



Journal of Educational Sciences

Journal homepage: <https://jes.ejournal.unri.ac.id/index.php/JES>



P-ISSN
2581-1657

E-ISSN
2581-2203

Performance-Based Assessment to Measure Narrative Reading Literacy Skills of Fourth-Grade Students at SDN Petunggu

Ratnawati*, Ida Nuraeni, Juniati, Gazali, Sukma

Master of Indonesian Language Education, University of Tadulako, Palu City, 94111, Indonesia

ARTICLE INFO

Article history:

Received: 07 Dec 2025

Accepted: 28 April 2026

Available Online: 05 May 2026

Keywords:

Performance Assessment

Reading Literacy

Narrative Text

* Corresponding author:

E-mail: ratnawatis274@gmail.com

Article Doi:

<https://doi.org/10.31258/jes.10.5.p.494-510>

This is an open access article under the [CC BY-
SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



ABSTRACT

This study aims to develop a valid, practical, and effective performance assessment instrument to measure fourth-grade students' reading literacy skills at SDN Petunggu. The study is motivated by students' low reading literacy and the limitations of conventional assessments that mainly measure surface-level understanding without capturing students' thinking processes authentically. This research employed a Research and Development (R&D) approach by adapting the Borg & Gall model, which was modified into several stages: preliminary study, needs analysis, curriculum analysis, product planning, instrument draft development, expert validation, limited trial, and field trial. The developed instrument includes fable texts, performance tasks, answer sheets, and analytic scoring rubrics designed to assess students' comprehension of narrative texts. Expert validation results indicated that the instrument was categorized as valid. Field trials involving 18 students showed a balanced score distribution, indicating the instrument's ability to differentiate students' literacy levels. The majority of students were categorized as Good and Very Good, with no indication of floor or ceiling effects. These findings demonstrate that the developed performance assessment effectively captures students' narrative reading comprehension in a more comprehensive and authentic manner. Therefore, the instrument is considered appropriate as an alternative assessment tool aligned with the demands of the Independent Curriculum to improve the quality of reading literacy evaluation in elementary schools.

1. Introduction

Reading literacy is a fundamental cornerstone of primary education, particularly for fourth-grade students. At this stage, learners transition from "learning to read" to "reading to learn." Early literacy skills strongly predict future academic success (Lesnick et al., 2010). Literacy is not merely a mechanical ability to decode text; it encompasses comprehension, interpretation, and the evaluation of meaning within

written material. However, despite its importance, current conditions reveal significant challenges. Data from the Programme for International Student Assessment (PISA) 2022 indicate that Indonesian students' reading proficiency remains low, ranking 72nd out of 81 countries, with an average score of 359—substantially below the OECD average of 476 (OECD, 2023). This gap highlights a critical deficiency in foundational literacy skills within Indonesia's primary education system.

One of the key factors contributing to low reading literacy, particularly in narrative texts, is the lack of comprehensive assessment approaches. Assessment instruments in primary schools tend to rely heavily on multiple-choice or short-answer formats that measure only surface-level cognitive skills (Wahdania, 2024). Previous studies indicate that conventional assessment practices often fail to capture students' higher-order and reflective thinking processes in reading activities, as many teachers still rely on limited and non-authentic assessment approaches (Ekoati et al., 2021). As a result, students' deeper comprehension and critical engagement with texts are not adequately assessed. Therefore, there is an urgent need to develop more holistic and comprehensive performance-based assessments.

Performance-based assessment aligns closely with the principles of the “Merdeka Belajar” (Independent Learning) curriculum, which emphasizes formative, authentic, and holistic evaluation. In this context, assessment is no longer viewed merely as a tool for measuring outcomes but as an integral part of the learning process (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2022). In line with this perspective, performance-based assessment provides a clearer representation of students' reading literacy skills and supports teachers in designing more effective instruction (Heydarnejad et al., 2022). The Indonesian Ministry of Education and Culture (Kemendikbud, 2022) further highlights that performance assessment, as a form of authentic assessment, enables students to demonstrate their understanding through real-world tasks such as summarizing texts, responding to content, and relating stories to everyday experiences. This perspective is further supported by empirical studies, which indicate that authentic assessment allows students to demonstrate their understanding in meaningful contexts and provides continuous formative feedback to improve learning outcomes (Puteri et al., 2023). Consequently, this approach allows students to express their understanding in a deeper and more creative manner (Darling-Hammond et al., 2020). Nevertheless, in Pasangkayu, the implementation of performance-based assessment remains suboptimal, particularly at the primary school level.

Preliminary observations in several elementary schools in Pasangkayu District, including SDN Petunggu, reveal that most teachers lack systematic and structured performance assessment instruments. As a result, reading assessments are still largely limited to literal comprehension, while interpretative and evaluative aspects are not fully addressed. In addition, assessment practices tend to focus on final outcomes rather than the students' cognitive processes during reading. This situation indicates the need to support teachers through the development of assessment instruments that are relevant to students' literacy needs. Furthermore, the narrative reading literacy level of fourth-grade students at SDN Petunggu

remains relatively low, reinforcing the urgency of developing performance-based assessments as part of meaningful learning.

