



# From Retrieval to Reflection: Profiling the Cognitive Levels and Stimulus Characteristics of Summative Reading Literacy Items in Indonesia

Astria Prameswari\*, Suyono, Nurhadi

Sastra Indonesia Department, Universitas Negeri Malang, Indonesia

\*Corresponding author's email: [astria.prameswari.2202119@students.um.ac.id](mailto:astria.prameswari.2202119@students.um.ac.id)

---

## Submission

### Track:

Received:

28 February 2025

Final Revision:

3 May 2025

Available online:

6 May 2025

## ABSTRACT

Indonesia faces significant literacy challenges, as evidenced by persistently low Programme for International Student Assessment (PISA) scores. To address this issue, a multifaceted analysis is required, particularly examining the characteristics of literacy questions featured in Indonesian language assessments administered to students. One critical factor contributing to poor literacy outcomes is the quality of instruction, which necessitates further scrutiny of assessment design to evaluate the efficacy of learning materials. This study investigates the structural and cognitive dimensions of reading literacy questions in Indonesian language assessments to identify potential shortcomings. Employing a descriptive research approach, the study analyzes the typology of questions, cognitive demand levels (including 1) information retrieval, 2) comprehension, and 3) evaluation and reflection), question quantity, and textual stimuli utilized in summative assessments for eleventh-grade high school students. The findings reveal that the assessments predominantly consist of multiple-choice questions, with a total of 30–50 items per test. The distribution of cognitive levels was uneven: Level 1 (basic information retrieval) accounted for 17% of questions, Level 2 (comprehension) constituted 66%, while Level 3 (evaluation and reflection) represented only 4%. Additionally, the textual stimuli were primarily presented as excerpts or fragmented continuous texts. These findings are subject to limitations, as the study did not account for teacher backgrounds or testing infrastructure conditions. Nevertheless, the results provide valuable insights for Indonesian language educators, serving as a foundation for refining assessment strategies and diversifying evaluation methods to enhance literacy development.

**Keywords:** literacy level, reading text, summative, assessment

---

DOI: 10.23917/varidika.v37i1.8819

## INTRODUCTION

The Programme for International Student Assessment (PISA) is a type of evaluation developed by the Organization for Economic Cooperation and Development (OECD) that is undertaken internationally with 70 countries serving as test subjects (Badan Standar, Kurikulum, [2020](#); Kemdikbud, [2019](#)). The PISA test materials are reading, mathematics and science (Başman, [2024](#)). PISA test materials include reading, math, and science. The test results are utilized as a diagnostic tool for each country, providing the foundation for improving the education system. Indonesia has participated in the PISA test since 2000. The PISA test is carried out periodically or within a certain period (Söyler et al., [2021](#)). PISA results can be a reference for a country in determining the quality of literacy of its students,

including Indonesia. According to the PISA test results, Indonesia's average reading literacy score remains lower than the international literacy average. It was further explained that in 2000, Indonesian students had an average reading literacy score of 371. In 2003, the average reading literacy score among Indonesian students was 382. In 2006, the average reading literacy score among Indonesian students was 393. In 2009, the average reading literacy score among Indonesian students was 402. The most recent result in 2022 was a reading literacy score of 359. The 2022 score is lower than the average international minimum literacy score of 476 (OECD, [2024](#)).

Literacy can be defined as the ability to read and write (Börner et al., [2019](#); Liu & Chen, [2021](#)). Reading literacy is the ability to understand various written information through written language tools. For example, the arrangement of ideas, the selection of titles, the arrangement of words and sentences to deliver the transfer of information to the reader. The understanding of written information is also closely related to thinking skills, reasoning skills, and useful creativity (Harsiati, [2018](#)). Reading literacy can also be extended to the ability to read various information which includes thinking skills using sources of knowledge in print, visual, and auditory forms (Fadli et al., [2020](#)). Literacy in reading and writing is closely related to the subjects of Indonesian language that literacy skills are part of language skills. Therefore, Indonesian learning outcomes affect students' literacy skills (Dwijayati & Rahmawati, [2021](#)). Indonesian language assessment is a part of learning that also contains aspects of literacy learning such as text comprehension and information inference.

