



*Original Research*

## Reconstructing Language Tests: A Comparative Study for Improved Assessment

**Ikrimah Ikrimah & Natalina Asi**

University of Palangka Raya, Palangka Raya, Indonesia

---

### Article Info

#### *Article history:*

Received 5 January 2024

Revised 28 January 2024

Accepted 19 February 2024

---

#### *Keywords:*

Bloom's taxonomy

Cognitive skills

Comparative study

Language testing

Test reconstruction



---

### Abstract

In the realm of educational systems, language testing plays a vital role in assessing a learner's language abilities. Ensuring the reliability and validity of such tests is crucial to provide an accurate gauge of proficiency and depth of learners' knowledge. However, modifications are necessary for testing methods to cope with the ever-evolving landscape of teaching methodologies while maintaining their effectiveness. This study employs a comparative analysis approach between an original and a reconstructed English language test, following the guidelines of the 'Kurikulum Merdeka' for 11th-grade students. Utilising Bloom's Taxonomy as a framework, both the original and reconstructed tests are evaluated with respect to cognitive complexity and coverage of various cognitive skills. The implemented modifications in the test demonstrated a significant enhancement in the overall effectiveness and alignment with broader cognitive skills. Two key modifications, expanding the multiple-choice answers from four to five and realigning questions to span various cognitive levels as per Bloom's Taxonomy, were found to be effective in increasing the complexity and scope of the test. This, in turn, led to a more comprehensive engagement of students in the test. The study concludes that the reconstructed test format, when compared to the original, offers a more effective overall assessment structure. Its more challenging nature stimulates critical thinking, deep understanding, and creativity among learners, thereby promoting improved language assessment and a better understanding of teaching-learning processes. It underscores the necessity of periodic improvements and modifications in test designs, ultimately contributing towards more precise and effective language testing.

---

**Corresponding Author:** Ikrimah, [ikri12jaliha@gmail.com](mailto:ikri12jaliha@gmail.com)

---

## 1. Introduction

Language testing forms the backbone of any linguistic educational curriculum and is absolutely fundamental when trying to accurately assess a learner's language abilities (Dooly & Vinagre, 2022). The main purpose of these tests, whether it be oral or written, informal or formal, is to delineate a learner's proficiency as well as the depth of their understanding in various facets of the language. However, for these tests to provide a fair and complex portrayal of a learner's capabilities and comprehension, they must incorporate two crucial attributes: reliability and validity (Sumarsono et al., 2023).

The original test under study was designed within this context. Designed by an experienced English teacher, it was devised as a comprehensive measure to evaluate learners' skill. The integral aspect of the test was to determine the students' understanding and proficiency level of English. The exam format made substantial use of traditional testing methodologies such as multiple-choice questions, fill-in-the-blanks and short essay prompts, which offered a structured and easy-to-mark means of assessing fundamental language skills. An appreciable emphasis was put on examining the students' understanding of grammar rules, vocabulary usage, listening comprehension, and basic writing abilities. While the test aimed to provide a

well-rounded view of the learners' language aptitude, it followed a somewhat linear approach. Predominantly, it relied on the information recall and comprehension viewpoints of learning, underestimating the role of higher-order thinking skills and the complexity of real-world language use.

In addition, an effective language test should account for a holistic view of a student's linguistic skills, such as reading, writing, listening, and speaking as interconnected elements, in various contexts (Runesi et al., 2022). Moreover, it is necessary to consider cultural competence and intercultural communication ability, which are important for communication and comprehension in today's interconnected world (Mulyanah & Krisnawati, 2022). To that end, task-based approaches or authentic assessments are often used to measure these skills in more realistic settings and situations, which can better reflect a learner's actual communicative competence (Karaman, 2021). Furthermore, the continuous advancement in technology also influences the design and implementation of language assessments. Computer-assisted language testing is an emerging area offering exciting potentials such as immediate feedback, engaging multimedia input, and adaptive testing according to the test-taker's level (Syathroh et al., 2021). To put it in perspective, effective language assessments should be multifaceted mirroring the complexity of language use and aim to capture the breadth and depth of a learner's linguistic repertoire in diverse circumstances.

Reliability is an important factor in achieving the tests' core function and essentially involves ensuring that the tests display a high level of consistency. Test results need to persist and hold true in different contexts and environments and must also yield the same outcome over time. This consistency in replicating results is paramount because it validates the quality of the test, boosting its credibility and its power to help in understanding a learner's proficiency level reliably (Sürücü & Maslakçi, 2020).

On the other hand, according to (Brown, 1996), validity is an essential consideration in language testing as it pertains to how accurately the test measures what it is intended to measure. To ensure validity, a language test must carefully construct questions that effectively examine the core language competencies it aims to evaluate. This involves developing a test that adequately covers the entire spectrum of language skills and understanding. Without such thorough coverage, the test runs the risk of losing its purpose and credibility, as it may not accurately assess the desired language abilities.

The consequentiality and complexity of these attributes, reliability and validity, potentialise the necessity of test reconstruction. This is because achieving such complex attributes requires a well-structured and periodically reviewed test mechanism that profoundly reflects the dynamism of language and real-world language use.

Considering the limitations of the original test, it becomes evident that a reconstructive approach is necessary to address these issues effectively. This approach should encompass a more comprehensive language acquisition model that aligns with the modern understanding of learning. In this context, the principles of Kurikulum Merdeka, emphasising learner autonomy and self-directed learning, and Bloom's Taxonomy, focusing on higher-order thinking skills, become significant factors to consider (Kamal, 2023). By incorporating these principles into the test construction process, it will be possible to create a more robust and relevant language assessment tool.

The Kurikulum Merdeka, which emphasises creativity and critical thinking, provides a framework for developing dynamic language proficiency tests that reflect modern understanding of language learning<sup>1</sup>. This curriculum also stresses the importance of digital literacy and intercultural skills, which are becoming increasingly valuable in today's globalised society (Anggraena et al., 2022). By incorporating these principles into language testing, it becomes possible to assess not only language proficiency but also digital literacy and intercultural competencies, giving a more comprehensive view of a candidate's language skills.

Bloom's Taxonomy offers a robust framework for classifying learning objectives across a range of cognitive levels. This taxonomy encourages language testing to move beyond simple recall and comprehension tasks and to incorporate higher-order thinking skills, including analysis, synthesis, and evaluation. By incorporating these higher-order thinking skills into language tasks, it becomes possible to assess learners' ability to analyse language structures, synthesise information from various sources, and evaluate arguments or opinions. Bloom's Taxonomy provides a structured approach that guides the development of language tests, ensuring they cater to a variety of diverse and complex language skills (Harahap et al., 2023).

Integrating the guiding principles of Kurikulum Merdeka and Bloom's Taxonomy in the process of reconstructing the language test enables researchers to design a comprehensive and robust testing model that improves upon the limitations of the original test. This model aims to not only enhance the reliability and validity of the assessment but also provide a comprehensive estimate of students' language skills across a range of proficiency levels, from basic understanding to critical analysis and creative output (Sadora et al., 2021). By aligning the test with these principles, it becomes possible to assess learners' ability to engage in higher-order thinking tasks, demonstrate creativity, and apply their language skills in various contexts.

Language, indeed, is a dynamic entity that consistently changes and evolves. As a result, in the fluid landscape of language teaching methodologies, the necessity for efficient and precise language assessment intensifies (Natsir et al., 2020). To stay pertinent, precise, and comprehensive, these tests need to evolve congruently with these changes. Regular revisions and modifications are paramount to keeping up with language developments, effectively assessing a broad spectrum of language skills, and accommodating learner diversity.

Moreover, in light of the increasing use of digital technologies in language teaching and learning, the methodology for assessment requires to be aligned with this digital shift (Armanda & Yosintha, 2022). Digital assessments can offer innovative ways of testing skills, such as collaborative problem solving and digital literacy, which are becoming integral to twenty-first-century communication. Today's learners are digital natives; thus, leveraging technology in testing can make assessments more engaging and relevant for them.

In addition to traditional forms of language use, the emergence of new communication platforms, such as online writing and social media discourse, necessitates their integration into language testing practices (Alfia et al., 2020). These new forms of communication require specific language skills that may not be effectively captured by traditional tests. For instance, online writing demands the ability to convey ideas concisely, use appropriate emoticons or emojis for expression, and understand the nuances of informal language use. Similarly, social media discourse requires skills in responding to comments, engaging in threaded discussions, and understanding language conventions unique to specific platforms. By acknowledging and incorporating these new forms of language use in testing, language assessments can better evaluate learners' proficiency in a wide range of language skills and their ability to adapt to evolving communication contexts.

Furthermore, a growing understanding of the complex interplay between language and cultural identity also necessitates the reconsideration of how language proficiency and intercultural competence are assessed (Subandowo, 2022). Hence, tests need to not only evaluate a learner's syntactic and semantic knowledge but also their ability to navigate through diverse cultural contexts and nuances.

Keeping these considerations in mind, this research aims to undertake a comparative analysis of an original language test and its subsequent reconstructed version. It particularly emphasises the enhancements implemented during the reconstruction process, a vital factor in language testing underscored earlier by Callaghan-Koru & Aqil (2022). As practitioners and researchers in the field, the core interest lies in understanding how these modifications can potentially increase reliability, enhance validity, and assure better alignment with the targeted language skills, as delineated by Effendi & Mayuni (2022).

Building upon the previous studies by Callaghan-Koru & Aqil (2022) and Effendi & Mayuni (2022), this research aims to contribute to the existing literature by conducting a comprehensive comparative analysis of an original language test and its subsequent reconstructed version. While previous research has emphasised the significance of enhancements made during the reconstruction process in achieving higher reliability, validity, and alignment with targeted language skills, there remains a gap in understanding how these modifications specifically impact the assessment's efficacy and potential biases. By conducting a thorough examination of the revised language test and analysing its alignment with targeted language skills, this study seeks to bridge this gap and shed light on the effectiveness of these modifications in improving overall language assessment quality.