Further insights from interviews with three teachers who have taught fourth grade at SDN Petunggu indicate that teachers often face difficulties in determining objective assessment indicators that align with the cognitive characteristics of 9–10-year-old students when measuring narrative reading literacy. This challenge is closely related to the lack of accurate assessment instruments, resulting in less meaningful measurement, as traditional approaches often fail to capture students' cognitive processes and deeper understanding (Pellegrino et al., 2016). In addition to assessment issues, contextual factors also play a role. Reading materials available in the school library are generally formal and lack contextual relevance, and textbooks often do not align with students' real-life experiences. This mismatch reduces students' interest in reading narrative texts, causing them to read out of obligation rather than curiosity or enjoyment (Sukma & Amurdawati, 2020), which in turn leads to weak comprehension of plot, characters, and moral messages in narrative texts. Although Putri (2024) found that performance-based assessment is frequently applied in science and mathematics contexts, it has not received sufficient attention in Indonesian language learning, particularly in relation to narrative texts. This imbalance underscores the need for assessment instruments specifically designed to evaluate students' thinking processes and deep comprehension of narrative reading, in line with curriculum demands for authentic assessment.

Previous studies have explored reading literacy and the development of performance-based assessments. Anggraeni and Mukhlis (2023), in their study on minimum competency assessment of reading literacy at SDN 09 Merangkai, concluded that literacy instruction still requires improvement and recommended the development of performance-based assessments. Amri (2020) examined the impact of performance assessment implementation on character and science literacy among fourth-grade students at MIN 2 Konawe Selatan, finding significant differences. Rahmawati et al. (2020) developed assessment instruments and rubrics to measure communication skills among fourth-grade students. However, these studies primarily focus on science literacy and communication skills rather than narrative reading literacy. This indicates a significant research gap in the development of performance-based assessment instruments specifically targeting narrative reading literacy for fourth-grade elementary students.

In addition to these issues, recent studies have emphasized the importance of structured and contextual approaches in literacy development. Sugiarsih et al. (2026) highlight that balanced literacy models play a crucial role in supporting students' comprehension processes. Furthermore, the integration of technology-based literacy programs has been shown to significantly improve students' reading literacy outcomes (Wato & Mustadi, 2026). These findings reinforce the need for more comprehensive and process-oriented assessment practices in primary education.

Based on the above discussion, there remains substantial opportunity for research integrating narrative literacy and performance-based assessment within the context of Indonesian language learning at the primary school level. The absence of validated and ready-to-use instruments represents a tangible gap in both educational practice and research. Sudianto and Kisno (2021) assert that effective assessment should encourage students to reflect on their understanding rather than merely answer questions. Therefore, the development of contextualized and standardized performance-based assessment is essential. The instrument should measure three key dimensions of reading literacy as defined by OECD (2019): locating information, interpreting content, and reflecting on or evaluating the message within a text, while also being tailored to the cognitive, linguistic, and instructional characteristics of fourth-grade students. Accordingly, this study aims to develop a performance-based assessment instrument for measuring fourth-grade students' narrative reading literacy that is valid, practical, and capable of capturing students' cognitive processes in a comprehensive and authentic manner.

2. Methodology

This study employs a Research and Development (R&D) approach aimed at designing, developing, and producing a valid, practical, and effective performance-based assessment instrument for reading literacy to be used by elementary school students. The R&D approach is selected because it not only focuses on testing theoretical constructs but also emphasizes the development of educational products that can be practically implemented in the learning process. The research methodology is systematically designed to ensure that each stage of instrument development is structured, measurable, and scientifically accountable. Therefore, the Borg and Gall model is modified to better align with the context of primary education, making it more operational and responsive to field-based needs.

The product developed in this study is a performance-based assessment instrument designed to measure students' narrative reading literacy authentically through tasks that reflect real reading processes and comprehension in meaningful contexts. The development model refers to the framework proposed by Gall et al. (2003), which emphasizes the integration of theoretical review, expert validation, and field testing in the process of product development and refinement. The stages of development implemented in this study include:

1. Identification of potential and problems to determine the need for reading literacy assessment in primary schools;
 2. Information gathering and planning through literature review and curriculum analysis;
 3. Development of the initial product draft, including the preparation of assessment blueprints, task items, and scoring rubrics;
 4. Expert validation to evaluate the content, language, and construct validity of the instrument;
 5. Limited field testing and product revision to obtain empirical data and refine the instrument based on field findings; and
-

6. Product effectiveness testing to determine the extent to which the performance-based assessment can measure and improve students' reading literacy skills. Through this approach, the resulting product is expected to possess a strong theoretical foundation, empirically validated reliability, and high practicality for implementation in primary education settings.

