

Innovating Portfolio Assessment for Listening Skills in Arabic Textbooks under the Merdeka Curriculum

Rahmad Khusein¹, Lalu Rahmat Sugiara^{2*}, Namiyah Fitriani³, Yogi Sopian Haris⁴, Ilham Dhya'ul Wahid⁵

¹ Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia; rahmadkhusein20@gmail.com

² Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia; rahmatlalu32@gmail.com

³ Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia; namiyahfitriani@gmail.com

⁴ Universitas Ahmad Dahlan, Indonesia; 2308052039@webmail.uad.ac.id

⁵ Al-Azhar University, Egypt; ilhamdhya14@gmail.com

ABSTRACT

This study aims to analyze the assessment forms used in Arabic Language Textbooks for 11th Grade Madrasah Aliyah under the Merdeka Curriculum and to propose an innovative and interactive portfolio assessment model for listening skills. Using a qualitative literature review, primary data were obtained from textbooks and secondary data from relevant academic sources. This finding reveals that current assessments primarily focus on objective tests such as matching questions, true-false statements, and short answers, which are less thorough in evaluating interpretive and reflective listening abilities. To address these limitations, this study introduces a portfolio-based assessment model that emphasizes four criteria: identifying the main idea, recognizing details, analyzing conclusions, and reflection. This model uses a rubric to assess word choice, content, and grammatical accuracy on a scale of 1-4. Innovation lies in integrating process-based reflection and self-assessment components, making evaluation more authentic and learner-centered. This model supports the goals of the Merdeka Curriculum to foster higher-order thinking and learner autonomy in Arabic listening comprehension.

Keywords: Assessment Innovation, Listening Skills, Portfolio

How to Cite Khusein, R., Sugiara, L. R., Fitriani, N., Haris, Y. S., & Wahid, I. D. (2026). Innovating Portfolio Assessment for Listening Skills in Arabic Textbooks under the Merdeka Curriculum. *Jurnal Al Burhan*, 6(1), 425-441. <https://doi.org/10.58988/jab.v6i1.651>

INTRODUCTION

Assessment in learning plays a very important and indispensable role, as it makes a significant contribution to the success of the teaching and learning process (Pellegrino, 2014) (Andayani & Madani, 2023). Assessing learning outcomes is important for understanding the extent to which students have achieved the established goals (Martin & Mahat, 2017) (Destiana et al., 2020). One of the main roles of educators in this context is to provide meaningful feedback that helps students reflect on the effectiveness of their learning process (Mamoon-Al-Bashir et al., 2016) (Rahmawati et al., 2022). Therefore, the assessment process not only serves to quantitatively measure learning outcomes, but also facilitates continuous

reflection, self-evaluation, and competency development (Widiatsih et al., 2020). Effective assessment is also needed to identify students' learning strengths and challenges and provide constructive feedback that supports improvement (Atikah & Amelia, 2024).

However, there is still a gap between the ideal concept of assessment and its actual practice in the field (Lee, 2006). In principle, assessment within the framework of the Merdeka Curriculum emphasizes authenticity, process orientation, and the holistic development of student competencies across cognitive, affective, and psychomotor domains (Khoerunnisa et al., 2025). However, in reality, many assessments used by educators and those found in textbooks are still limited to outcome-oriented tests, such as multiple choice, true-false, or matching, which only focus on the final product and not on the learning process and reflection (Lemu et al., 2024). This gap indicates a mismatch between the philosophical goals of the curriculum, such as independent learning and higher-order thinking skills, and the operational form of the assessment instruments used in the classroom.

Portfolio assessment offers an alternative that can bridge this gap. This assessment can be defined as a continuous evaluation process based on a dataset that documents the improvement and development of students' competencies over a specific period (Munib, 2017). Portfolio assessment can function well both formatively and summatively (Marzuki, 2023). More than just a collection of values, a portfolio serves as an instrument for mapping students' growth, trends, and learning reflections (Nisrofah, 2024). Through this assessment, educators can comprehensively evaluate learning, covering cognitive, affective, and psychomotor aspects. (Magdalena et al., 2023). Assessment practices like this not only evaluate learning outcomes but also provide teachers with pedagogical information to refine their teaching strategies for greater effectiveness (Magdalena et al., 2021).