The low reading literacy scores of Indonesian students must be carefully assessed so that PISA results do not become the primary reference. In other words, PISA results should not include exaggerated statements about high- and low-performing countries (Jerrim et al., [2023](#)). Several other factors must be investigated in order to determine the causes of students' low literacy skills (Hopfenbeck, [2023](#)). The causes of low reading literacy levels can arise from several things. One of them is the lack of teacher habituation for students to face reading literacy questions. This is not in accordance with the fact that the success of learning is in terms of teacher competence in conducting classroom assessments (Asamoah et al., [2023](#)). The reading motivation of Indonesian students is also considered very low (Suandewi et al., [2019](#)). Reading literacy materials provided to students are also of lower quality as learning materials. To prove this, an analysis of assessment questions used for Indonesian students was conducted.

The didactic purpose of assessment is to measure students' cognition or knowledge as objects (Meijer et al., [2020](#)). This is consistent with the statement that assessment must meet two objectives: 1) measuring competence and 2) having a beneficial effect on the learning process and outcomes (Sudrajat, [2019](#)). These objectives were implemented by all Indonesian language teachers in the SAS/PAS questions for grade XI in Java and Bali. Confirmed grade XI Indonesian teachers gave statements that the SAS/PAS aims to measure the knowledge competence of grade XI students during the odd semester.

The Indonesian teachers also stated that the questions in the reading texts had been directed to literacy questions. To prove this statement, it is necessary to review 6 questions in the SAS/PAS of grade XI Indonesian. The questions analyzed were end-of-semester summative questions or end-of-semester assessments at the high school level in Java and Bali. The end-of-semester summative questions were reviewed to describe the characteristics of the questions. The characteristics in question are the form of the question, the form of the text, the type of literacy level used, and the number of questions.

This research considers the results of similar previous studies. The research objectives are based on similarities with previous studies. The first study was a study that captured the characteristics of reading literacy questions in the PISA study (Harsiati, [2018](#)). The results of this research are in line with the research objectives, namely examining the characteristics of reading literacy questions. The second research is a description of reading and writing literacy learning practices and teachers' constraints in practicing literacy learning at SMAN 1 Pangkalan Bun (Dwijayati & Rahmawati, [2021](#)). The results of this research are the basis for comparing literacy learning practices with the samples reviewed. The third study is a description of the assessment components in language literacy assessment and solutions to the obstacles faced in language literacy learning (Coombe et al., [2020](#); Levi & Inbar-Lourie, [2020](#)). The results of this research become the basis for reviewing the assessment components in language literacy questions. The fourth study, Prasetyo ([2018](#)), confirmed that language teachers' teaching experience determines the quality of language literacy assessments used in the classroom. This research became the basis for a study that looked at the quality of Indonesian literacy items conducted by teachers.

The research objectives also departed from the shortcomings of previous studies. For example, Baidoo-Anu et al. ([2023](#)) found an era relationship between good classroom assessment practices and student learning outcomes in Canada. Researchers have not found a link between the success of literacy learning and the quality of the use of literacy questions. The sixth study is the result of a needs survey on language learning assessment, which states that successful language learning assessment requires the involvement of language teachers (K. Meijer et al., [2023](#); Takkaç & Yastibas, [2018](#)), language researchers, and language learning policy makers (Forzani et al., [2020](#); Kremmel & Harding, [2020](#); Levi & Inbar-Lourie, [2020](#)). This research has not shown the specific role of teacher involvement in the aspect of developing assessments. The second study is Meijer et al. ([2020](#)) who examined the accuracy of literacy assessment methods in higher education and their influence on collaborative learning. This research can be continued to examine the form of Indonesian language assessment as a method of literacy assessment in high school.

Harsiati ([2018](#)) stated that the results of the analysis of PISA reading literacy questions were dominated by higher-level thinking skills in the form of interpretation, reflection, and evaluation abilities. Coombe et al. ([2020](#)) added that teachers often have a literacy assessment model profile that is different from the ideal assessment model. The two statements above are different from the condition of

students' literacy quality in schools. Indonesian students' literacy skills are still at the bottom of the rankings and students' interest in reading is low. Thus, the research objective was formulated to find out the characteristics of reading literacy questions in Indonesian summative questions used in Indonesia. Specifically, the objective was formulated to examine the characteristics of literacy questions in the Indonesian summative in terms of level of cognitive literacy such as 1) finding information, 2) understanding information, and 3) evaluating and reflecting in questions, number of questions, and characteristics of the text used as stimulus. This analysis will be a valuable input for Indonesian language teachers to improve the literacy quality of Indonesian students through practice in summative exams.

## METHOD

### *Research design*

This research used a qualitative descriptive approach. Descriptive research is research that uses data to solve existing problems by presenting, analyzing, and interpreting it. The research object is the text of the End of Semester Summative Assessment (SAS) or End of Odd Semester Assessment (PAS) for grade XI in SMA and SMK, in eight Java and Bali schools.