Research Design (Modified Borg and Gall Model)

The Borg and Gall model in Gall et al. (2003) is one of the most widely used research and development frameworks in the field of education. It focuses on the development of educational products through a sequence of validation, revision, and empirical evaluation stages before large-scale implementation. In this study, the model is adapted into nine general stages and further operationalized into eleven steps to better suit field conditions and time constraints. To provide a clearer understanding of the development process employed in this study, the adapted stages of the Borg and Gall R&D model can be seen in Figure 1.

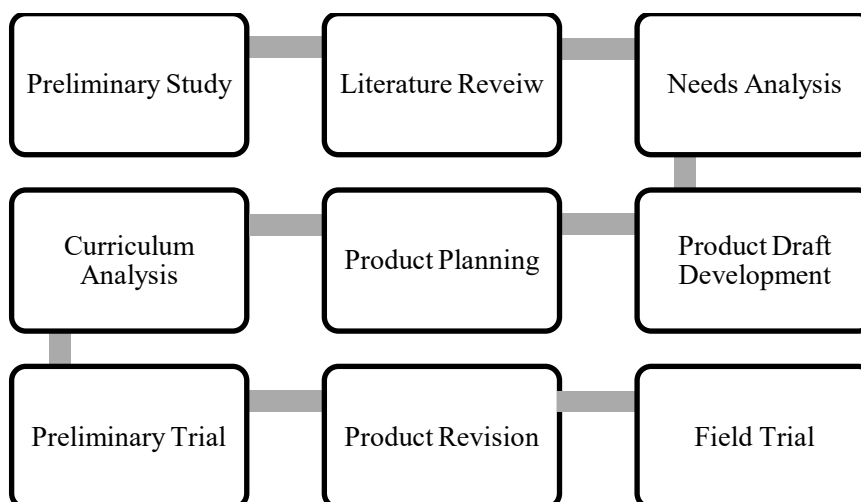


Figure 1. Stages of the Modified Borg and Gall R&D Model

In this study, the Research and Development (R&D) model referring to Borg and Gall is employed due to its alignment with the research objective, namely developing an educational product that is both theoretically sound and empirically tested. Considering the time limitations and the primary school research context, the model is further simplified into eight main stages that are more concise and relevant to the development of performance-based assessment instruments.

3. Results and Discussion

Preliminary Study

The preliminary study in this research was conducted through a series of initial observations and in-depth interviews with fourth-grade teachers at SDN Petunggu. This stage aimed to obtain a comprehensive understanding of the actual conditions

of reading assessment practices in the field, so that the direction of instrument development is not only grounded in theory but also rooted in authentic empirical needs. This approach aligns with the principles of Research and Development (R&D), which place needs assessment as a crucial foundation before a product is designed and tested.

The results of the preliminary study indicate that reading literacy assessment practices in schools are still dominated by traditional test formats, such as multiple-choice questions and short-answer items. These conventional tests generally measure only lower-order literacy skills, particularly in recalling factual information and answering literal questions. Teachers stated that such assessment forms are chosen because they are more practical, easier to score, and have become a habitual approach in both daily and summative evaluations.

However, from a pedagogical perspective, such assessment forms have significant epistemological limitations. Conventional tests are not designed to explore the complex thinking processes that occur when students read and comprehend narrative texts. As a result, various higher-order literacy competencies are not adequately measured. Some competencies that cannot be captured through conventional tests include:

1. The ability to deeply identify intrinsic elements, such as character development, plot dynamics, and setting details.
2. The ability to understand cause–effect relationships, both explicit and implicit within the story structure.
3. The ability to make inferences by integrating explicit information and implicit meaning.
4. The ability to evaluate moral values, including assessing characters' actions, providing justification, and relating them to real-life contexts.

Teachers also revealed that conventional assessment results often do not reflect students' actual understanding. Students who can interpret stories well through oral discussion sometimes receive low scores due to incorrect multiple-choice answers. Conversely, students who guess answers may obtain high scores without demonstrating deep understanding. This indicates that assessment fails to capture the quality of thinking processes and instead focuses only on final outputs. In addition, observations showed a wide disparity in reading abilities among students. Some students demonstrate strong literal comprehension but struggle to produce interpretive conclusions, while others can explain moral meanings but have difficulty reconstructing the sequence of events chronologically. This variation requires an instrument with high sensitivity and strong discriminatory power to capture diverse student performance.

These findings clearly indicate that static, pseudo-objective, and multiple-choice-based assessments are no longer adequate within the context of the Merdeka Curriculum. This curriculum emphasizes the development of literacy competencies authentically through activities that encourage analysis, reflection, and meaning construction, thus requiring assessment forms that align with these demands.

Therefore, this preliminary study provides a strong scientific basis for the need to develop performance-based assessment instruments that are able to:

1. capture students' thinking processes authentically,
2. assess higher-order reading skills,
3. provide a comprehensive profile of literacy ability, and
4. encourage students to interpret texts critically and reflectively.