Assessment can also be conducted through various learning media, including textbooks (Lau et al., 2018) (Hasibuan & Afrinaldi, 2023). However, not all textbooks provide assessments that are appropriate for authentic, reflective, and process-oriented learning needs. This inconsistency is evident in the Arabic Language Textbook for Grade XI of Madrasah Aliyah (Islamic Senior High School) under the Merdeka Curriculum, published by Tiga Serangkai (2024). The listening skills (*maharah istima'*) assessment section is still dominated by matching, true/false, and short answer questions. Structurally, this type of question does not fully meet the principles of good assessment. For example, in matching exercises, the number of questions and answers is always the same, which reduces item variation and test validity.

This condition indicates a clear discrepancy between the ideal assessment goals of the Merdeka Curriculum and the listening assessment format used in the Arabic language textbook. Ideally, listening skills (*maharah istima'*) assessments should be able to measure students' ability to interpret meaning, infer implied messages, and reflect on the communication context (Tamphu et al., 2024). However, the current instruments only capture understanding at a superficial

level. Therefore, assessment innovation is needed that aligns with the vision of holistic and authentic learning as mandated by the Merdeka Curriculum.

Based on these conditions, this research proposes the development of an interactive and adaptive portfolio assessment model for listening skills (*maharah istima'*) integrated into the Arabic Language Textbook for Class XI of Madrasah Aliyah (Trilogy Series, 2024). This model emphasizes process-based evaluation thru the stages of identification, analysis, reflection, and discussion, drawing from textbook materials. This assessment design not only aims to comprehensively measure learning outcomes but also encourages students to engage in self-reflection, critical thinking, and collaborative learning.

Some previous studies have also highlighted the weaknesses of receptive skills assessment in Arabic language learning. Among them is the research conducted by Imawan (2023) showing that receptive skills assessment in Arabic textbooks generally consists of sound imitation tests and essays for listening, and pronunciation tests, multiple choice, true/false, and essays for reading. Similar findings were made by Utami et al., (2024) which revealed that diagnostic assessments of listening skills are still dominated by objective test formats such as multiple choice, true-false, and sentence completion. These assessment formats only measure literal comprehension abilities and do not yet touch upon students' interpretive, analytical, and reflective abilities in understanding spoken texts. Faizmailatus Sofa et al. (2023) affirming that receptive skills such as reading and listening should be assessed in the realm of skills-based assessment, not merely in the realm of knowledge. However, the reality is that the instruments used still focus on the accuracy of answers, not on the context of meaning and real communication.

These results suggest that the present Arabic listening skills assessment's primary weaknesses are its product orientation, its shortcomings with regard to lower-order thinking skills, and its lack of reflecting and participatory elements. Exercises such matching, true/false, and fill in the blank merely measure the capacity to identify information; they do not evaluate, interpret, or analyze the message that was received. Additionally, as stressed in the Merdeka Curriculum, children are not habituated to self-assessment or gaining learning independence due to the lack of formative feedback methods and reflective activities.

Based on this gap, this research focuses on developing an interactive adaptive portfolio assessment model for listening skills integrated into the Arabic Language Textbook for Class XI of Madrasah Aliyah (Trilogy Series, 2024). This model was built by combining the principles of authentic assessment and a process-based approach thru the stages of identification, analysis, reflection, and discussion, sourced directly from the material in the textbook. This approach not only assesses students' listening comprehension but also stimulates their active engagement in the learning process thru reflective and collaborative activities.

The novelty of this research lies in the development of an adaptive portfolio model, which is an assessment model that adjusts to the individual learning progress of students. This makes assessment not just a tool for measuring learning outcomes, but also a means of comprehensively understanding the dynamics of each student's learning process. Additionally, this research also presents the

integration of Pancasila Student Profile values such as critical thinking, independence, cooperation, and reflection into the portfolio assessment system. Thus, assessment not only serves as an academic evaluation but also as a means of character development and strengthening students' social-emotional competencies. Then, the novelty of this research is also evident in the contextualization of portfolio assessment with the transformative spirit of the Merdeka Curriculum. The developed model promotes interactive, authentic, and real-world learning experience-oriented assessment, thus supporting the principles of independent learning and student-centered instruction.

With these characteristics, this research has a high level of novelty because it offers a portfolio assessment model that is not only adaptive to the needs and development of students, but also relevant to the direction of national education policy based on the Pancasila Student Profile and the values of the Merdeka Curriculum transformation. This model is expected to make a significant contribution to strengthening more humanistic, reflective, and contextual Arabic language assessment practices in relation to the realities of learning in madrasas.