Moreover, the paper seeks to explore the impact of these modifications on the test takers' performance and their perceptions of the assessment process. A detailed analysis of candidate responses and feedback can provide valuable insights into the effectiveness of test changes (Sumarsono et al., 2023). Investigating the test takers' perspective is an essential step in validating a test and understanding its potential biases and blind spots.

A further layer of scrutiny will be given to the potential influence of the test reconstruction on the alignment of the assessment with curriculum standards and learning outcomes, a crucial aspect as underlined by [Ulfie & Permana \(2021\)](#). This evaluation will shed light on how well the test matches the intended learning objectives and informs instructional practices.

Lastly, the study will probe how the revised test caters to a diverse range of learners - in terms of language proficiency levels, learning preferences, and needs. It will bring to the fore the inclusivity aspect of the language test, an area often overlooked but of growing significance in contemporary education ([Kusumaningsih, 2021](#)).

This paper will provide an in-depth analysis of both the original test and its reconstructed version. This will involve evaluating the alterations made and their effectiveness in meeting the intended objectives. Being aware of the process and implications of reconstructing a language test can guide educators, curriculum designers, and all stakeholders to make efficient decisions about future development. Consequentially, the knowledge gained from this analysis will contribute significantly to ensuring that language tests adhere to their core principles and remain relevant, accurate, and efficacious both immediately and in the long term.

## 2. Literature Review

Assessing reading comprehension is a crucial aspect of language testing, as it determines a learner's ability to understand, interpret, and analyse written texts ([Shea, 2020](#)). This process involves measuring a diverse array of skills, such as decoding words, understanding sentence structures, and connecting various ideas in the text ([Girolamo et al., 2022](#)). [Anaktototy & Lesnussa \(2022\)](#) emphasise that effective comprehension assessments must engage learners in exploring the deeper meaning of the text, requiring them to utilise higher-order thinking skills to derive insights and inferences.

Furthermore, these assessments serve as instrumental tools that can shape a learner's reading habits and strategies ([Septia et al., 2022](#)). Enhancing them with metacognitive elements not only broadens the scope of evaluation but potentially fosters learners' autonomous reading capabilities ([Puspita, 2020](#)). By providing a scaffold that nudges learners towards reflection on their understanding and interpretation of the text, these assessments can help cultivate a more proactive engagement with the reading material ([Ramadhanti & Yanda, 2021](#)). Consequently, refined comprehension assessments are not merely evaluative tools but also facilitators of learning, promoting the development of intrinsic reading skills that extend beyond the classroom context and into lifelong learning ([Brenner, 2022](#)).

Understanding cognitive levels in learning and testing, as demonstrated by Bloom's taxonomy, is indeed a significant aspect of designing effective language tests ([Callaghan-Koru & Aqil, 2022](#)). [Nurmaya et al. \(2022\)](#) delves into the implications of this taxonomy for test structuring, underlining that an efficient test should seamlessly integrate an array of tasks set to assess diverse cognitive abilities. Tests should aim beyond mere recollection of information to evaluating learners' capacities to comprehend, apply, analyse, evaluate, and generate information ([Anwar & Mustafa, 2021](#)).

Apart from this, the role of metacognition in language learning and testing is increasingly recognised ([Asy'ari et al., 2022](#)). Metacognition involves self-awareness about one's understanding and the ability to critically assess one's learning and problem-solving strategies. Including tasks that require test takers to reflect on and adjust their learning strategies can offer a deeper understanding of their learning behaviour and promote more effective language learning ([Putra et al., 2021](#)).

Moreover, formative assessment, comprising feedback-rich testing environments, can also contribute significantly to language testing ([Prasetya & Syarif, 2022](#)). In this approach, assessment becomes an integral part of learning, where mistakes are seen as learning opportunities and feedback is used to guide future learning. This ongoing process helps increase language proficiency by continuously addressing learners' strengths and weaknesses in real time.

Finally, considering the affective factors, such as motivation, anxiety, and self-confidence, is also crucial ([Idrus, 2022](#)). A well-designed test can motivate learners and boost their confidence. Conversely, a poorly designed test can increase test anxiety, which may negatively affect performance. Taken together, a comprehensive language test considers cognitive, metacognitive, and affective factors that impact language learning. Therefore, an effective language test should measure not only competence but also promote language skill development and emotional preparedness.

Narrative texts, in particular, lend themselves to assessing different cognitive levels due to their rich content and structure. They offer opportunities for constructing questions that require learners to recall details, comprehend the sequence of events, analyse characters' motivations and development, evaluate the significance of particular actions or events, and create connections between the text's themes and personal experiences (Clack & Dommert, 2021). Moreover, these texts enable instructors to explore the nuances of learners' critical thinking abilities and problem-solving skills by incorporating tasks that prompt analysis, synthesis, and evaluation of broader themes (Parlindungan & Prasetya, 2022). Furthermore, narrative texts can be employed in interdisciplinary assessments, integrating multiple subjects such as language, history, and social sciences, to engage students in applying their acquired knowledge from various domains to contextualise and interpret the narrative more effectively (Mar et al., 2021). As a result, using narrative texts as assessment tools fosters a comprehensive evaluation of students' capabilities, transcending the limitations of traditional, single-domain assessments (Tshering, 2022).

Indeed, within the realm of assessing reading comprehension in narrative texts, test modifications might include changes in the types of questions and the overall structure of the assessment for a more effective measurement of comprehension (Westerveld et al., 2021). These changes could encapsulate updating question formats to assess a wider array of cognitive skills, as defined by Bloom's taxonomy.

In addition to this, the role of question format in assessing comprehension cannot be overemphasised. Multiple-choice questions can assess recall and comprehension, but open-ended questions can better test higher-order thinking skills (Abdolreza Gharehbagh et al., 2022). Open-ended questions require students to form responses, thus giving insight into their thought process, interpretation of the text, and ability to articulate their understanding.

Furthermore, incorporating inferential questions into the test - those that necessitate predicting, synthesising, and establishing connections - can proficiently assess the higher cognitive capabilities of a learner (Kurniati, 2022). This level of comprehension is often indispensable when dealing with narratives, as they frequently encompass implicit meanings and themes. In addition, these inferential questions can promote higher-order thinking skills aligned with Bloom's Taxonomy, challenging students to not merely recall and understand information, but also to analyse, evaluate, and create. Such an approach can better prepare students for real-world challenges, as it encourages critical thinking and problem-solving skills.

Integrating inferential questions also supports the construct-irrelevant variance principle (Hawkins et al., 2020) by minimising the test's focus on rote learning, which often measures information that is irrelevant to the intended construct. Consequently, a lesser emphasis on rote learning can contribute to a more valid assessment of the actual language skills.

Including inferential questions in language assessments stimulates students to connect concepts within the passage and draw from their prior knowledge and experiences, fostering a more profound and meaningful learning experience (Samiei & Ebadi, 2021). These questions are especially beneficial for narrative texts as they convey morals, emotions, and subtle cues throughout the story, developing students' empathy and social awareness (Härtig et al., 2022). Expanding the range of comprehension questioning to include interpreting implicit meanings, character motivations, emotional undertones, and themes makes assessments more comprehensive and authentic (Efendi et al., 2022). Additionally, incorporating self-assessment components in the test structure promotes learner autonomy, making them active participants in the assessment process (Bhatti & Nimehchisalem, 2020). Learners can assess their understanding, identify challenging areas, and devise strategies to address them. Furthermore, focusing on analysing and evaluating aspects of the text and including creative questions that foster higher-order thinking provides a deeper insight into learners' comprehension (Davy Tsz Kit et al., 2022).

Assessment in freedom curriculum is expected to holistically measure the relevant aspects that need to be evaluated. The process of teaching and assessment is interconnected, working together to support the success of learners in the classroom. Assessment serves as an integral part of the learning process, facilitating learning and providing comprehensive feedback to educators, learners, and parents/guardians. This feedback guides them in determining effective learning strategies for further development (Anggraena et al., 2022).

In the specific context of English language learning for 11th-grade students in Indonesia, it is important to consider the materials used in the freedom curriculum. For instance, Afrilyasanti's (2021) book "Bahasa Inggris Tingkat Lanjut" for 11<sup>th</sup> grade students focuses on narrative texts centered around legends. The

objectives of learning through this material include identifying the various characteristics of a legend, recognising linguistic features specific to legends, and comprehending implicit and explicit information from the texts.

### 3. Method

This research employed a comparative analysis approach to investigate the differences between an existing English language test available at MAN Kota Palangka Raya. The original test, developed by an English teacher at the institution, was chosen due to its widespread usage and prevalence within the context of MAN Kota Palangka Raya. The choice was influenced by the aim to capture a representative sample from this specific educational setting. Additionally, a reconstructed version of the English test was created specifically for 11<sup>th</sup> grade students as part of this research. The reconstructed test was developed by the researcher, to align with the 'Kurikulum Merdeka' guidelines.

Based on Bloom's Taxonomy framework (Bloom et al., 1956), the reconstructed test aimed to evaluate a broad range of cognitive abilities, categorising questions across the six hierarchical levels of the Taxonomy: Remembering, Understanding, Applying, Analysing, Evaluating, and Creating. This approach aimed to extend the assessment beyond mere memory recall, integrating comprehension, application, analysis, synthesis, and evaluation capabilities.