The findings at this stage serve as a fundamental basis for designing the initial product of a reading literacy performance-based assessment instrument that is not only curriculum-relevant but also functional in everyday classroom practice.

Literature Review

The literature review was conducted to establish a theoretical foundation related to three main aspects: reading literacy, performance-based assessment, and Indonesian language learning focused on narrative texts in fourth-grade elementary school. The findings indicate that reading literacy is not limited to word recognition but also includes the ability to comprehend content, draw conclusions, identify story elements, and provide critical responses. Theories of reading literacy proposed by Rosenblatt (1978) and the framework from Wiggins and McTighe (2005), along with guidelines from Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (2021), serve as key references. These perspectives emphasize that meaning comprehension can be measured through indicators such as identifying main ideas, summarizing information, connecting experiences with texts, and evaluating the effectiveness of story elements. In addition, recent developments in literacy learning indicate that the use of interactive and digital-based learning media can support students' engagement and comprehension more effectively. Imran et al. (2025) found that e-module-based learning can improve students' reading literacy through more contextual and meaningful learning experiences. Furthermore, Pujiningsih et al. (2026) emphasize that literacy programs require systematic evaluation to ensure their effectiveness in improving student learning outcomes, particularly in primary education contexts.

Initial observations and in-depth interviews with fourth-grade teachers at SDN Petunggu show that reading assessment practices remain limited to conventional tests, such as multiple-choice and short-answer formats. These instruments generally measure factual knowledge and recall abilities, thus failing to represent reading literacy comprehensively. Teachers acknowledged that such assessments are easier to prepare and score but do not provide deep insight into students' thinking processes when understanding texts. The limitations of conventional assessments result in the inability to measure higher-order literacy skills, such as identifying intrinsic elements of narrative texts (characters, setting, plot), understanding cause-effect relationships, reconstructing story sequences, interpreting meaning, and drawing moral conclusions. These competencies are essential in reading literacy as emphasized in the Merdeka Curriculum and international standards such as PISA. Furthermore, teachers stated that assessments focusing only on final answers cannot capture students' thinking processes. Students who understand texts but struggle with answer selection often receive low scores, while those who guess correctly may achieve high scores without genuine

understanding. This highlights the need for assessments that evaluate not only outcomes but also cognitive processes during reading.

Teachers also reported wide variations in students' reading abilities. Some students can explain stories orally but struggle to write them, while others can answer literal questions but fail to interpret implicit meanings. This condition requires assessment instruments capable of capturing students' abilities more comprehensively rather than focusing solely on recall. Based on these findings, conventional assessment is insufficient to meet literacy learning needs in fourth grade. Teachers require performance-based assessment instruments that are authentic, contextually relevant, and capable of accurately assessing higher-order thinking skills. Performance-based assessment allows teachers to evaluate how students interpret texts, organize information, construct meaning, and connect moral messages with personal experiences. Thus, the need for developing performance-based assessment instruments emerges as a response to curriculum demands, teacher needs, and real classroom challenges. This needs analysis serves as the primary foundation for designing an initial product that is more representative, authentic, and aligned with the characteristics of elementary school students.

Needs Analysis

The needs analysis was conducted through teacher interviews, classroom observations, and questionnaire administration. The overall findings indicate that fourth-grade elementary school teachers face several challenges in assessing students' reading literacy skills. First, teachers still predominantly rely on multiple-choice written tests, even though the current literacy curriculum emphasizes deep comprehension skills. Teachers expressed the need for authentic assessment instruments that can more naturally reflect students' abilities in reading narrative texts.

Second, students tend to experience difficulties in understanding story elements such as characters, plot, and moral messages when reading narrative texts. Without appropriate assessment instruments, teachers are unable to accurately identify students' levels of difficulty. Third, teachers reported that they do not yet have clear scoring rubrics to assess narrative reading performance. As a result, assessments are often conducted subjectively and are not systematically documented.

From the students' perspective, observational results show that they find it easier to understand reading materials when they are assigned performance-based tasks such as summarizing texts, illustrating plot sequences, or retelling stories. This finding highlights the need for performance-based assessment instruments that can directly capture students' thinking processes. Overall, the results of the needs analysis indicate the necessity of developing assessment instruments that are practical, valid, and aligned with the characteristics of Indonesian language learning in elementary schools.

The needs analysis was also conducted to identify gaps between existing reading literacy assessment practices at SDN Petunggu and the demands of the curriculum,

student characteristics, and principles of authentic assessment. This analysis was derived from classroom observations, assessment document analysis, and in-depth interviews with fourth-grade teachers. The findings were categorized into five main aspects that provide a strong foundation for developing performance-based reading literacy assessment instruments.