METHOD

This research method uses a descriptive qualitative approach aimed at providing a detailed description of assessment phenomena in Arabic Language textbooks for Class XI Merdeka Curriculum and designing interactive portfolio assessment innovations for listening skills. The data sources consist of two types: primary data in the form of the content of the Arabic Language textbook for Class XI Merdeka Curriculum published by Tiga Serangkai (2024), and secondary data obtained from relevant literature such as books, journal articles, and curriculum policy documents. The primary data collection technique is the researcher's documentation of all assessment items related to listening skills (instructions, test items, existing assessment criteria, and related activities) as well as relevant learning materials.

To ensure procedure transparency, data analysis was carried out in a progressive and methodical manner. In order to identify units of analysis in the form of assessment instrument formats and listening task types, the researcher first used document analysis techniques to arrange and minimize the document transcript data. Second, deductive-inductive combination coding was carried out. The first framework (deductive) made reference to the general assessment categories of performance tasks, objective tests, and subjective tests. The coding results were then mapped into a findings matrix that displayed frequency, cognitive depth, and alignment with listening skill objectives (*maharah istima'*).

As for the development of the portfolio model, it follows an explicit instructional design procedure. The development steps include: (1) formulating portfolio assessment objectives aligned with the core competencies of the Merdeka Curriculum; (2) designing portfolio activities (identifying main ideas, exploring specific details, drawing conclusions/analyzing meaning); (3) developing detailed assessment rubrics for three aspects (word choice, content, and grammar) with a scale of 1-4; and (4) creating a guide for teachers and students on how to use

portfolios. Each step is documented with objectives, task instructions, and assessment instruments to facilitate review.

Model validation was conducted thru content validity triangulation. For content validity, the portfolio design and rubric were reviewed by at least two relevant experts (Arabic language education specialists, educational assessment experts, and Arabic language teacher practitioners). They were asked to assess the suitability, relevance of competencies, and clarity of assessment items; expert feedback was integrated into the final revisions. Analysis of the trial results includes simple descriptive statistics (e.g., score distributions, range) to assess scale suitability, as well as qualitative analysis of teacher and student feedback to improve instruction and rubrics. To maintain the credibility and robustness of the findings (trust), the researcher used source triangulation strategies (textbooks, policy documents, literature), technique triangulation (documentation, expert review), and audit trails (recording analytical decisions and instrument revisions). Here is the research flow visualized thru an image.

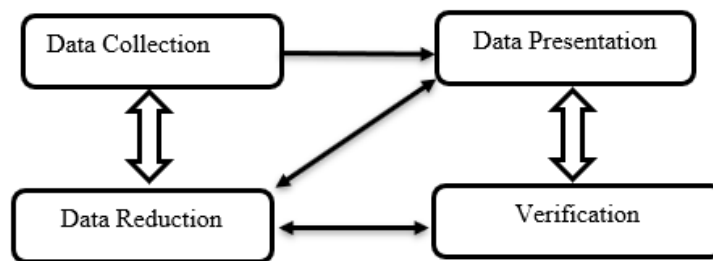


Figure 1. Qualitative Data Analysis Design

By using a combination of document analysis, expert validation, and source and technique triangulation, this research ensures the credibility and validity of the findings. Thru this structured process, this research not only provides an in-depth overview of assessment practices in Arabic language textbooks but also produces a validated and interactive portfolio assessment model that teachers can use to improve students' listening skills (*maharah al-istima'*) in accordance with the principles of the Merdeka Curriculum.

RESULTS AND DISCUSSION

Portfolio Assessment

Portfolio assessments can be categorized as formative assessments, so educators are more interested because these assessments provide different alternatives in the learning evaluation process (Wibowo, 2020). Portfolio assessment is an evaluation method that assesses work or documents to reflect individual achievements, as well as an interactive learning model that uses real problems as a starting point. This method involves students actively in problem-solving and knowledge development through hands-on experience (Kurnila et al., 2022). A portfolio assessment is very supportive of determining the achievement of student competencies and is an evaluation material for teachers in developing the learning process as a solution to the problems faced by students so that they are given motivation and support to achieve learning goals (Sumardi, 2017). In this way, portfolio assessments can serve as a substitute for traditional assessments that

tend to assess only cognitive aspects but also include psychomotor and affective aspects, which are the main targets of education (Najmudin & Ain, 2024).

Portfolio-based learning puts students at the center (student-oriented). In the learning process, students need to be motivated to be willing and able to carry out actions to expand their work experience by strengthening socialization with the environment. Through this socialization, can provide knowledge related to the surrounding environment, increase confidence, and form the personality of students who understand the surrounding diversity so that they can develop individuals who are tolerant of the diversity and differences of the social environment (Magdalena et al., 2021). Thus, a portfolio can provide an overview of the development of students' abilities and potential through their works, including written works, collections of poems or short stories, paintings, handicrafts, lesson plans, books, and others.