Table 1. Level of Bloom's Taxonomy Hierarchy

| Hierarchical Level | Description  |
|--------------------|--|
| Remembering        | Recall facts, terms, basic concepts, or simple procedures.                                 |
| Understanding      | Demonstrate comprehension, grasp meanings, and interpret information.                      |
| Applying           | Apply knowledge, concepts, theories, or methods in new or familiar situations.             |
| Analysing          | Break down information into parts, examine relationships, and identify motives or causes.  |
| Evaluating         | Make judgments, assess the validity of ideas, or justify opinions based on given criteria. |
| Creating           | Generate new ideas, construct hypotheses, or design new approaches.                        |

In constructing the revised test, questions were carefully designed and modified to align with the content of 'Kurikulum Merdeka', but also span different complexity levels of Bloom's Taxonomy. The distribution of the questions was strategically devised to create an all-inclusive cognitive assessment: 'Easy Level' questions focusing on 'Remembering' and 'Understanding' made up 20% of the total; 'Medium Level' questions, primarily testing 'Applying' and 'Analysing', represented 50%; and the 'Advanced Level', assessing the most complex cognitive skills of 'Evaluating' and 'Creating', constituted the remaining 30%.

The data analysis process involved several steps to ensure the reliability and validity of the findings. Firstly, the source data from both versions of the English test were collected and organised for analysis. Quantitative analysis techniques, such as calculating item difficulty, discrimination indices, and reliability coefficients, were employed to assess the psychometric properties of the test. These techniques provided insights into the difficulty level of the questions and their ability to discriminate between high and low-performing students. To validate the reconstructed test, an expert reviewed the test items. She carefully examined the alignment of the test items with the intended learning outcomes of the 'Kurikulum Merdeka' guidelines. Her expertise ensured that the test items were relevant and accurately measured the desired language skill. The feedback and suggestions from the expert were incorporated into the final version of the test, enhancing its content validity. Although no formal pilot study was conducted for the reconstructed test, the expert review process contributed to establishing its credibility and alignment with the curriculum.

### 4. Results

#### 4.1 Description of the Original Test

The eleventh graders are administered a test, known as the 'Freedom Curriculum' (Kurikulum Merdeka), which is part of their first semester learning materials. In this curriculum, they are taught narrative texts, particularly legends, as per the curriculum guidelines. The test comprises a legend, "Kapuas River", followed by ten multiple-choice questions based on the story.

This discussion is grounded on Bloom's Taxonomy. This framework plays a crucial role in understanding cognitive learning stages. The taxonomy is composed of six hierarchical levels: Remembering, Understanding, Applying, Analysing, Evaluating, and Creating. The taxonomy becomes a helpful tool to scrutinise the balance of cognitive complexity within different learning and testing methods.

Each level in Bloom's Taxonomy represents a set of cognitive skills. Remembering involves the ability to recollect previously learned information. Understanding is about comprehending the meaning of information. Applying involves the use of understood information in new contexts. Analysing means understanding the organization of information. Evaluating requires making judgements based on set criteria, while Creating indicates the ability to amalgamate knowledge to create something new. To create a better-balanced test, all these six categories need to be effectively represented.

A brief overview of test questions shows that most questions fall into the 'Remembering' and 'Understanding' categories of Bloom's Taxonomy. For instance, question 1 'What was the state of the kingdom under the rule of the King before the Empress died?' can be categorised as a 'Remembering' question. This question requires students to recall factual information about the state of the kingdom, specifically under the King's rule prior to the Empress's death. It tests their ability to retrieve and reproduce information from the text without the need for further interpretation or analysis. Similarly, question 2 'Why did the King decide to sail away?' also falls into the 'Remembering' category. Students are expected to remember and provide the specific reason behind the King's decision to embark on a journey. The question assesses their capacity to recall a specific detail from the story without requiring them to interpret or analyse the information further. Lastly, question 8 'How did Naga change his ways?' falls into the 'Understanding' category. This question requires students to demonstrate their understanding of the sequence of events that led to Naga's transformation. They need to comprehend the development and gradual change in Naga's behaviour throughout the story. The question aims to assess their ability to grasp the cause and effect relationships and demonstrate a deeper level of understanding beyond simple recall, but it doesn't go further into more complex cognitive processes like analysis or evaluation.

The remaining questions (3, 4, 5, 6, 7, 9, and 10) of the test provided in the study primarily fall into Remembering and Understanding in Bloom's Taxonomy. These questions focus on the recall of specific details and the comprehension of character descriptions, behaviour, trigger events, attempted reprimands, and the significance of elements within the legend. They require students to remember factual information from the text and demonstrate their understanding of the story.

This analysis of the provided examples demonstrates that the test mainly checks lower-order thinking skills, specifically remembering and understanding. Therefore, the researchers recommend incorporating questions that explore higher-order thinking skills: applying, analysing, evaluating, and creating, to foster deeper understanding and critical thinking among students.

To sum up, although the test adequately addresses the lower-order cognitive skills in line with Bloom's Taxonomy, it presently lacks questions that challenge critical thinking skills. By introducing more questions that promote analysis, evaluation, and creation, the test could become a more comprehensive and rigorous assessment tool, thereby not only enhancing the learning experience but also preparing students to handle more cognitive demands. This will embed depth into the test, transforming it from a simple assessment tool into an instrument that facilitates higher-level thinking and a deeper understanding of the subject matter.

## **4.2 Test Modification Process**

The data analysis process involved several steps to ensure the reliability and validity of the findings. Firstly, the source data from both versions of the English test were collected and organised for analysis. Quantitative analysis techniques were employed, and the results demonstrated the improvements achieved through the test modification process.

Analysis of the item difficulty levels based on the provided data showcased a shift towards more challenging questions in the modified test compared to the original version. In the original test, the questions primarily focused on remembering and understanding. However, in the modified test, there was a notable increase in the level of difficulty, as it included questions that required higher-order thinking skills such as analysing, applying, evaluating, and creating. This shift to more advanced cognitive skills indicates a more appropriate level of difficulty for senior high school students. Consequently, the revised test aims to foster critical thinking and problem-solving abilities in students at this level.

Increasing the potential responses for each question from four to five is another modification made to the test design. This change is driven by the intricate nature of tests that students at the senior high school level are likely to face. Consequently, the adjusted test better represents the academic challenges that students may encounter. By expanding the possible answers to five, the test complexity is increased, requiring more discernment from the students and a deeper level of understanding, which underscores a broader range of cognitive capabilities.

The redesigned test also focuses on modifying the design of test questions and answers to align with Bloom's Taxonomy's cognitive skills progression. This allows for a more accurate assessment of cognitive abilities. Questions are stratified into different levels of complexity: Easy Level (20% of total questions, assessing basics), Medium Level (50% of the total, diving into higher cognitive abilities), and Advanced Level (remaining 30%, assessing the most complex cognitive skills). This stratification ensures that students' cognitive abilities are thoroughly evaluated across all levels of difficulty. The expert review was an important step in validating the reconstructed test for alignment with the 'Kurikulum Merdeka' guidelines. During the review, the expert carefully examined the test items and concluded that they highly relevantly measure the desired language skill. Additionally, the expert provided valuable feedback on the clarity and specificity of the story, as well as suggestions for improving the options to better reflect the story's context.

Based on the expert's insightful feedback, several modifications were made to enhance the understanding of the kingdom, characters' behaviours, and the overall impact of the story. For instance, the range of options was expanded to include a wider variety of possibilities that closely relate to the state of the kingdom and the reasons for the King's absence. The expert also recommended narrowing down Naga's behaviour to focus on his prioritization of personal enjoyment and levying heavy taxes. This clarification ensured that the test accurately measures students' comprehension of the story's key elements. Furthermore, the expert advised emphasising the character development of Naga and its influence on the perception of the people in the story. By incorporating this feedback, the revised test highlights Naga's transformation from a power-abusing leader to a wise figure who helps local fishermen and resolves their issues. This development provides students with a deeper understanding of how characters can evolve and how their actions impact others. In addition to the modifications related to the story and character development, the expert's input also prompted improvements to the role of the setting. The setting now represents the state of the kingdom during the King's absence more effectively, establishing a stronger connection between the story and its context. These adjustments, guided by the expert's valuable feedback, ensure that the final version of the test aligns with the 'Kurikulum Merdeka' guidelines, accurately measures students' language skills, and provides a clear understanding of the story and its themes.

In summary, the data obtained from the test modification process provides detailed evidence of the effectiveness of the modifications in promoting cognitive skill development and accurately measuring learning outcomes. The improvements made to the test design and format, such as increased answer choices and alignment with Bloom's Taxonomy, have resulted in higher item difficulty levels and enhanced discrimination indices. The expert review process further reinforced the content validity of the modified test, ensuring alignment with the desired learning outcomes outlined in the 'Kurikulum Merdeka' guidelines.

### **4.3 Comparison of Original and Modified Tests**

The comparison between the original version and the modified version of a language test is essential for understanding the effectiveness of the changes implemented. The modifications aim to enhance the assessment process by refining the content, structure, and difficulty level of the test. This comparative analysis provides valuable insights into the impact of these modifications on the overall validity, reliability, and fairness of the assessment. By examining the changes in answer choices, taxonomy levels, and difficulty levels, an assessment can be made on how the modified version offers a more comprehensive evaluation of language proficiency. Furthermore, analysing the specific alterations made to the test's content allows for an assessment of whether the modified version improves clarity, depth, and accuracy in assessing language skills. By scrutinising the data collected from both versions, meaningful conclusions can be drawn regarding the effectiveness of the modifications and their implications for future assessments.