1. The limitations of conventional assessments in evaluating reading literacy skills were identified as the most fundamental issue. Teachers continue to use multiple-choice and short-answer formats that primarily measure surface-level skills, such as recalling literal information. In contrast, reading literacy in the Merdeka Curriculum and PISA standards requires more complex abilities, including interpreting cause-effect relationships, making inferences from implicit information, and evaluating moral values within texts. Conventional formats are unable to capture these processes authentically, resulting in assessments that do not reflect students' true understanding.
 2. Teachers require assessment instruments that are more accurate, standardized, and practical to implement. Based on interviews, teachers reported difficulties in formulating objective indicators aligned with Phase B learning outcomes, developing consistent scoring rubrics, selecting reading texts appropriate to students' cognitive development, and conducting assessments without excessive time demands. Teachers also emphasized that right-wrong scoring systems tend to misrepresent actual abilities, as students who understand the material may answer incorrectly, while those who guess may achieve high scores. This highlights the need for performance-based instruments with analytical rubrics.
 3. The wide variation in students' reading abilities requires assessment instruments that are sensitive in distinguishing performance levels. Observations show that some students can explain texts orally but struggle in writing; others recall literal information but fail to grasp implicit meanings; while some can sequence events but cannot draw conclusions. These differences require assessments that not only assign scores but also capture students' thinking processes, such as interpreting plot structures, organizing information, and evaluating moral content.
 4. Assessment practices need to be aligned with the Merdeka Curriculum and the principles of Merdeka Belajar, which emphasize holistic, authentic, process-oriented, and contextually relevant evaluation. Current practices at SDN Petunggu remain focused on final outcomes without assessing the depth of students' comprehension. The curriculum emphasizes authentic tasks to measure literacy skills in depth, making performance-based assessment development both relevant and necessary.
 5. The contextual conditions of the school indicate limitations in the availability of standardized and relevant assessment instruments, both in terms of reading texts and scoring rubrics. School documentation shows that no instruments have been developed through rigorous procedures with clear validity and reliability, nor are they tailored to fourth-grade students. Reading texts tend to be formal and less contextual, reducing students' reading interest. Teachers also lack operational
-

analytical rubrics. These conditions strengthen the urgency of developing performance-based assessment instruments that are practical, relevant, and capable of comprehensively measuring reading performance.

Overall, this needs analysis provides a strong empirical and theoretical foundation for developing performance-based reading literacy assessment instruments to improve the quality of learning evaluation and provide an authentic representation of students' literacy skills at SDN Petunggu.

Curriculum Analysis

The curriculum analysis was conducted by referring to the Learning Outcomes of the Merdeka Curriculum at Phase B (Grades 3–4), where students are expected to understand, evaluate, and respond to the content of narrative texts both orally and in writing, demonstrating appropriate reading literacy skills. The analysis focused on the competency of reading literacy in narrative texts. The results of the analysis indicate that the competencies that fourth-grade students must master include:

1. Understanding information contained in narrative texts, both explicit and implicit.
2. Identifying intrinsic elements such as characters, setting, plot, conflict, and moral message.
3. Interpreting the meaning of the text based on students' personal experiences.
4. Retelling the content of the story in oral, written, or visual forms.

The researcher also aligned the analysis with the Merdeka Curriculum, which emphasizes a deep learning approach and authentic assessment, aimed at capturing students' abilities in real-life contexts. Based on the curriculum mapping, this served as a reference for the researcher in developing performance-based assessment instruments that focus on the following competencies:

1. The ability to comprehend reading content.
2. The ability to draw conclusions.
3. The ability to identify story elements.
4. The ability to re-express the meaning of the story through performance tasks.

Thus, the researcher concludes that the developed performance assessment will be aligned with curriculum demands and assist teachers in obtaining a comprehensive picture of students' reading literacy development.

Product Planning

Product planning was carried out as a strategic stage to design the components of the performance assessment instrument to be developed. This stage was structured based on the results of the needs analysis, curriculum analysis, and a review of theories related to authentic assessment, reading literacy, and the characteristics of fourth-grade students. At this stage, a comprehensive design was formulated, covering the content of the instrument, task formats, assessment indicators, and the rubric to be used, ensuring that the initial product is aligned with the competencies of the Merdeka Curriculum and the research objectives. This product planning includes four core components as follows:

1. Main Text, used in the instrument is planned to take the form of a fable narrative text. The selection of fables is based on their linear plot structure, context that is close to children's experiences, and the presence of both explicit and implicit moral values, making them ideal for fourth-grade students. From a cognitive development perspective, students at the concrete operational stage (ages 10–11) are able to understand simple cause-and-effect relationships, follow plot structures, and identify characters in narrative texts. Fables also provide opportunities to assess literal, inferential, and evaluative comprehension comprehensively. Therefore, the selection of fables as the main text is grounded in theoretical, pedagogical, and curricular alignment.

