halverson, R. (2020). Distributed leadership for personalized learning. *Journal of Research on Technology in Education*, *52*(3), 371–390.

Kang, J. (2022). Interrelationship between inquiry-based learning and instructional quality in predicting science literacy. *Research in Science Education*, *52*(1), 339–355.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299.

Kasuga, STUDENT WORKSHEET., Maro, STUDENT WORKSHEET., & Pangani, I. (2022). Effect of Problem-Based Learning on Developing Science Process Skills and Learning Achievement on the Topic of Safety in Our Environment. *Journal of Turkish Science Education*, *19*(3), 872–886.

Lessy, D. (2022). Peningkatan Kemampuan Numerasi Bagi Guru Mi Melalui Bimtek Tindak Lanjut Hasil Asesmen Kompetensi Madrasah Indonesia. *Jurnal Pendidikan Dan Pengabdian Masyarakat*, *5*(1). https://doi.org/10.29303/jppm.v5i1.3323

Mellegård, I., & Pettersen, K. D. (2016). Teachers' response to curriculum change: balancing external and internal change forces. *Teacher Development*, *20*(2), 181–196.

Narvaez Rojas, C., Alomia Peñafiel, G. A., Loaiza Buitrago, D. F., & Tavera Romero, C. A. (2021). Society 5.0: A Japanese concept for a superintelligent society. *Sustainability*, *13*(12), 6567.

NILOA, N. I. for L. O. A. (2020). *Planning for Effective Communication of Assessment : A Toolkit for Practice*. *April*.

Nugraha, D., & Octavianah, D. (2020). Diskursus Literasi Abad 21 di Indonesia. *Jurnal Pendidikan Edutama*, *7*(1). https://doi.org/10.30734/jpe.v7i1.789

OECD. (2019). PISA Result In Focus. In *PISA 2009 at a Glance: Vol. I*.

Oliver, M., McConney, A., & Woods-McConney, A. (2021). The efficacy of inquiry-based instruction in science: A comparative analysis of six countries using PISA 2015. *Research in Science Education*, *51*, 595–616.

Pellegrino, J. STUDENT WORKSHEET. (2014). Assessment as a Positive Influence on 21st-Century Teaching and Learning: A Systems Approach to Progress. *Psicología Educativa*, *20*(2), 65–77.

Permata, J. I., Sukestiyarno, STUDENT WORKSHEET. L., & Hindarto, N. (2017). Analisis Representasi Matematis Ditinjau dari Kreativitas dalam Pembelajaran Cps dengan Asesmen Diagnostik. *Unnes Journal of Mathematics Education Research*, *6*(2).

Plomp, T., & Nieveen, N. (2013). Educational design research part A: An introduction. *Enchede: SLO*.

Pratiwi, I. (2019). Efek Program Pisa Terhadap Kurikulum Di Indonesia. *Jurnal Pendidikan Dan Kebudayaan*, *4*(1), 51–71.

Rachmawati, A., & ... (2022). Penerapan Model Pembelajaran Inovatif Melalui Asesmen Diagnostik Dalam Menguatkan Literasi Anak Kelas 1 di SDN Banjaran 5. *Prosiding SEMDIKJAR*

Rosana, D., Widodo, E., Setianingsih, STUDENT WORKSHEET., & Setyawarno, D. (2020). Developing Assessment Instruments of the PISA Model to Measure Students' Problem-Solving Skills and Scientific Literacy in Junior High Schools. *Jurnal Pendidikan Sains Indonesia*, *8*(2). https://doi.org/10.24815/jpsi.v8i2.17468

Sari, A. dkk. (2020). Penghapusan Ujian Nasional Tahun 2021 dalam Perspektif Guru SMA di Kota Tebing Tinggi. *Prosiding Seminar Nasional PBSI-III*.

Suhandi, A. M., & Robi'ah, F. (2022). Guru dan tantangan kurikulum baru: Analisis peran guru dalam kebijakan kurikulum baru. *Jurnal Basicedu*, *6*(4), 5936–5945.

Sułkowski, Ł., Kolasińska-Morawska, K., Seliga, R., & Morawski, P. (2021). Smart Learning Technology in the Economy 5.0—the Polish Perspective. *Applied Sciences*, *11*(11), 5261.

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. Ascd.

Trilling and Fadel. (2009). 21st-century skills: Learning for life in our times. Jossey Bass: USA. *Journal of Chemical information and Modeling*, *53*(9).

Turiman, P., Omar, J., Mohd Daud, A., & Osman, K. (2012). Fostering the 21st Century Skills through Scientific Literacy and Science Process Skills. *Procedia-Social and Behavioral Sciences*, *59*, 110–116. https://doi.org/10.1016/j.sbspro.2012.09.253

Wagner, T. (2010). *The global achievement gap: Why even our best schools do not teach the new survival skills our children need-and what we can do about it*—readHowYouWant.com. com.

Wagner, T., & Dintersmith, T. (2015). *Most likely to succeed: Preparing our kids for the innovation era*. Simon & Schuster.

Wibowo, A. M. (2022). Pergeseran Paradigma Pembelajaran: Analisis Dampak Penerapan Asesmen Nasional Dalam Bingkai Teori Kognitif Sosial. *Madrasah: Jurnal Pendidikan Dan Pembelajaran Dasar*, *14*(2), 137–152. https://doi.org/10.18860/mad.v14i2.16023

Wong, S. S. STUDENT WORKSHEET., Kim, M., & Jin, Q. (2021). Critical literacy practices within problem-based learning projects in science. *Interchange*, *52*(4), 463–477.

Zhao, STUDENT WORKSHEET. (2012). Flunking innovation and creativity. *Phi Delta Kappan*, *94*(1), 56–61.

Zhao, STUDENT WORKSHEET. (2018). *Reach for greatness: Personalizable Education for all children*. Corwin Press.

Zhao, STUDENT WORKSHEET., & Watterston, J. (2021). The Changes We Need: Education Post-COVID-19. *Journal of Educational Change*, *22*(1), 3–12.