### *Method of collecting data*

This study employs qualitative analysis of eleventh-grade summative Indonesian language assessments from the even semester, utilizing content analysis and structured note-taking techniques. The content analysis systematically examines question formats (e.g., multiple-choice), text structures (e.g., excerpts or continuous passages), cognitive literacy levels (categorized as information retrieval, comprehension, and evaluation/reflection), and question quantity. Complementary note-taking documents observed patterns from both direct examination and written sources. Data collection instruments include a validated question-item assessment sheet and semi-structured teacher interview protocols, reviewed by learning evaluation experts. Gathered data is categorized thematically and subjected to comprehensive analysis to identify assessment design trends.

### *Source data*

The source of the research data was the SAS/PAS questions of the Indonesian language in class XI majoring in science or social studies at high schools and vocational schools in West Java, Central Java, East Java, and Bali. A total of six questions from five high schools and 1 vocational school were analyzed. Two schools from East Java, one school from West Java, one school from Central Java, one school from Bali, and one school from the Special Capital Region of Jakarta. The number of question papers that became the data source amounted to six question papers. The questions were selected from six favorite and accomplished schools spread across Java and Bali. Sample number can be a limitation

of the study because bias is possible. They were chosen because they are favorite schools and have Indonesian language teachers who are outstanding teacher finalists. All schools voluntarily participated in assisting with the research.

### ***Data analysis***

Data analysis was carried out by data selection, categorization, presentation of data, drawing conclusions, and triangulation (Miles et al., [2014](#)). The question scripts were read and then the items were classified into sub forms of questions, sub levels of literacy, sub forms of stimulus, and sub number of items. After being grouped, each item was coded, reviewed and summarized. The last stage was triangulated through re-interviewing the teachers regarding the results of the review.

## **RESULTS & DISCUSSION**

### ***Result***

#### *Characteristics of reading literacy problem forms in the final summative of the XI grade odd semester*

According to the results of a review of 6 manuscripts of SAS/PAS questions in the Indonesian Language of Grade XI, the form of questions used was multiple choice with one correct answer of up to 100 percent. The answer options used as answer choices range from 1 to 5, with the letters A, B, C, D, and E. The teachers chose multiple choice questions with one correct answer because the school uses the Computer Based Test (CBT) method to implement SAS/PAS.

#### *Characteristics of reading literacy text forms in the odd semester final summative of grade XI*

The following is a screen capture of the text form used in the Indonesian Language SAS/PAS of Grade XI.

#### *Narration text*

##### **Observe the following short story excerpt!**

If there was a match in the early morning, my father and I worked together to wake each other up. 2) The two of us would stay awake, or not sleep until dawn, sitting glued to the 14-inch Grundig TV with its plywood shutters, accompanied by glasses of coffee. 3) At Ullevi Stadium in Gothenburg, this blond-haired team shattered the Dutch defense in just the first 5 minutes with a header from Larsen: 1 – 0. 4) I raised my fist high in the air, “Yes!” I shouted. 5) I glanced at my father; he shook his head while clearing his throat.

**The evidence of the *place setting* in this novel excerpt is shown in sentence ....**

- a. 1)
- b. 2)
- c. 3)
- d. 4)
- e. 5)

*Argumentation text***Observe the following text!**

Based on the explanation above, it cannot be denied that Indonesia is one of the countries with the greatest biodiversity. Much of this biodiversity is exploited by other nations, and we receive nothing in return. Amidst the ongoing climate change, it is time for us to focus on local food sources to ensure our survival. (*Kompas*, modified-2019)

**The excerpt above is part of the structural section...**

- a. Thesis
- b. Argumentation
- c. Reaffirmation
- d. Coda
- e. Opinion

*News text***Observe the following news texts!****News Text 1**

The death of Theresa has angered the public. How could it not? This little girl, who was reported to have improved, was instead found buried in her own backyard. Ironically, Theresa's death was caused by someone close to her. During her life, the little girl received no attention from her adoptive mother. She was even treated like a servant.

**News Text 2**

The police have finally identified a lecturer suspected of neglecting his four children. According to investigators, the children were left to live in uninhabitable conditions. One of them was even forbidden from entering the house and had to sleep in a guard post. Further investigation revealed that the suspect was also a drug addict.