2. Performance Tasks, Based on the characteristics of authentic assessment, students are not only required to answer closed-ended questions but also to demonstrate their thinking processes through performance-based tasks. At the planning stage, it was determined that students would read a fable text and then complete a series of tasks designed to measure overall literacy skills. These tasks include:

1. Identifying explicit information stated in the text,
2. Identifying characters, traits, and settings (time and place),
3. Reconstructing the sequence of events logically,
4. Drawing conclusions based on explicit and implicit information,
5. Evaluating the moral message and reflecting on the characters' actions.

Each task is designed to assess students' cognitive processes, not merely the final outcomes. Thus, performance tasks provide an authentic picture of how students understand, interpret, and analyze narrative texts. These tasks form the basis for developing instruments that are ready for pilot testing.

3. Assessment Indicators, the planning of assessment indicators was based on the reading literacy competencies outlined in the Merdeka Curriculum for Phase B. The five core indicators were mapped from the learning outcomes, narrative literacy skills, and higher-order cognitive demands.

4. Analytical Scoring Rubric, the rubric is designed as an analytical rubric with a scale of 1–4, containing performance descriptions for each indicator. The use of an analytical rubric aims to ensure that the assessment of student performance is conducted objectively, systematically, and consistently. Each score level includes descriptors that represent levels of student mastery, ranging from unable, less capable, fairly capable, to highly capable. At the planning stage, the rubric was developed by considering (1) the measurability of indicators, (2) the clarity of performance descriptions, (3) the use of teacher-friendly language, and (4) alignment with the characteristics of performance tasks. With the use of an analytical rubric, the assessment process becomes more transparent and capable of capturing detailed dimensions of students' abilities.

This product planning serves as the foundation for developing the initial draft of the reading literacy performance assessment instrument, which will be validated by experts. All components are designed to be interconnected, resulting in an

instrument that is relevant, authentic, and capable of providing a comprehensive picture of fourth-grade students' reading literacy skills.

Product Draft Development

At this stage, the performance assessment will be revised based on feedback from expert validators, which will later be used in the trial phase. This stage consists of several activities, including distributing validation sheets to three expert validators. These validators include a postgraduate lecturer from the Indonesian Language Education program at Tadulako University, a school supervisor in Pasangkayu Regency, and a lecturer from the Faculty of Teacher Training and Education (FKIP) at Tomakaka with expertise in different fields.

The product draft development stage is a continuation after product planning and the mapping of assessment indicators. At this stage, all previously formulated components are compiled into a complete, integrated Draft Product 1, which is ready for expert validation. This draft is developed based on the results of the needs analysis, curriculum analysis, reading literacy theories, and principles of authentic assessment relevant to the characteristics of fourth-grade students. Draft Product 1 consists of four main components: an adapted fable text, performance task instructions, student answer sheets, and an analytical scoring rubric.

The first component is the fable text, which has been adjusted to match the readability level of fourth-grade students. This adjustment process considers sentence structure, paragraph length, vocabulary complexity, and alignment with students' cognitive development at the concrete operational stage. The text is selected based on curricular and pedagogical criteria, including a clear plot, explicit and implicit moral messages, and characters that can be analyzed by students. This adjustment ensures that the text remains challenging without exceeding students' capacity to comprehend the story as a whole.

The second component is the performance-based task instructions, designed to guide students in completing a series of reading and text analysis activities. These instructions are written clearly, systematically, and in communicative language to ensure they are easily understood by students. Each instruction aligns with the five previously mapped reading literacy indicators, ensuring that the tasks effectively measure students' ability to identify information, recognize intrinsic elements, reconstruct plot sequences, make inferences, and evaluate moral messages. The task instructions are structured to guide students' thinking processes progressively, from literal to evaluative levels.

The third component is the student answer sheet, which serves as a medium for students to record their responses to the performance tasks. The answer sheet is designed to help students express their understanding in a coherent and systematic manner. Each section corresponds to specific assessment indicators, ensuring that students' responses align with the task structure and facilitating teachers in conducting the assessment. This component also supports authentic assessment by

providing space for students to demonstrate their reasoning processes, not just final answers.

The fourth component is a complete analytical rubric for the five assessment indicators. The rubric is designed using a 1–4 scale, with operational, clear, and teacher-friendly performance descriptors. The analytical rubric enables objective and consistent assessment, as each level of achievement describes specific characteristics of student responses. The rubric development considers the principle of content validity, ensuring that each descriptor accurately represents the measured indicator. It also ensures that the assessment results provide a comprehensive picture of students' reading literacy skills, from literal to interpretative and evaluative aspects.

Thus, the development of Draft Product 1 results in a structured, systematic performance assessment instrument aligned with the research objectives. This draft is now ready to proceed to the expert validation stage to obtain theoretical and practical feedback before being tested in schools. This validation stage is essential to ensure that the instrument demonstrates strong academic, pedagogical, and operational feasibility.