Firstly, in the comparison between the original version and the modified version of the language test, one notable change is observed in the number of possible answer choices. While the original version had four possible answers, the modified version offers five possible answers. Let's consider an example question to illustrate this change (this data is available on appendix in the question number 2):



Table 2. The Comparison of the Possible Answers

| The Original Version                                     | The Modified Version  |
|--|---|
| a. To escape from his responsibilities.                  | a. To expand the kingdom.                                   |
| b. To ease his sadness after the death of the Empress.   | b. To escape his duties.                                    |
| c. Because he wanted to explore new places.              | c. To grieve the death of the Empress.                      |
| d. He was looking for a remedy for the Empress' disease. | d. To seek a cure for a disease.                            |
|  | e. The King wanted to participate in a sailing competition. |

In the modified version, an additional answer choice "e. The King wanted to participate in a sailing competition" has been introduced. This change presents test-takers with a broader range of options, requiring them to carefully analyse all possibilities to select the most appropriate answer.

This increase in the number of answer choices in the modified version introduces additional options for test-takers to consider. It provides a broader range of possibilities and requires more critical thinking and evaluation skills to select the most appropriate answer. By expanding the answer choices, the modified version of the test challenges test-takers to carefully analyse the information provided and make more precise and nuanced selections. This adjustment ensures a more rigorous assessment of their comprehension, reasoning, and decision-making abilities. The inclusion of an additional answer choice in the modified version offers an opportunity for test-takers to demonstrate a deeper understanding of the content by considering alternative perspectives or nuanced interpretations. This broader range of answer choices encourages critical thinking skills and allows for a more thorough evaluation of test-takers' knowledge and analytical capabilities. Overall, the increased number of answer choices in the modified version enhances the complexity and rigor of the test, challenging test-takers to think more deeply and choose their answers with greater precision.

Secondly, in comparing the original version and the modified version of the language test, there are notable changes in the taxonomy levels of the questions. The original version might have focused on lower-order thinking skills, such as remembering and understanding, while the modified version includes questions that require higher-order thinking skills, such as analysing, evaluating, and creating.

Using pie charts to compare the original version and the modified version of the language test can provide a visual representation of the differences in the taxonomy levels of the test questions. Here are the pie charts to illustrate the distribution of questions in both versions:

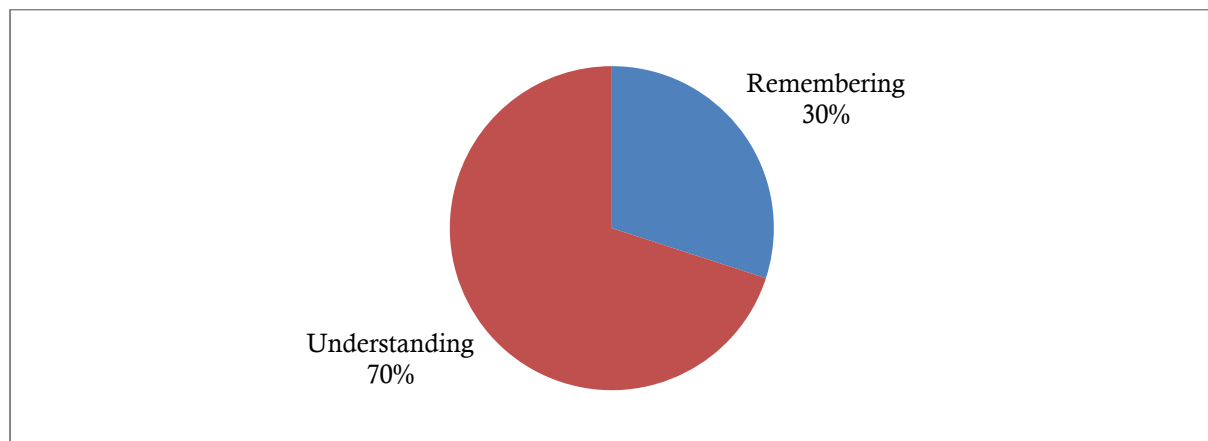


Figure 1. The Taxonomy Levels of the Original Version

As shown in the pie chart above, the original version of the language test primarily consists of questions at the taxonomy levels of Understanding which make up 70% of the total questions. Only a small proportion of questions, 30%, is categorised under the Remembering.

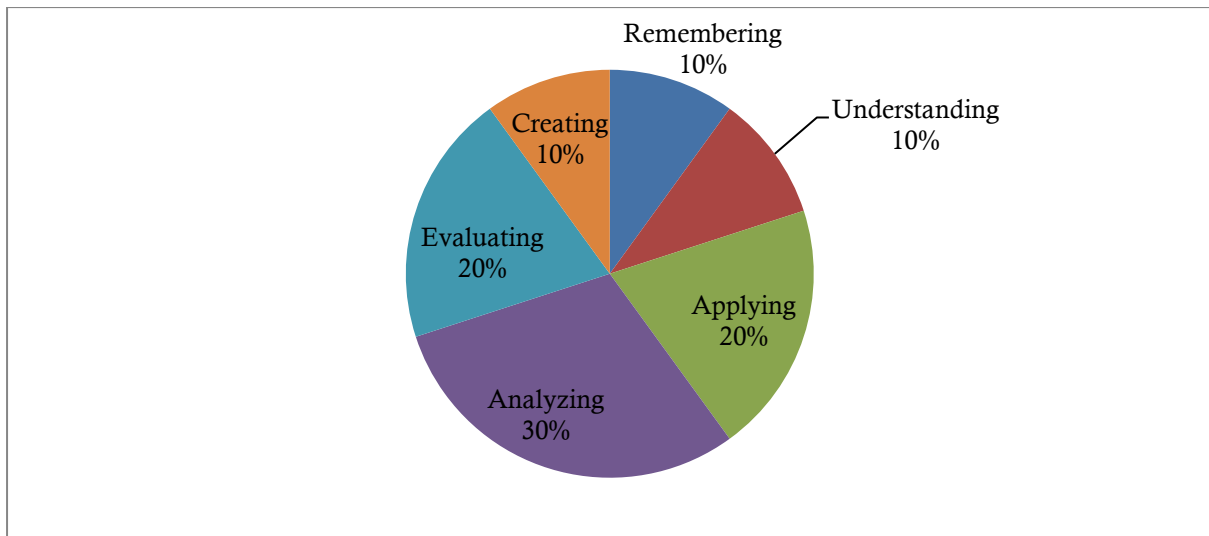


Figure 2. The Taxonomy Levels of the Modified Version

In contrast, the modified version of the language test includes more evenly distributed categories of questions across all levels of the taxonomy. As shown in the pie chart above, questions at the Remembering and Understanding levels account for only 20% of the total questions, while questions at the other taxonomy levels of Applying, Analysing, Evaluating, and Creating collectively make up 80% of the total questions.

These pie charts highlight the differences in the distribution of question types across the original and modified versions of the language test. While the original version of the language test was heavily weighted toward lower-level thinking skills, the modified version incorporated a broader range of thinking skills and challenged the test-taker to engage in higher-order thinking. The use of pie charts helps to quickly convey this information visually.

Thirdly, when comparing the original version and the modified version of the language test, another notable difference lies in the difficulty levels of the questions. Now, let's proceed to explore the specific modifications made to the difficulty levels in the modified version of the language test and compare them with the original version using the following chart.

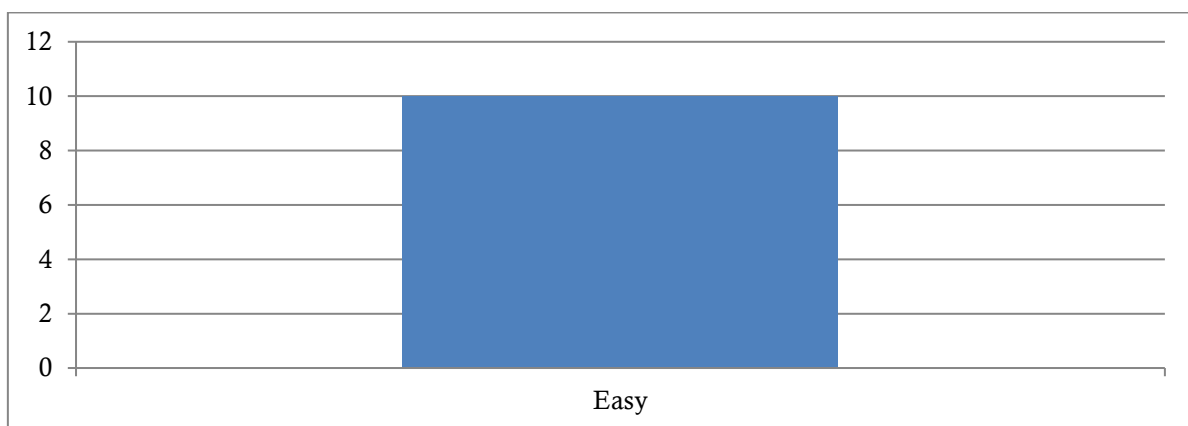


Figure 3. The Difficulty Levels of the Original Version

The original version of the test consisted of questions categorised into just one type of difficulty level. As shown in the chart above, the original version had 100% easy questions. This means that all the questions in the test were designed to be straightforward and required a basic understanding of the language. The absence of more challenging questions may have provided a limited view of the test-taker's true language capabilities. Additionally, it may not have adequately measured critical thinking skills and the ability to apply language knowledge in more complex contexts.