**The similarity in the content of the two news excerpts above is?**

- a. Child abuse by biological parents
- b. Cases of child neglect
- c. Cases of child murder
- d. Cases of drug abuse
- e. Cases of wiretapping

*Procedure Text***Read the following procedure text!**

Cincau leaves are placed in a bucket, then poured with boiling water. During the soaking process, the leaves are stirred with a tool to ensure even softening. The boiling water will soften the initially hard and stiff leaves, making them limp and less shiny. The hot water also aids in the evaporation process, reducing the strong aroma of the cincau leaves.

**The excerpt above is part of...**

- A. Purpose
- B. Materials
- C. Steps
- D. Conclusion
- E. Interpretation

The use of text in the summative questions above also does not use text in other forms such as audiovisual text. The form of text used is still continuous text which can only be read in words and sentences. This is contrary to the findings of Bayazit & Akçapinar (2023) who found that the effect of video-assisted assessment would increase students' understanding of the questions to be solved.

*Characteristics of reading literacy level in the odd semester final summative of grade XI*

Based on the Indonesian language summative questions of grade XI that were reviewed, the percentage results of each level were obtained. The results of the analysis can be observed in the Table

**Table 1.** Percentage of Reading Literacy Level

School Code	Information Retrieval Level	Understanding Level	Evaluate and Reflect Level	Not Literacy Level Questions
1	68	4	0	28
2	62	15	4	19
3	72	12	24	12
4	43	22	0	35
5	90	6	0	4
6	70	6	0	23

The data in the table shows the average percentage results for four aspects, namely Information Retention Level, Understanding Level, Evaluate and Reflect Level, Not Literacy Level Questions. The average is calculated from the total number of items and questions that show indicators of each aspect of the review.

Of the six schools studied, there were two results that showed the level of literacy in Bahasa Indonesia questions. The highest percentage was at the low literacy level (finding information). The percentage of the six schools is above 50 percent compared to other levels. Thus, most Bahasa Indonesia questions in the six schools are still at a low level. The second result shows that the high literacy level (evaluating and reflecting) got the lowest percentage. There were four out of six schools that did not have any questions at the high literacy level. So, Bahasa Indonesia subject questions need improvement by adding literacy questions with evaluation and reflection levels.

The following is the question that contains the level of finding information in one of the analyzed grade XI Indonesian summative questions.

**Analyze the following excerpt!**

The fiber in sweet potatoes can reduce the risk of heart disease, stroke, and diabetes.  
Vitamin A in sweet potatoes is very important for eye health.  
Vitamin C in sweet potatoes is essential for the immune system.  
Vitamin B6 in sweet potatoes is crucial for body metabolism.  
Potassium in sweet potatoes helps maintain blood pressure.

**Which of the following statements is correct regarding the linguistic aspects of the argumentative text above?**

- A. The paragraph is weak because it contains non-standard words, such as "*serat*" (fiber).
- B. There are five non-standard words in the text.
- C. The dominant conjunctions in the paragraph are causal conjunctions.
- D. The paragraph is weak due to incoherence, caused by the second sentence.
- E. "*Resiko*" (risk) is one of the non-standard words in the paragraph.



The following is the question that contains the level of understanding information in one of the analyzed grade XI Indonesian Language summative questions.

**Analyze the following two short story excerpts!**

Excerpt 1:

*"The poor man once worked as a clown. His costume was tattered and pitiful. Every time he entertained at birthday parties, the children watching his act would cry in fear. (Excerpt from "Orang Miskin yang Bahagia" by Agus Noor)*

Excerpt 2:

*"He was not a wealthy man. Circumstances forced him to migrate. When he first employed. His employer quickly took a liking to him because he was diligent and honest. He was taught to save at a bank. right—he was as tenacious as rattan.*

**The similarity between the two short story excerpts is:**

The following is the question that contains the level of evaluating and creating information in one of the analyzed grade XI Indonesian Language summative questions.

**1 portion of rice is equivalent to:**

- 3 ears of corn
- 2 potatoes
- 2 bananas
- 15 pieces of cassava
- 1/2 taro
- 100 grams of sago/8 tablespoons of sago flour

**Based on the poster above, the most appropriate call to action is:**

- A. Let's eat a portion of rice to fill our stomachs.
- B. Rice is no longer healthy; let's switch to other foods.
- C. Overcome rice dependency; let's switch to alternative foods.
- D. Let's abandon rice; its price is now unaffordable.
- E. Let's eat corn to feel as full as eating rice.

The last category is questions that are not included in the literacy level questions. The questions used did not meet the requirements of the three-reading literacy thinking levels. A total of 17 percent of the questions fell into this category. This is due to the fact that the questions only asked about the memorization aspect of certain theories. For instance, shown in the following screenshot.

A news text is a text that presents information that is...