Preliminary Trial

The instrument developed during the design stage was subsequently validated by two subject-matter experts and one educational evaluation expert. The involvement of experts from these two domains was intentionally selected because a reading literacy performance assessment instrument requires not only accurate content representation but also sound assessment construction, alignment of indicators, language clarity, and the accuracy of the scoring rubric used to measure student performance. The validity technique employed in this study is content validity using the Aiken's V index. The selection of Aiken's V is based on the characteristics of the instrument, which uses a rating scale and is assessed by more than two validators. The Aiken's V results for each item are presented in Table 1.

Table 1. Calculation of Aiken's V

Item	Validator 1	Validator 2	Validator 3	r - l _o	Σs	V
P1	4	5	4	3,4,3	10	0.83
P2	3	4	4	2,3,3	8	0.67

The Aiken's V index is calculated using the following formula:

$$V = \frac{10}{3 \times (5-1)} = 0,83$$

The validation process was conducted to ensure that each component of the instrument aligns with the assessment objectives, does not create ambiguity, and is truly capable of measuring reading literacy skills according to the established indicators. This validation is crucial, as performance-based assessment instruments are inherently complex and may lead to multiple interpretations if their components are not formulated clearly and operationally. Therefore, feedback from experts

serves as a scientific basis for determining the feasibility of the instrument before it is implemented with students.

Field Trial

The field trial was conducted with 18 fourth-grade students as a continuation of the limited trial phase. The purpose of this field trial was to examine the effectiveness of the instrument in an authentic learning context, to observe a wider variation of students' abilities, and to ensure that the scoring rubric functions consistently when applied to a larger group of participants. In addition, this stage was intended to evaluate the practicality and sensitivity of the instrument in distinguishing different levels of reading literacy performance among students. The results of the field trial are presented in Table 2.

Table 2. Field Trial Results

Score Range	Category	Frequency	Percentage
16–20	Very Good	6	33.3%
11–15	Good	8	44.4%
6–10	Fair	3	16.7%
1–5	Poor	1	5.6%

The results of the field trial show that most students fall into the Good category (44.4%), followed by the Very Good category (33.3%). This finding indicates that the instrument is capable of appropriately representing students' reading literacy abilities and producing a proportional score distribution. The spread of students across different categories, rather than being concentrated in a single level, demonstrates that the instrument has a high level of sensitivity in distinguishing students' performance across varying ability levels. The instrument does not produce a ceiling effect (where most scores cluster in the highest category) or a floor effect (where scores cluster in the lowest category).

Thus, the instrument is able to measure students' abilities more accurately. The presence of students in the Fair (16.7%) and Poor (5.6%) categories also indicates that the tasks are not overly easy, allowing the instrument to assess higher-order thinking skills such as drawing conclusions from reading passages, evaluating moral messages, and identifying relationships among intrinsic elements of the text. This strengthens the conclusion that the instrument measures not only surface-level understanding but also more complex cognitive skills. These findings are in line with previous studies indicating that literacy development requires structured and contextual learning support to enhance students' comprehension abilities (Sugiarsih et al., 2026). In addition, the integration of innovative learning media has been shown to significantly improve students' reading literacy performance (Wato & Mustadi, 2026; Imran et al., 2025). However, several studies also reveal that literacy programs in elementary schools still face challenges, particularly in developing students' higher-order thinking skills and deep comprehension (Pujiningsih et al., 2026). This is further supported by Hidayat et al. (2025), who found that students often experience difficulties not only in understanding texts but also in explaining

their reasoning processes. Therefore, performance-based assessment becomes essential to capture students' cognitive processes more comprehensively.

4. Conclusion

This study produced a performance-based assessment instrument that was systematically designed based on field needs, curriculum demands, and the cognitive developmental characteristics of fourth-grade elementary school students. The instrument consists of a fable text, a series of performance tasks, student answer sheets, and an analytical scoring rubric that enables teachers to assess students' thinking processes more comprehensively. The results of expert validation indicate that the instrument demonstrates strong content validity, while the field trial results confirm that it is capable of measuring reading literacy authentically and clearly distinguishing students' ability profiles. The findings of this study affirm that performance-based assessment is superior to conventional assessment, as it more accurately captures students' literal, inferential, and evaluative abilities.

References

- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959.
- Amri, N. (2020). Pengaruh Implementasi Asesmen Kinerja Terhadap Karakter Dan Literasi Sains Siswa Kelas Iv Min 2 Konawe Selatan. *Jurnal Pendidikan Dasar Flobamorata*, 1(2), 40-48.
- Anggraeni, M., & Mukhlis, M. (2023). Asesmen kompetensi minimum literasi membaca siswa di sd negeri 09 merangkai. *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*, 9(1), 313-325.
- Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, 24(2), 97–140. <https://doi.org/10.1080/10888691.2018.1537791>
- Ekoati, E. S., Rakhman, F., Zulaeha, I., & Subyantoro. (2021). *The necessity of the Indonesian language authentic assessment model development based on the teacher interview results*. Proceedings of the International Conference on Science, Education and Technology. <https://doi.org/10.2991/assehr.k.211125.020>
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston: Allyn & Bacon.
- Heydarnejad, T., Tagavipour, F., Patra, I., & Khafaga, A. F. (2022). The impacts of performance-based assessment on reading comprehension achievement, academic motivation, foreign language anxiety, and students' self-efficacy. *Language Testing in Asia*, 12(1), 51. <https://doi.org/10.1186/s40468-022-00202-4>
- Hidayat, Z., Syamsuddin, A., & Mukhlis, M. (2025). The numeracy literacy process of fifth-grade elementary school students in solving word problems. *Journal*
-