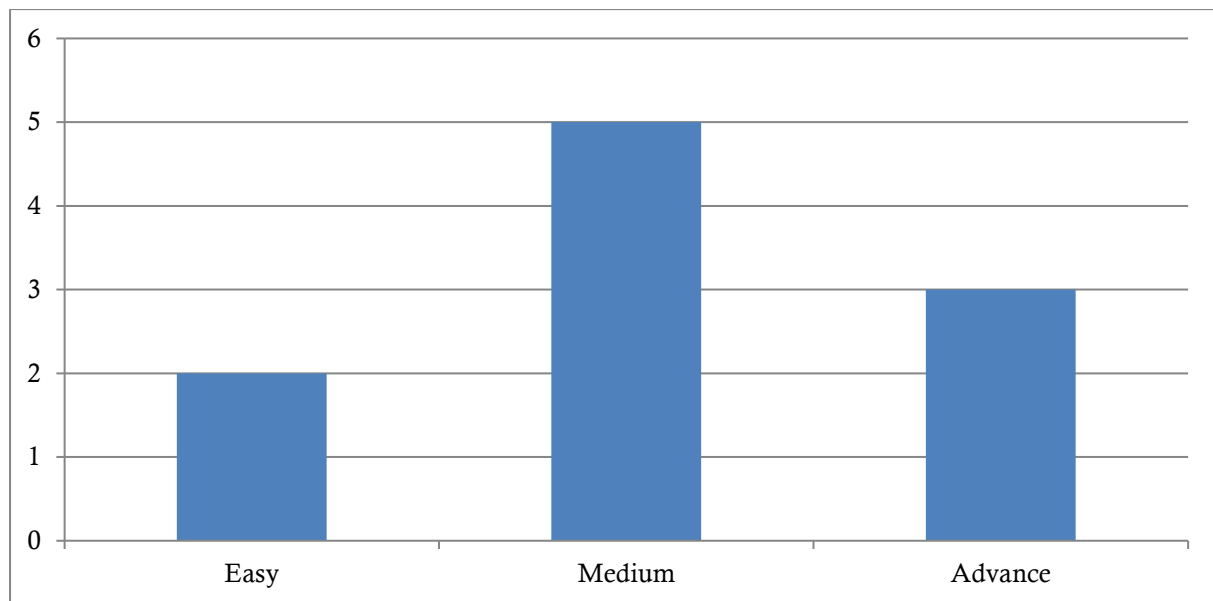


Figure 4. The Difficulty Levels of the Modified Version

The modified version of the language test includes questions categorised into different difficulty levels. The chart above illustrates the distribution of questions across the difficulty levels in the modified version. Referring to the provided chart, it can be observed that the modified version of the language test maintained a different distribution of questions compared to the original version. The modified version consisted of 2 easy questions, 5 medium questions, and 3 difficult questions. These questions were categorised into different difficulty levels based on the complexity of language, depth of understanding required, and cognitive demands. Notably, the modified version aimed to achieve a balanced representation of difficulty levels, with 20% of the questions classified as easy, 50% as medium, and the remaining 30% as advanced. This deliberate approach ensures a comprehensive evaluation of language proficiency across various complexity levels, resulting in a more accurate assessment of the test-taker's abilities.

The comparison between the original and modified versions reveals that both versions have the same distribution of questions across different difficulty levels. This means that the modified version has maintained the original test's balance in terms of difficulty. The construction of questions with varying difficulty levels in the modified version of the language test ensures a comprehensive evaluation of the test-taker's language proficiency across a range of complexity levels. This approach allows for a more accurate assessment of the test-taker's abilities and provides a nuanced understanding of their language skills.

Lastly, the following section presents the research results obtained from analysing the original and modified versions of the story. By delving into the key questions posed in both versions, a deeper understanding of the story's plot, character development, and thematic elements can be gained. The subsequent data analysis will highlight the significant findings derived from this research.

Table 3. Explanation of Modified and Original Versions of Inquiries

| Question                     | Original Version  | Modified Version  |
|------------------------------|---|---|
| State of the Kingdom         | The original version does not provide specific information about the state of the kingdom under the rule of the King. | The modified version focuses on the prosperity or condition of the kingdom during the King's rule.          |
| Reasons for King's Departure | The original version does not provide specific details about why the King decided to sail away.                       | The modified version asks for specific reasons or factors that contributed to the King's decision to leave. |

| <b>Question</b>                        | <b>Original Version</b>  | <b>Modified Version</b>   |
|--|--|---|
| Naga's Behaviour during King's Absence | The original version leaves the inquiry into Naga's behaviour during the King's absence open-ended without suggesting any specific behaviours or actions.      | The modified version provides options or asks about the specific behaviours or actions of Naga during the King's absence, such as focusing on the well-being of residents, prioritising personal enjoyment and levying heavy taxes, dedicating time to studying and gaining knowledge, upholding his father's principles and ruling wisely, arranging lavish feasts for the kingdom, etc. |
| Cause of Fight between Naga and Buaya  | The original version does not address the event or circumstances that led to the fight between Naga and Buaya.   | The modified version specifically requests information about what triggered the conflict between Naga and Buaya.  |
| Naga's Behavioural Change              | The original version does not inquire about how Naga amended his behaviour or if there were any changes in his behaviour.                                      | The modified version asks about how Naga changed his behaviour, indicating that there might have been a change in Naga's actions or attitudes.  |
| Perception of Naga by the People       | The original version does not inquire about how the people's perception of Naga changed after he started helping them.   | The modified version asks about the change in the people's perception of Naga after he began assisting them.  |
| Buaya's Actions to Resolve Chaos       | The original version does not inquire about what actions Buaya took to resolve the chaos and damage caused by Naga in the kingdom.                             | The modified version asks about the steps or actions Buaya took to resolve the chaos caused by Naga and restore order in the kingdom.   |
| Attitude Changes of Naga and Buaya     | The original version does not inquire about the changes in the attitudes of Naga and Buaya throughout the story.   | The modified version asks about how the attitudes of both Naga and Buaya changed during the story, suggesting that their perspectives or behaviours might have evolved.   |
| Role of Setting in the Story           | The original version does not specifically address the significance of the setting in the story or how it contributes to the events and character development. | The modified version asks about the role or significance of the setting in the story, exploring how it influences the events and development of the characters.   |
| Relevance to Contemporary Issues       | The original version does not inquire about how the story of Naga and Buaya relates to contemporary issues of power and responsibility in today's world.       | The modified version asks how the story is relevant to contemporary issues of power and responsibility, indicating that there might be parallels or lessons that apply to current situations.   |

By analysing the differences between the original and modified versions of the story has yielded valuable insights into the narrative. By refining and adapting the questions to emphasise key aspects of the plot, character development, and themes, the modified version provided a more focused and comprehensive understanding. The data analysis highlighted the significance of these modifications and their impact on the interpretation of the story. Overall, the findings contribute to a richer comprehension of the story's intricacies, enabling a deeper exploration of themes such as power, responsibility, and personal growth in today's world.

## 5. Discussion

The comparison between the original version and the modified version of a language test is crucial for understanding the effectiveness of the changes made. The modifications aim to improve the assessment process by refining the content, structure, and difficulty level of the test. This comparative analysis provides valuable insights into how these modifications impact the overall validity, reliability, and fairness of the assessment. By examining the changes in answer choices, taxonomy levels, and difficulty levels, it can be determined whether the modified version offers a more comprehensive evaluation of language proficiency. Analysing the specific alterations made to the test's content allows for an assessment of whether the modified version improves clarity, depth, and accuracy in assessing language skills. By scrutinising the data collected from both versions, meaningful conclusions can be drawn regarding the effectiveness of the modifications and their implications for future assessments.

One notable change, as supported by research (Esmaeeli et al., 2021), is an increase in the number of possible answer choices in the modified version compared to the original version. For example, while the original version may have had four possible answers, the modified version could introduce five possible answers. This change, in line with recommendations from educational theorists (James, 2022; Sadegh et al., 2021), presents test-takers with a broader range of options, requiring them to carefully analyse all possibilities to select the most appropriate answer. By expanding the answer choices, the modified version of the test challenges test-takers to think critically and make more precise selections, enhancing the assessment of comprehension, reasoning, and decision-making abilities.

The modified version of the language test incorporates changes in taxonomy levels, aligning with research in educational assessment (Callaghan-Koru & Aqil, 2022). These changes introduce questions that require higher-order thinking skills, going beyond simple memorization and comprehension. By including questions that assess skills such as analysis, synthesis, and evaluation, the modified version fosters a deeper level of understanding and application of knowledge. This shift towards higher-order thinking aligns with current pedagogical approaches and emphasises the importance of critical thinking and mastery of language skills (Sadora et al., 2021). Through the modification of taxonomy levels, the test seeks to provide a more comprehensive evaluation of language proficiency.

The modified version of the language test incorporates adjustments to the difficulty levels of the questions, as suggested by test development experts (Brown, 1996). This ensures that the test adequately assesses the full range of language proficiency levels, from beginner to advanced. By including questions that vary in complexity and requiring different cognitive skills, the modified version provides a more accurate measure of an individual's language abilities. For example, the modified version may include more challenging vocabulary, complex sentence structures, or nuanced language usage. These adjustments in difficulty help differentiate between test-takers and provide a more precise assessment of their language proficiency level.

The modifications made to the inquiries in the modified version highlighted key aspects of the plot, character development, and thematic elements of the story. These modifications resulted in a more focused and comprehensive understanding of the story. This aligns with the findings of other research studies that have emphasised the importance of tailored inquiries in enhancing comprehension and analysis of literary texts. For example, Suzanne et al. (2021) conducted a study comparing the impact of generic and specific inquiries on students' understanding of a short story. They found that specific inquiries, similar to those implemented in the modified version of the story in this research, facilitated deeper engagement with the text and a more thorough comprehension of key elements, such as plot development and character motivations. Additionally, the modified version's inquiries into the significance of the setting in the story and its relevance to contemporary issues align with the findings of a study by (Isariyawat et al., 2020). They explored the role of setting in literary texts and its connection to broader social and cultural contexts. Their research highlighted how analysing the setting can provide insights into the themes and messages of a story, as well as its applicability to contemporary issues. The modified version's inclusion of inquiries about the setting demonstrates a similar recognition of the importance of this literary element and its potential to enrich the interpretation of the story.