- A. Objective and subjective
- B. Accurate and argumentative
- C. Factual and current
- D. Informative and systematic
- E. Argumentative and objective



### *Characteristics of the number of reading literacy questions in the odd semester final summative of grade XI*

The six Indonesian questions analyzed reveal information about the number of questions used in the final summative assessment of the odd semester for grade XI Indonesian Language. The end-of-semester summative test lasted 90 minutes. Three schools used summative assessments totaling 50 items. Two other schools used summative assessments with a total of 40 items. The last school used a summative assessment with 30 items.

### **Discussion**

#### *Characteristics of reading literacy problem forms in the final summative of the XI grade odd semester*

According to the results of a review of 6 manuscripts of SAS/PAS questions in the Indonesian Language of Grade XI, the form of questions used was multiple choice with one correct answer of up to 100 percent. The answer options used as answer choices range from 1 to 5, with the letters A, B, C, D, and E. The teachers chose multiple choice questions with one correct answer because the school uses the Computer Based Test (CBT) method to implement SAS/PAS. In addition to computers, teachers use applications such as Google Forms and Moodle-based Learning Management System (LMS) to conduct the test. The multiple-choice form with one correct answer is preferred because it is easier to correct the answers. The assistance app used during the exam can display the results of student answers up until the final score stage. If the teacher uses other types of questions, such as essays, the teacher will have to correct them manually, which takes longer.

The use of digital technology in conducting assessments is permissible. This is because it can train students to be more technologically literate and train students to improve their competence in reading text visually. According to Börner et al. (2019), the ability to read and write visual data is as important as the ability to read and write plain text. In addition, the use of digital technology can also improve students' abilities more broadly (McDougall et al., 2018).

However, the use of only one form of question in an assessment has disadvantages. First, the use of one form of question is considered a conventional or less innovative assessment and has limitations (Aybek, 2023). At the same time, aspects of innovation and creativity are indispensable in the era of the Industrial Revolution 4.0 (Wayan Widana, 2020). Second, the use of one form of question in the assessment will make students limited in terms of mastering their knowledge and expertise (Childress et al., 2020). The possibility of interactive answers will also be closed. Dialogue in the process of answering questions will appear in essay questions that allow students to answer in various perspectives. Questions with open answers give students more freedom in providing answers (Kocaoğlu, 2024) compared to multiple choice questions. This dialogue is significant because it engages students in the process of creating meaning together (Baartman & Quinlan, 2023) and allows for feedback between

student and teacher responses (de Kleijn, [2023](#); Deeley & Bovill, [2017](#)). Furthermore, the problem's limited form will make it difficult for students to find alternative problem-solving methods. This difficulty is consistent with the opinion of Ideyani Vita et al., ([2020](#)) who argue that reading assessment as reading literacy must improve students' cognitive aspects as readers. Assessment must also be a fair method of assessing student learning (Childress et al., [2020](#)).

To mitigate the overreliance on multiple-choice questions in literacy assessments, teachers can reduce the number of multiple-choice items. Two, gradually decrease the quantity of multiple-choice questions in assessments to encourage deeper engagement with texts. Three, diversify question formats. Four, introduce acceptable alternative formats that align with digital assessment tools, such as complex multiple-choice questions and matching exercises.

Based on the findings, triangulation through teacher interviews revealed that multiple-choice questions were primarily selected for two reasons. First, efficiency in test construction, as they significantly reduce item development time. Two, ease of scoring, particularly for digital assessments. Also additionally, institutional regulations mandate this format for standardized digital testing.

#### *Characteristics of reading literacy text forms in the odd semester final summative of grade XI*

The types of literacy texts used varied across the six SAS/PAS question papers reviewed. In the Indonesian SAS/PAS, reading literacy questions were based on narrative, argumentation, news, explanation, and procedure texts. Baghana et al. ([2019](#)) classify Aristotle's text types as narrative, argumentative, and descriptive. Another division includes poetry, other literary texts, and scientific work texts. Another classification is continuous and non-continuous texts. Continuous texts include descriptions, narratives, expositions, and arguments. Non-continuous text includes simple lists, mixed lists, column lists, network lists, combination lists), with specific formats such as vouchers, certificates, advertisements, charts/graphs, tables and matrices, lists of things, maps (Harsiati, [2018](#)).