-
- of *Educational Sciences*, 9(4), 2409–2418.
<https://doi.org/10.31258/jes.9.4.p.2409-2418>
- Imran, M. A., Akhir, M., & Arif, T. (2025). Development of a prototype of children's literature e-module media to improve reading literacy. *Journal of Educational Sciences*, 9(4), 2678–2686.
<https://doi.org/10.31258/jes.9.4.p.2678-2686>
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2021). *Asesmen Kompetensi Minimum (AKM): Literasi membaca*. Jakarta: Pusat Asesmen dan Pembelajaran.
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2022). *Panduan pembelajaran dan asesmen Kurikulum Merdeka*. Jakarta: Kemdikbudristek.
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 21 Tahun 2022 tentang Standar Penilaian Pendidikan*. Jakarta: Kemendikbudristek.
- Lesnick, J., Goerge, R. M., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level by third grade: How is it related to high school performance and college enrollment?*. Chicago, IL: Chapin Hall at the University of Chicago.
- OECD. (2019). *PISA 2018 assessment and analytical framework*. Paris: OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. Paris: OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Pellegrino, J. W., Wilson, M. R., Koenig, J. A., & Beatty, A. S. (2016). *Developing assessments for the next generation science standards*. Washington, DC: National Academies Press. <https://doi.org/10.17226/21858>
- Pujiningsih, A., Nyoman, N. A., & Prayito, M. (2026). Evaluating the effectiveness of literacy and numeracy programs using educational report analysis. *Journal of Educational Sciences*, 10(1), 1011–1024.
<https://doi.org/10.31258/jes.10.1.p.1011-1024>
- Puteri, A. N., Yoenanto, N. H., & Nawangsari, N. A. F. (2023). Efektivitas asesmen autentik dalam pembelajaran. *Jurnal Pendidikan dan Kebudayaan*, 8(1), 77–87. <https://doi.org/10.24832/jpnk.v8i1.3535>
- Putri, S. H. (2024). Analisis Implementasi Kurikulum Merdeka Di SMA Negeri 1 Palipi. *Cognoscere: Jurnal Komunikasi dan Media Pendidikan*, 2(2), 102–107. <https://doi.org/10.61292/cognoscere.180>
- Rahmawati, A., Rosmalina, I., & Anggraini, H. W. (2020). Prosodic reading and reading comprehension in university. *Journal of English Education, Literature, and Culture*, 5(1), 89-108.
- Rosenblatt, L. M. (1978). *The reader, the text, and the poem: The transactional theory of the literary work*. Carbondale, Illinois: Southern Illinois University Press.
- Sudianto, S., & Kisno, K. (2021). Potret kesiapan guru sekolah dasar dan manajemen sekolah dalam menghadapi asesmen nasional. *Jurnal Akuntabilitas Manajemen Pendidikan*, 9(1), 85-97.
- Sugiarsih, S., Suhardi, S., Syamsi, K., & Anggito, A. (2026). Needs analysis for the development of a balanced literacy approach model for beginning reading in elementary schools. *Journal of Educational Sciences*, 10(1), 530–543.
<https://doi.org/10.31258/jes.10.1.p.530-543>
-

- Sukma, H. H., & Amurdawati, G. (2020). Pengembangan modul Bahasa Indonesia dengan pendekatan local wisdom pada materi menulis karangan narasi untuk siswa SD. *Jurnal Kiprah*, 8(2), 143-150.
- Wahdania, A. (2024). *Penerapan strategi pembelajaran directed reading activity untuk meningkatkan keterampilan membaca pemahaman siswa kelas IV SDN Inpres Mallengkei Bertingkat 1 Kota Makassar*. (Skripsi). Universitas Negeri Makassar. <http://eprints.unm.ac.id/id/eprint/35194>
- Wato, I. M., & Mustadi, A. (2026). The effectiveness of QR code-based smart access library as a literacy program in improving elementary school students' reading literacy. *Journal of Educational Sciences*, 10(4), 760–775. <https://doi.org/10.31258/jes.10.4.p.760-775>
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Alexandria, VA: ASCD.

How to cite this article:

Ratnawati., Nuraeni, I., Juniati., Gazali., & Sukma. (2026). Performance-Based Assessment to Measure Narrative Reading Literacy Skills of Fourth-Grade Students at SDN Petunggu. *Journal of Educational Sciences*, 10(5), 494-510.