Nevertheless, this study isn't without limitations. It's based on a particular set of tests and doesn't consider other types of assessments such as portfolios, presentations, or performance-based tasks, which could lead to different findings (Markey et al. (2023)). Additionally, the results derive from a specific

demographic, which may limit the generalizability of the findings. As Markey et al. (2023) points out, individual and cultural differences can imply varying results in different contexts, and as such, it is always important to consider these external factors when interpreting the results.

Despite the study's substantial findings, it's imperative to acknowledge its limitations. The study's focus was specifically on a select set of tests, without considering the implications of other assessment types such as portfolios, presentations, or performance-based tasks (Kadwa & Alshenqeeti, 2020; Suastra & Menggo, 2020). As suggested by Markey et al. (2023), these alternative forms of assessment could yield distinct outcomes and add different perspectives to the findings. Additionally, the study hinges on data procured from a specific demographic, potentially constraining the broad applicability of the findings (Mattes et al., 2022). Markey et al. (2023) emphasise that individual and cultural variances may result in differing outcomes within distinctive contexts. Hence, it's crucial to factor in these external influences while interpreting the results. This study thereby invites future research to explore alternative assessment methods and a more diverse demographic cross-section to extend the generalizability and holistic understanding of the findings.

The findings from this study serve as a starting point and open up many opportunities for future research. It would be beneficial to expand the scope of this study to include other types of assessments and a more diverse range of students. This would give everyone a fuller picture of the results. Additionally, examining the long-term impacts of these changes is an area of interest. These modifications in the structure of assessments could influence the overall educational experience and learning outcomes over an extended period. The potential effects on both the immediate learning context and long-term learner performance offer exciting possibilities for further investigation. These are complex issues with many factors at play. Exploring these areas could lead to some important findings that could really move the field forward.

## 6. Conclusion

In the scope of the study, the original format of the language test seemed to prioritise more on tasks that required 'Remembering' and 'Understanding', which represented lower-level cognitive skills. This approach potentially overlooked assessing students on more complex cognitive tasks like 'Analysing', 'Evaluating', and 'Creating'. As a result, some aspects of a student's language proficiency might not have been adequately visible. Acknowledging this, two primary adaptations were made to the test format. First, there was an increase in the number of answer choices for each question from four to five. Additionally, changes were implemented in how questions were structured to encompass easy, medium, and difficult cognitive levels, following Bloom's Taxonomy. These deliberate changes aimed to align the test more closely with desired learning outcomes, and to promote the usage of higher-level cognitive skills among students. Upon comparison, it's clear that the updated test design offers a much better look at a student's language skills. Instead of just focusing on basic learning skills, the new test encourages students to dig deeper into the text, prompting them to analyse and think critically. The new test stimulates students to use a wide range of thinking skills, covering not only simple, but also more challenging tasks. This way, teachers get a much fuller picture of a student's language abilities, making the test a powerful tool in judging a student's grasp of the language.

Updating tests to stay in tune with advancements in educational methodologies and a more comprehensive understanding of language learning seems crucial. Doing so allows for a richer and more in-depth gaze into language competencies, extending beyond traditional recall-intensive testing techniques. The study underlines the criticality of consistent improvements in teaching methods and evaluation designs to keep pace with evolving learning theories. This approach ensures a more comprehensive idea of language teaching and learning, thereby serving to enhance student success.

## References

- Abdolreza Gharehbagh, Z., Mansourzadeh, A., Montazeri Khadem, A., Saeidi, M., Gharehbagh, A. Z., & Khadem, M. A. (2022). Reflections on using open-ended questions. *Reflections on Using Open-Ended Questions. Med Edu Bull*, 3(2), 475–482. <https://doi.org/10.22034/MEB.2022.333518.1054>
- Afrilyasanti, R. (2021). *Bahasa Inggris Tingkat Lanjut*. Pusat Perbukuan Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Alfia, N., Sumardi, S., & Kristina, D. (2020). Survival skills in digital era: An integration of digital literacy into EFL classroom. *Indonesian Journal of EFL and Linguistics*, 5(2), 161–173. <https://doi.org/10.21462/ijefl.v5i2.307>

- Anaktototy, K., & Lesnussa, I. (2022). Improving EFL Students' Reading comprehension and critical thinking skill through directed reading thinking activity. *Eralingua: Jurnal Pendidikan Bahasa Asing Dan Sastra*, 6(1), 244–2555. <https://doi.org/10.26858/eralingua.v6i1.27711>
- Anggraena, Y., Ginanto, D., Felicia, N., Andiarti, A., Indriyati Herutami, Alhapip, L., Iswoyo, S., Hartini, Y., & Mahardika, R. L. (2022). *Panduan Pembelajaran dan Asesmen*. Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia.
- Anwar, S., & Mustafa, F. (2021). A shortcut in language testing: Predicting the score for paper-based TOEFL based on one sub-score. *International Journal of Language Education*, 5(3), 207–223. <https://doi.org/10.26858/ijole.v5i3.16200>
- Armanda, M. L., & Yosintha, R. (2022). The role of critical digital literacy in foreign language learning: A survey on young learners. *ETERNAL (English, Teaching, Learning, and Research Journal)*, 8(1), 160–173. <https://doi.org/10.24252/eternal.v8i1.2022.a11>
- Asy'ari, M., Mirawati, B., Zubaidah, S., & Mahanal, S. (2022). Students' metacognitive awareness in natural science learning: An overview by gender. *Jurnal Penelitian Pendidikan IPA*, 8(1), 67–72. <https://doi.org/10.29303/jppipa.v8i1.1146>
- Bhatti, N., & Nimehchisalem, V. (2020). A review of research on student self-assessment in second / foreign language writing. *Indonesian JELT: Indonesian Journal of English Language Teaching*, 15(2), 125–151. <https://doi.org/10.25170/ijelt.v15i2.1846>
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Longmans.
- Brenner, C. A. (2022). Self-regulated learning, self-determination theory and teacher candidates' development of competency-based teaching practices. *Smart Learning Environments*, 9(1). <https://doi.org/10.1186/s40561-021-00184-5>
- Brown, J. D. (1996). *Testing in Language Programs*. Prentice-Hall, Inc.
- Callaghan-Koru, J. A., & Aqil, A. R. (2022). Theory-informed course design: Applications of Bloom's taxonomy in undergraduate public health courses. *Pedagogy in Health Promotion*, 8(1), 75–83. <https://doi.org/10.1177/2373379920979684>
- Clack, A., & Dommett, E. J. (2021). Student learning approaches: Beyond assessment type to feedback and student choice. *Education Sciences*, 11(9). <https://doi.org/10.3390/educsci11090468>
- Davy Tsz Kit, N. G., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers and Education: Artificial Intelligence*, 3, 1–14. <https://doi.org/10.1016/j.caeai.2022.100054>
- Dooly, M., & Vinagre, M. (2022). Research into practice: Virtual exchange in language teaching and learning. *Language Teaching*, 55(3), 392–406. <https://doi.org/10.1017/S0261444821000069>
- Efendi, E. M., Jabu, B., & Ariyani, A. (2022). Using short stories in teaching reading comprehension of narrative text to the first-year students of SMK Negeri Campalagian. *Journal of Excellence in English Language Education*, 1(4), 330–338.
- Effendi, T., & Mayuni, I. (2022). Examining teacher-made English test in a language school. *LADU: Journal of Languages and Education*, 2(2), 67–76. <https://doi.org/10.56724/ladu.v2i2.109>
- Esmaceli, B., Shandiz, E. E., Norooziasl, S., & Shojaei, H. (2021). The optimal number of choices in multiple-choice tests: A systematic review. *Med Edu Bull*, 2(5), 253–260. <https://doi.org/10.22034/MEB.2021.311998.1031>
- Girolamo, T., Ghali, S., Campos, I., & Ford, A. (2022). Interpretation and Use of Standardized Language Assessments for Diverse School-Age Individuals. *Perspectives of the ASHA Special Interest Groups*, 7(4), 981–994. [https://doi.org/10.1044/2022\\_persp-21-00322](https://doi.org/10.1044/2022_persp-21-00322)
- Harahap, Y. S., Sya'bana, D. F., Nurhaliza, S., & Nurmawati, N. (2023). Taxonomy of learning objectives. *Ta'dib Jurnal Pendidikan Islam*, 12(1), 159–170. <https://doi.org/10.29313/tjpi.v12i1.11927>
- Härtig, H., Bernholt, S., Fraser, N., Cromley, J. G., & Retelsdorf, J. (2022). Comparing reading comprehension of narrative and expository texts based on the direct and inferential mediation model. *International Journal of Science and Mathematics Education*, 20, 17–41. <https://doi.org/10.1007/s10763-022-10302-5>
- Hawkins, M., Elsworth, G. R., Hoban, E., & Osborne, R. H. (2020). Questionnaire validation practice within a theoretical framework: A systematic descriptive literature review of health literacy assessments. *BMJ Open*, 10(6), 1–12. <https://doi.org/10.1136/bmjopen-2019-035974>
- Idrus, M. (2022). Speaking anxiety experienced by Indonesian students in an international university.
-