#### *Characteristics of reading literacy level in the odd semester final summative of grade XI*

Reading literacy has three levels. The first level is access and retrieve. Thinking activities carried out at the access level are finding, identifying and describing an idea or explicit information contained in the text. As for the retrieve aspect, the activity that can be done is to describe the process of selecting the information needed, whereas access is more about how the process reaches the place or existence of the information needed. The second level is understanding (interpret and integrate). Activities include comparing ideas or making inferences, categorizing, and combining ideas and information within texts or between texts. The third level is evaluated and reflect. Activities include analyzing, predicting, and assessing content, language, and elements in the text, reflecting or creating a picture or opinion.

From the data in the table of percentage reading literacy, it is shown that the percentage of the level of retrieving information in Grade XI Indonesian summative questions in Java and Bali is 66 percent. The questions used were multiple choice with one correct answer. The question instructions that appear are such as 1) *the mandate contained in the fragment of the text above is ...*, 2) *the sentence of fact contained in the body of the text argument above is ...*, 3) *the following is a correct statement regarding the language of the argumentation text excerpt above ....* The type of text used as a question stimulus were adjusted to the learning target for one semester. The form of text used was also in the form of fragments or quotations. This is not in accordance with the principle of literacy text which should be a whole text.

The second percentage result comes from the level of understanding the information in the summative questions of Indonesian Language Grade XI in Java and Bali at 13 percent. At this second level, most question instructions were in the form of summarizing information in the stimulus text and converting information into another form. The questions used were multiple choice with one correct answer. The question instructions that appeared were such as 1) *the similarity between the two fragments in the short story excerpt is ...* 2) *the conclusion of the text fragment above is ....* The type of text used as a question stimulus is adjusted to the learning target for one semester. The form of text used is also in the form of fragments or quotations. This is not in accordance with the principle of literacy text which should be a whole text.

The third percentage result comes from the level of evaluating and creating information in the Indonesian Language summative questions of Grade XI in Java and Bali at 4 percent. At this level, there are few question instructions that indicate the activities of evaluating and creating information in the information. Questions demonstrating the level of evaluation and creation only appear in some types of questions, such as multiple-choice questions with one correct answer. The question instructions that appear are as follows. *According to the poster above, the most appropriate invitation is as follows.* Multimodal text, such as posters, are used as a stimulus for third-level questions. Students are asked to convert the poster's brief information into narrative information.

The last category is questions that are not included in the literacy level questions. The questions used did not meet the requirements of the three-reading literacy thinking levels. A total of 17 percent of the questions fell into this category. This is since the questions only asked about the memorization aspect of certain theories.

#### *Characteristics of the number of reading literacy questions in the odd semester final summative of grade XI*

The analysis reveals that summative assessments for Grade XI Indonesian language consistently comprise 30–50 questions, a quantity teacher has calibrated to student ability levels and testing

constraints. These findings carry significant pedagogical implications, equipping educators with empirical insights to refine their approach to literacy-question design. By prioritizing higher-quality, cognitively balanced assessments, teachers can foster more robust literacy engagement, ultimately cultivating students' critical competencies. While such improvements promise to elevate Indonesia's broader educational quality, the study's geographic limitation to Java and Bali necessitates caution in generalizing these outcomes nationally. Future research should expand its scope to include a more diverse, nationally representative sample of schools, ensuring findings capture the full spectrum of Indonesia's educational landscape and yield more universally applicable recommendations for literacy assessment reform.

## CONCLUSION

This examination of Grade XI Indonesian language summative assessments across Java and Bali reveals a concerning disparity between assessment design and literacy competency development. The evaluations predominantly employ restrictive response formats—notably multiple-choice and constrained essay questions—paired with textual stimuli ranging from conventional prose to digital media, all ostensibly aligned with national curricular standards. However, a critical analysis of cognitive demand exposes a fundamental imbalance: assessments disproportionately emphasize rudimentary information retrieval (Level 1) while underrepresenting intermediate comprehension (Level 2) and nearly neglecting higher-order evaluation and reflection (Level 3). Such skewed cognitive distribution, coupled with the logistical constraints of digital administration platforms, demonstrates how current practices prioritize logistical efficiency over pedagogical substance. This misalignment not only fails to prepare students for the complex literacy demands of international benchmarks but perpetuates a cycle of superficial engagement with texts. The findings necessitate immediate curricular intervention to recalibrate question design, specifically through intentional scaffolding of Level 2 and Level 3 cognitive tasks, which would simultaneously address short-term assessment readiness and cultivate the critical literacy competencies essential for navigating Indonesia's evolving information ecosystem.

## REFERENCES

- Alkan, M., & Doğan, N. (2023). A Comparison of Different Designs in Scoring of PISA 2009 Reading Open Ended Items According to Generalizability Theory. *Journal of Measurement and Evaluation in Education and Psychology*, 14(2), 106–117. <https://doi.org/10.21031/epod.1210917>
- Asamoah, D., Shahrill, M., & Abdul Latif, S. N. (2023). Towards developing classroom assessment literacy: Exploring teachers' approaches to assessment across cultures. *Cogent Education*, 10(2). <https://doi.org/10.1080/2331186X.2023.2280301>
- Badan Standar, Kurikulum, dan A. P. K. (2020). *PISA*. <https://bskap.kemdikbud.go.id>
- Baghana, J., Buzina, E. I., Glamazda, S. N., V, T., Khvesko, & Lazareva, O. P. (2019). Literary Text as Knowledge Format. *Journal of Research Applied Linguistics*, 10, 1186–1196.

- Baidoo-Anu, D., Rasooli, A., DeLuca, C., & Cheng, L. (2023). Conceptions of classroom assessment and approaches to grading: teachers' and students' perspectives. *Education Inquiry*, 00(00), 1–29. <https://doi.org/10.1080/20004508.2023.2244136>
- Başman, M. (2024). Measurement Invariance of Achievement Motives Model: PISA 2018 Turkey Sample. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 15(1), 18–34. <https://doi.org/10.21031/epod.1302574>
- Bayazit, A., & Akçapınar, G. (2023). Design and Development of an Interactive Video Player for Supporting Formative Assessment in Online Learning. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 14(Özel Sayı), 320–331. <https://doi.org/10.21031/epod.1286077>
- Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America*, 116(6), 1857–1864. <https://doi.org/10.1073/pnas.1807180116>
- Can Aybek, E. (2023). The Relation of Item Difficulty between Classical Test Theory and Item Response Theory: Computerized Adaptive Test Perspective. *Journal of Measurement and Evaluation in Education and Psychology*, 14(2), 118–127. <https://doi.org/10.21031/epod.1209284>
- Childress, J., Backman, A. C., & Lipson, M. Y. (2020). Reframing Literacy Assessment: Using Scales and Micro-Progressions to Provide Equitable Assessments for All Learners. *Journal of Adolescent and Adult Literacy*, 63(4), 371–377. <https://doi.org/10.1002/jaal.1016>
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Correction to: Language assessment literacy: what do we need to learn, unlearn, and relearn? (Language Testing in Asia, (2020), 10, 1, (3), 10.1186/s40468-020-00101-6). *Language Testing in Asia*, 10(1). <https://doi.org/10.1186/s40468-020-00114-1>
- de Kleijn, R. A. M. (2023). Supporting student and teacher feedback literacy: an instructional model for student feedback processes. *Assessment and Evaluation in Higher Education*, 48(2), 186–200. <https://doi.org/10.1080/02602938.2021.1967283>
- Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: enhancing assessment literacy through democratic practices. *Assessment and Evaluation in Higher Education*, 42(3), 463–477. <https://doi.org/10.1080/02602938.2015.1126551>
- Dwijayati, C. D. C., & Rahmawati, L. E. (2021). Kendala Literasi Baca Tulis Sebagai Implementasi Gerakan Literasi Nasional Di Sma Negeri 1 Pangkalan Bun. *Tabasa: Jurnal Bahasa, Sastra Indonesia, Dan Pengajarannya*, 2(1), 17–32. <https://doi.org/10.22515/tabasa.v2i1.2685>
- Fadli, R. I., Nugraha, A. S., Raharjo, R. P., Sulton, A., & Sari, R. H. (2020). Model Pembelajaran Inovatif Guru Sma Abdul Hadi. *Jurnal ABIDUMASY*, 01(01), 1–12.
- Forzani, E., Corrigan, J. A., & Slomp, D. (2020). Reimagining Literacy Assessment through a New Literacies Lens. *Journal of Adolescent and Adult Literacy*, 64(3), 351–355. <https://doi.org/10.1002/jaal.1098>
- Harsiati, T. (2018). Karakteristik Soal Literasi Membaca Pada Program Pisa. *Litera*, 17(1), 90–106. <https://doi.org/10.21831/ltr.v17i1.19048>
- Hopfenbeck, T. (2023). Cultivating realistic self-appraisal: examining student overclaiming and fair assessment through PISA and classroom data. *Assessment in Education: Principles, Policy and Practice*, 30(5–6), 347–350. <https://doi.org/10.1080/0969594X.2023.2294428>
- Ideyani Vita, N., Agus Zainal, M., STIK Pembangunan Medan, D., & STIK Pembangunan Medan, A. (2020). Gerakan Literasi Membaca: Studi Fenomenologi tentang Gerakan Literasi Membaca Siswa SMA Negeri 2 Medan. *Jurnal Ilmiah Ilmu Komunikasi Communique*, 3(1), 41–47. <https://ejurnal.stikpmedan.ac.id/index.php/JIKQ/article/view/41>
- Jerrim, J., Parker, P. D., & Shure, N. (2023). Overclaiming. An international investigation using PISA data. *Assessment in Education: Principles, Policy and Practice*, 30(5–6), 351–371. <https://doi.org/10.1080/0969594X.2023.2238248>
- Kemdikbud. (2019). *Hasil PISA 2018: Akses Makin Meluas, Saatnya Tingkatkan Kualitas*.
- Kocaoğlu, S. (2024). Investigating the Effect of Testlets Consisting of Open-Ended and Multiple-Choice Items on Reliability via Generalizability Theory \*. 15(1), 65–78.
- Koh, K., & Chapman, O. (2019). Problem-Based Learning, Assessment Literacy, Mathematics Knowledge, and Competencies in Teacher Education. *Papers on Postsecondary Learning and*