- Interference: Journal of Language Literature, and Linguistics*, 3(1), 34–43. <https://doi.org/10.26858/interference.v3i1.31407>
- Isariyawat, C., Yenpdech, C., & Intanoo, K. (2020). The role of literature and literary texts in an EFL context: Cultural awareness and language skills. *Journal of Language and Linguistic Studies*, 16(3), 1320–1333. <https://doi.org/10.17263/jlls.803748>
- James, J. (2022). New assessments and teacher accountability: Lessons for teachers' practice. *American Educational Research Journal*, 59(2), 252–283. <https://doi.org/10.3102/00028312211024596>
- Kadwa, M. S., & Alshenqeeti, H. (2020). Students' perception of alternative assessment: A systematic literature review. *International Journal of Linguistics, Literature and Translation (IJLLT)*, 3(11), 55–67. <https://doi.org/10.32996/ijllt>
- Kamal, S. (2023). Kurikulum merdeka in two senior high schools: Its implementation in EFL classroom. *Global Expert*, 11(2), 25–28. <https://doi.org/10.36982/jge.v11i2.3426>
- Karaman, P. (2021). The Effect of Formative Assessment practices on student learning: A meta-analysis study. *International Journal of Assessment Tools in Education*, 8(4), 801–817. <https://doi.org/10.21449/ijate.870300>
- Kurniati, S. R. (2022). The effect of using inference strategy on reading comprehension at second years students of SMP Negeri 35 Makassar. *Teaching English as a Foreign Language Overseas Journal*, 10(1), 42–49. <https://doi.org/10.47178/teflo.v10i1.1621>
- Kusumaningsih, S. (2021). Incorporating inclusive assessment principles and universal design for learning in assessing multicultural classroom: An autoethnography. *LLT Journal*, 24(1), 337–348. <https://doi.org/10.24071/llt.v24i1.3656>
- Mar, R. A., Li, J., Nguyen, A. T. P., & Ta, C. P. (2021). Memory and comprehension of narrative versus expository texts: A meta-analysis. *Psychonomic Bulletin and Review*, 28(3), 732–749. <https://doi.org/10.3758/s13423-020-01853-1>
- Markey, K., Graham, M. M., Tuohy, D., McCarthy, J., O'Donnell, C., Hennessy, T., Fahy, A., & O'Brien, B. (2023). Navigating learning and teaching in expanding culturally diverse higher education settings. *Higher Education Pedagogies*, 8(1). <https://doi.org/10.1080/23752696.2023.2165527>
- Mattes, R. D., Rowe, S. B., Ohlhorst, S. D., Brown, A. W., Hoffman, D. J., Liska, D. J., Feskens, E. J. M., Dhillon, J., Tucker, K. L., Epstein, L. H., Neufeld, L. M., Kelley, M., Fukagawa, N. K., Sunde, R. A., Zeisel, S. H., Basile, A. J., Borth, L. E., & Jackson, E. (2022). Valuing the diversity of research methods to advance nutrition science. *Advances in Nutrition*, 13(4), 1324–1393. <https://doi.org/10.1093/advances/nmac043>
- Mulyanah, A., & Krisnawati, E. (2022). Cross-cultural communication of foreign students in the Indonesian language regarding cultural expressions. *International Journal of Science and Applied Science: Conference Series P-ISSN*, 6(2), 2549–4635. <https://doi.org/10.20961/ijscasc.v6i2.73914>
- Natsir, M., Saragih, B., Mira A., P. S., & Dewi, R. (2020). Language change in social media (Instagram). *Randwick International of Social Science Journal*, 1(2), 394–401. <https://doi.org/10.47175/rissj.v1i2.70>
- Nurmaya, E., Rusilowati, A., & Astuti, B. (2022). Development of cognitive ability test instrument based on revision Bloom taxonomy on dynamic electricity materials for students of senior high school. *Physics Communication*, 6(2), 50–60. <https://doi.org/10.15294/physcomm.v6i2.37028>
- Parlindungan, F., & Prasetya, W. (2022). *Literature review on trends of comprehension instruction for elementary school students*. 7(1), 42–54. <https://doi.org/10.33086/ehdj.v7i1.2589>
- Prasetya, R. E., & Syarif, A. (2022). Assessment and feedback to enhance online English language learning. *Metathesis: Journal of English Language, Literature, and Teaching*, 6(1), 85–100. <https://doi.org/10.31002/metathesis.v6i1.185>
- Puspita, A. M. I. (2020). Metacognitive analysis of elementary school students. *Indonesian Journal of Elementary Teachers Education*, 1(1), 1–8. <https://doi.org/10.25134/ijete.v1i1.2871>
- Putra, R. A. A., Riwayatningsih, R., & Setyarini, S. (2021). Portraying teacher's metacognitive knowledge to promote efl young learners' critical thinking in indonesia. *International Journal of Language Education*, 5(1), 552–568. <https://doi.org/10.26858/IJOLE.V5I1.13043>
- Ramadhanti, D., & Yanda, D. P. (2021). Students' metacognitive awareness and its impact on writing skill. *International Journal of Language Education*, 5(3), 193–206. <https://doi.org/10.26858/ijole.v5i3.18978>
- Runesi, Y., Moge, T., & Liando, N. (2022). Students' perception in using Common European Framework of Reference (CEFR) application as an English language learning assesment. *International Journal of Education, Information Technology and Others (IJEIT)*, 5(4), 226–234. <https://doi.org/10.5281/zenodo.7076471%0D>
-



- Sadegh, M., Yazdi, G., Shoar, S. M. H., Sobhani, G., Sani, F. V., Khoshkholgh, R., Bazaz, N. M., & Mansourzadeh, A. (2021). Factors affecting students' guesswork in multiple choice questions and corrective strategies. *Med Edu Bull*, 2(6), 341–349. <https://doi.org/10.22034/MEB.2021.312176.1032>
- Sadora, E., Faridi, A., & Suwandi. (2021). Formative assessment based on revised Bloom's Taxonomy in merdeka curriculum English textbook. *The Journal of Educational Development*, 9(2), 139–148. <https://doi.org/10.15294/JED.V9I2.66658>
- Samiei, F., & Ebadi, S. (2021). Exploring EFL learners' inferential reading comprehension skills through a flipped classroom. *Research and Practice in Technology Enhanced Learning*, 16(1), 1–18. <https://doi.org/10.1186/s41039-021-00157-9>
- Septia, N. W., Indrawati, I., Juriana, J., & Rudini, R. (2022). An analysis of students' difficulties in reading comprehension. *EEdJ: English Education Journal*, 2(1), 11–22. <https://doi.org/10.32923/eedj.v2i1.2519>
- Shea, M. (2020). Assessing Reading Fluency. *Running Records*, 83–92. <https://doi.org/10.4324/9780203120408-14>
- Suastra, I. M., & Menggo, S. (2020). Empowering students' writing through performance assessment. *International Journal of Language Education*, 4(3), 432–441. <https://doi.org/10.26858/ijole.v4i3.15060>
- Subandowo, D. (2022). A detailed overview on sociolinguistic competence in pedagogical applications. *International Journal of Humanity Studies (IJHS)*, 5(2), 238–246. <https://doi.org/10.24071/ijhs.v5i2.4411>
- Sumarsono, D., Arrafii, M. A., & Imansyah. (2023). Evaluating the quality of a teacher's made test against five principles of language assessment. *JOLLT Journal of Languages and Language Teaching*, 11(2), 225–237. <https://doi.org/10.33394/jollt.v%25vi%25i.7481>
- Sürücü, L., & Maslakçı, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, 8(3), 2694–2726. <https://doi.org/10.15295/bmij.v8i3.1540>
- Suzanne, Hazenberg, B., Evers-Vermeul, J., & Bergh, H. van den. (2021). A meta-analysis on the effects of text structure instruction on reading comprehension in the upper elementary grades. *International Literacy Association*, 56(3), 435–462.
- Syathroh, I. L., Kareviati, E., Lestari, A., & Fitria, N. (2021). Exploring the potentials of technology integration for teaching language skills: A literature review. *PROJECT (Professional Journal of English Education)*, 4(3), 488–496. <https://doi.org/10.22460/project.v4i3.p488-496>
- Tshering, S. (2022). The impact of using Understanding by Design (UbD) model on class 10 student's achievement in chemistry. *IJCER (International Journal of Chemistry Education Research)*, 6(1), 29–33. <https://doi.org/10.20885/ijcer.vol6.iss1.art4>
- Ulfie, F., & Permana, S. (2021). The alignment between learning outcomes and assessment tasks in writing: A BRT analysis. *STAIRS*, 2(2), 52–62. <https://doi.org/10.21009/stairs.2.2.6>
- Westerveld, M. F., Filiatrault-Veilleux, P., & Paynter, J. (2021). Inferential narrative comprehension ability of young school-age children on the autism spectrum. *Autism and Developmental Language Impairments*, 6, 1–13. <https://doi.org/10.1177/23969415211035666>

## Appendices

### Appendix 1. The Original Version

#### KAPUAS RIVER

Long ago, on Mintin Island, there was a small kingdom led by a very wise king. The lives of the people in the kingdom are guaranteed and prosperous. The people really loved the King and the Empress.

But one day, the Empress was attacked by a strange disease and finally died. The king was very sad. He lost his enthusiasm and intended to sail to erase his sadness.

Knowing this, the royal advisor asked, "Who will run the government while you are away?" "I think my twin sons, Naga and Buaya, are capable of carrying out this responsibility," said the King. After that, the King summoned his two twin sons. "My children, I ask you to run the government while I am away."

Even though Naga and Buaya were twins, their characteristics were not same. Buaya have a kind and generous personality, while Naga was the opposite. Naga liked to go on a spree and never thought about other people's interests.

After the King left, Naga started acting up. He did not want to accompany Buaya in running the government. He was just having fun and force people to pay large taxes. Annoyed by his brother's behaviour, Buaya went to the royal advisor to reprimand Naga. However, Naga did not heed all the advice.

Day by day, Naga behaviour became more and more violent and became more evil. A fight between Naga and Buaya was inevitable. Buaya was determined to stop all the Naga actions. However, the Naga actually fought Buaya. The Battle was inevitable.

In the middle of the journey, the King's heart was restless and ordered the crew to go home. Arriving at the kingdom, he saw many corpses of the royal troops. Before his astonishment had disappeared, he saw his two twin sons fighting. He also rebuked his two sons.