- Teaching*, 3, 74–80.
- Kremmel, B., & Harding, L. (2020). Towards a Comprehensive, Empirical Model of Language Assessment Literacy across Stakeholder Groups: Developing the Language Assessment Literacy Survey. *Language Assessment Quarterly*, 17(1), 100–120. <https://doi.org/10.1080/15434303.2019.1674855>
- Levi, T., & Inbar-Lourie, O. (2020). Assessment Literacy or Language Assessment Literacy: Learning from the Teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Liu, Y., & Chen, S. (2021). Physical literacy in children and adolescents: Definitions, assessments, and interventions. *European Physical Education Review*, 27(1), 96–112. <https://doi.org/10.1177/1356336X20925502>
- McDougall, J., Readman, M., & Wilkinson, P. (2018). The Uses of (Digital) Literacy. *Learning, Media and Technology*, 43(3), 263–279. <https://doi.org/10.1080/17439884.2018.1462206>
- Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J. W. (2020). Unfolding collaborative learning assessment literacy: a reflection on current assessment methods in higher education. *Assessment and Evaluation in Higher Education*, 45(8), 1222–1240. <https://doi.org/10.1080/02602938.2020.1729696>
- Meijer, K., Baartman, L., Vermeulen, M., & de Bruijn, E. (2023). Teachers' conceptions of assessment literacy. *Teachers and Teaching: Theory and Practice*, 29(7–8), 695–709. <https://doi.org/10.1080/13540602.2023.2190091>
- OECD. (2024). *GPS Pendidikan Indonesia-Prestasi Siswa (PISA 2022)*. <https://Gpeducation.Oecd.Org>. <https://gpeducation.oecd.org>
- Prasetyo, A. H. (2018). Language Assessment Literacy Development: a Student-Teacher's Experiences in Teaching Practice Program. *LLT Journal: A Journal on Language and Language Teaching*, 21(2), 199–206. <https://doi.org/10.24071/llt.v21i2.1224>
- Söyler, P. B., Aydın, B., & Atilgan, H. (2021). Pisa 2015 reading test item parameters across language groups: A measurement invariance study with binary variables\*. *Journal of Measurement and Evaluation in Education and Psychology*, 12(2), 112–128. <https://doi.org/10.21031/epod.800697>
- Suandewi, P. M., Putrayasa, I. B., & Gunatama, G. (2019). Hubungan Budaya Literasi (Baca-Tulis) Dengan Hasil Belajar Bahasa Indonesia Siswa Kelas Xisma Negeri 7 Denpasar. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia Undiksha*, 9(2). <https://doi.org/10.23887/jjpbs.v9i2.20453>
- Sudrajat, D. (2019). Asesmen Pembelajaran Bahasa Inggris: Model dan Pengukurannya. *Jurnal Pendidikan Dan Pembelajaran: Intelegensia*, 4(1), 1–20. <http://intelegensia.web.id/index.php/intelegensia/article/view/27>
- Takkaç, M., & Yastibas, A. E. (2018). Understanding language assessment literacy: Developing language assessment. *Journal of Language and Linguistic Studies*, 14(1), 178–193.
- Wayan Widana, I. (2020). The Effect of Digital Literacy on the Ability of Teachers to Develop HOTS-based Assessment. *Journal of Physics: Conference Series*, 1503(1). <https://doi.org/10.1088/1742-6596/1503/1/012045>