Naga and Buaya turned their heads at the same time. They immediately stopped fighting. "You have wasted Father's trust in you. Look, how many victims have fallen?" said the King.

Suddenly, thunder boomed. Instantly, Buaya changed into a crocodile. Seeing his younger brother turned into a crocodile, Naga was afraid. The king looked at his son with regret. In an instant, Naga turned into a dragon and left his father.

The King was very sorry to see this incident. He felt that his decision to entrust the government to Naga and Buaya was a big mistake. However, he couldn't change what had happened. After that, Buaya ran the government wisely. He repaired all the damage and chaos caused by Naga. He also apologized to the people for his brother's mistakes.

Meanwhile, Naga felt lonely and isolated. He pondered and lamented his fate alone in the Kapuas River. However, as time went by, Naga began to realize his mistakes and felt sorry for wasting his father's trust.

One day, Naga met a monk. He told the monk about his life and his regrets. Hearing Naga's story, the monk gave Naga advice to do good and help others. Naga decided to follow the monk's advice and started helping fishermen who were having difficulty catching fish in the Kapuas River.

Because of his kindness, Naga began to be respected by the local people. They ask the Naga to help them solve problems and give wise advice. Naga began to feel happy because he could help other people and be useful.

Several years later, the elderly King decided to visit Naga on the Kapuas River. After meeting Naga, the King was happy to see the changes in Naga. He felt proud to see Naga, who used to be troubled and like to go on a spree, had now become a wise figure and helped other people.

The King decided to apologize Naga for the bad treatment he had given him. Naga graciously accepted the King's apology.

From there, the Kapuas River was known as a river full of local wisdom. This legend continues to be remembered by the public as a reminder that every action would have an impact in the end.

**Questions:**

1. What was the state of the kingdom under the rule of the King before the Empress died?
    - a. The kingdom was in turmoil.
    - b. The people were impoverished.
    - c. The people were prosperous and loved the King and Empress.
    - d. The King was disliked by the people.
  2. Why did the King decide to sail away?
    - a. To escape from his responsibilities.
    - b. To ease his sadness after the death of the Empress.
    - c. Because he wanted to explore new places.
    - d. He was looking for a remedy for the Empress' disease.
  3. How does the text describe the Naga's character?
    - a. Kind and considerate.
    - b. Reckless and inconsiderate.
    - c. Humble and generous.
    - d. Wise and responsible.
  4. What was Naga's behaviour when the King left?
    - a. He governed wisely.
    - b. He collected large taxes and prioritized his own fun.
    - c. He took care of the people.
    - d. He did nothing and neglected his duties.
  5. Who attempted to reprimand Naga for his bad behaviour?
    - a. The King
    - b. Buaya
    - c. The royal advisor
    - d. The twin sons
  6. What event caused a fight between Naga and Buaya?
    - a. Naga's increasing evil behaviour.
    - b. A misunderstanding between them.
    - c. Naga refused to help Buaya.
    - d. A disagreement over who should rule.
  7. What transformations did Naga and Buaya undergo?
    - a. Naga turned into a dragon, Buaya turned into a crocodile.
    - b. Naga turned into a dragon, Buaya turned into a bird.
    - c. Naga turned into a lion, Buaya turned into a crocodile.
    - d. Naga turned into a bird, Buaya turned into a dragon.
  8. How did Naga change his ways?
    - a. He became a monk.
    - b. He helped fishermen and solved problems for the local people.
    - c. He took over the government from Buaya.
    - d. He began paying large taxes to the King.
  9. What caused the King to feel pride towards Naga?
    - a. Naga's transformation into a dragon.
    - b. Naga's realization of his mistakes and his actions to fix them.
    - c. Naga's defeat in a fight.
-

- d. Naga's return to the kingdom.
10. Why is the Kapuas River remembered by the public?
- It is a reminder of Naga's mistakes and their consequences.
  - It reminds them of the King's wise rule.
  - It is a reminder that every action has an impact in the end.
  - It is associated with the legend of the monk.

## **Appendix 2. The Modified Version**

### **KAPUAS RIVER**

Long ago, on Mintin Island, there was a small kingdom led by a very wise king. The lives of the people in the kingdom are guaranteed and prosperous. The people really loved the King and the Empress.

But one day, the Empress was attacked by a strange disease and finally died. The king was very sad. He lost his enthusiasm and intended to sail to erase his sadness.

Knowing this, the royal advisor asked, "Who will run the government while you are away?" "I think my twin sons, Naga and Buaya, are capable of carrying out this responsibility," said the King. After that, the King summoned his two twin sons. "My children, I ask you to run the government while I am away."

Even though Naga and Buaya were twins, their characteristics were not same. Buaya have a kind and generous personality, while Naga was the opposite. Naga liked to go on a spree and never thought about other people's interests.

After the King left, Naga started acting up. He did not want to accompany Buaya in running the government. He was just having fun and force people to pay large taxes. Annoyed by his brother's behaviour, Buaya went to the royal advisor to reprimand Naga. However, Naga did not heed all the advice.

Day by day, Naga behaviour became more and more violent and became more evil. A fight between Naga and Buaya was inevitable. Buaya was determined to stop all the Naga actions. However, the Naga actually fought Buaya. The Battle was inevitable.

In the middle of the journey, the King's heart was restless and ordered the crew to go home. Arriving at the kingdom, he saw many corpses of the royal troops. Before his astonishment had disappeared, he saw his two twin sons fighting. He also rebuked his two sons.

Naga and Buaya turned their heads at the same time. They immediately stopped fighting. "You have wasted Father's trust in you. Look, how many victims have fallen?" said the King.

Suddenly, thunder boomed. Instantly, Buaya changed into a crocodile. Seeing his younger brother turned into a crocodile, Naga was afraid. The king looked at his son with regret. In an instant, Naga turned into a dragon and left his father.

The King was very sorry to see this incident. He felt that his decision to entrust the government to Naga and Buaya was a big mistake. However, he couldn't change what had happened. After that, Buaya ran the government wisely. He repaired all the damage and chaos caused by Naga. He also apologized to the people for his brother's mistakes.

Meanwhile, Naga felt lonely and isolated. He pondered and lamented his fate alone in the Kapuas River. However, as time went by, Naga began to realize his mistakes and felt sorry for wasting his father's trust.

One day, Naga met a monk. He told the monk about his life and his regrets. Hearing Naga's story, the monk gave Naga advice to do good and help others. Naga decided to follow the monk's advice and started helping fishermen who were having difficulty catching fish in the Kapuas River.

Because of his kindness, Naga began to be respected by the local people. They ask the Naga to help them solve problems and give wise advice. Naga began to feel happy because he could help other people and be useful.

Several years later, the elderly King decided to visit Naga on the Kapuas River. After meeting Naga, the King was happy to see the changes in Naga. He felt proud to see Naga, who used to be troubled and like to go on a spree, had now become a wise figure and helped other people.

The King decided to apologize Naga for the bad treatment he had given him. Naga graciously accepted the King's apology.

From there, the Kapuas River was known as a river full of local wisdom. This legend continues to be remembered by the public as a reminder that every action would have an impact in the end.

**Questions:**

1. Was the kingdom prosperous under the King's rule before the Empress died?
  - a. Yes
  - b. No
  - c. I don't know
  - d. The text doesn't mention it
  - e. It was neither prosperous nor impoverished
2. Why did the King decide to leave the kingdom?
  - a. To expand the kingdom
  - b. To escape his duties
  - c. To grieve the death of the Empress
  - d. To seek a cure for a disease
  - e. The King wanted to participate in a sailing competition
3. How did Naga behave when the King was away?
  - a. He focused on the well-being of his residents
  - b. He prioritized his own enjoyment and levied heavy taxes
  - c. He dedicated his time to study and gain knowledge
  - d. He upheld his father's principles and ruled wisely
  - e. He arranged lavish feasts every day for the kingdom
4. What was the trigger point that led to the fight between Naga and Buaya?
  - a. A disagreement about who should rule
  - b. Naga's refusal to aid Buaya
  - c. A misunderstanding between them
  - d. Naga's increasing irresponsible behaviour
  - e. They fought over the love of a princess
5. How did Naga amend his behaviour?
  - a. He took over the kingdom from Buaya
  - b. He started helping the local fishermen and resolved their issues
  - c. He started paying larger taxes
  - d. He renounced all worldly pleasures and became a monk
  - e. He enrolled in a royal etiquette training program
6. How did the people's perception of Naga change after he started helping them?
  - a. They started despising him
  - b. They acknowledged him as a wise figure and sought his advice
  - c. They feared him more
  - d. They asked him to leave the kingdom
  - e. They considered him a nuisance impossible to get rid of
7. What action did Buaya take to resolve the chaos and damage caused by Naga in the kingdom?
  - a. He blamed Naga publicly and took revenge
  - b. He ran the government wisely and apologized to the people for his brother's mistakes
  - c. He left the kingdom, leaving the mess for someone else to handle

- d. He joined Naga in his misrule
  - e. He decided to throw a massive party to distract people from the chaos
8. How did Naga's and Buaya's attitudes change through the story?
- a. They remained the same throughout
  - b. They became arrogant with power
  - c. They started as humble but became reckless
  - d. They started reckless but learned from their mistakes
  - e. They became scared of their own power
9. What role does the setting play in the story?
- a. The setting has no significance
  - b. It amplifies the character's transformations
  - c. It only serves as a background for the plot
  - d. It represents the state of the kingdom during the King's absence
  - e. It's used to justify the character's decisions
10. How does the story of Naga and Buaya relate to issues of power and responsibility in today's world?
- a. It highlights the abuse of power in leadership positions
  - b. It doesn't relate to the current world scenario
  - c. It emphasizes the importance of personal transformation
  - d. It reflects on the consequences of reckless decision making
  - e. All of the above