Several rules need to be considered and used as guidelines in implementing portfolio assessments in schools, including:

Table 1. Portfolio Assessment Implementation Rules

Criteria	Explanation
Mutual Trust	Teachers and students build a relationship of mutual trust, so that assessment is seen as a means of improving learning, not just getting grades.
Openness	The assessment process is carried out transparently, teachers provide constructive criticism, and students learn from feedback with a sense of security.
Confidentiality	Portfolio documents are kept confidential, comments are submitted privately to maintain ethics and student feelings.
Joint Ownership	Portfolios are seen as shared property between teachers and students, thus fostering a sense of responsibility in students for their work.
Satisfaction	Portfolios provide a sense of satisfaction for teachers and students because they reflect learning success according to the curriculum.
Kerelevance	Portfolio results must be relevant to the learning objectives set out in the curriculum.
Assessment of Process and Results	Assessment does not only focus on the final results, but also includes the learning process, behavior, habits, and experiences of students.

Innovation of Interactive Portfolio Assessment of Arabic Language Listening Skills Class XI Aliyah Published in Triad

In the Arabic Language Student Book for Class XI Madrasah Aliyah Approaches to the Differentiation of the Independent Curriculum published in a trilogy, the assessment of listening skills (*mahārah al-istima'*) is carried out through two types of assessment, namely formative assessment and diagnostic assessment. The exercises provided before and after the listening text in each chapter are designed to support this evaluation process. The Arabic textbook for grade XI Madrasah Aliyah of the differentiation approach in this independent curriculum applies two assessment methods, namely test and non-test. In the test technique, various instruments are provided, namely matchmaking, true-false questions,

short fills or answers, descriptions, and multiple choices. In this book, the assessment technique for listening skills (*mahārah al-istima'*) uses objective tests. There are three forms of objective tests used: *matching test*, *true-false*, and *objective description* (Nurhikmah, 2024).

The Arabic textbook for grade XI Madrasah Aliyah produced by Tiga Serangkai emphasizes the value of innovation in portfolio assessment as an addition to the several ways of assessment. As a result, the researcher discusses portfolio evaluation improvements in the Arabic textbook produced by the Triad of Class XI, which are found in each chapter's *maharah istima* sub-chapter. (Bok et al., 2018). Portfolio assessment, through performance-related information collected, supports continuous feedback and monitoring of competency development over time, improving formative learning and summative decision-making within the framework of competency-based education. In this study, the questions are arranged with the following criteria: Identification of Main Information, Identifying Specific Details, Concluding and Analyzing Meaning, and Discussion and Reflection. The portfolio assessment for listening skills in the grade XI Arabic textbook is designed to measure students' ability to understand a variety of spoken texts, ranging from general information to deeper meaning analysis. Each of the above criteria supports students in developing their listening skills gradually while providing room for continuous reflection and feedback.

The following are the innovations of portfolio assessment instruments in *maharah istima'*, namely:

Table 2. Portfolio Assessment Information Identification

Subject : Arabic

Class : XI

Assignment : I

Material : التسوق(Shopping)

Student name:

No	Competence	Criterion			Total Sector
		Word Choice	Content of the article	Gramatika	
1	Identify Information. What are the main topics discussed in this text?				

Table 3. Portfolio Valuation Identifies Specific Details

Subject : Arabic

Class : XI

Assignment : II

Materi : التسوق(Shopping)

Student name:

No	Competence	Criterion			Total Sector
		Word Choice	Content of the article	Gramatika	
2	Identify specific details. What types are sold at the AIDI Hargono Market?				

Table 4. Portfolio Assessment Concludes and Analyzes Meaning

Subject

: Arabic

Class

: XI

Assignment

: II

Materi

: التسوق(Shopping)

Student name:

No	Competence	Criterion			Total Sector
		Word Choice	Content of the article	Gramatika	
3	Concluding and Analyzing Meaning. What is the significance of the Aidi Hargono Market for the city of Solo?				

Table 5. Portfolio Assessment, Discussion, and Reflection

Subject

: Arabic

Class

: XI

Assignment

: II

Materi

: التسوق(Shopping)

Student name:

No	Competence	Criterion			Total Sector
		Word Choice	Content of the article	Gramatika	
4	Discussion and Reflection. What do you think makes a place like an ordinary market a cultural icon of a city?				

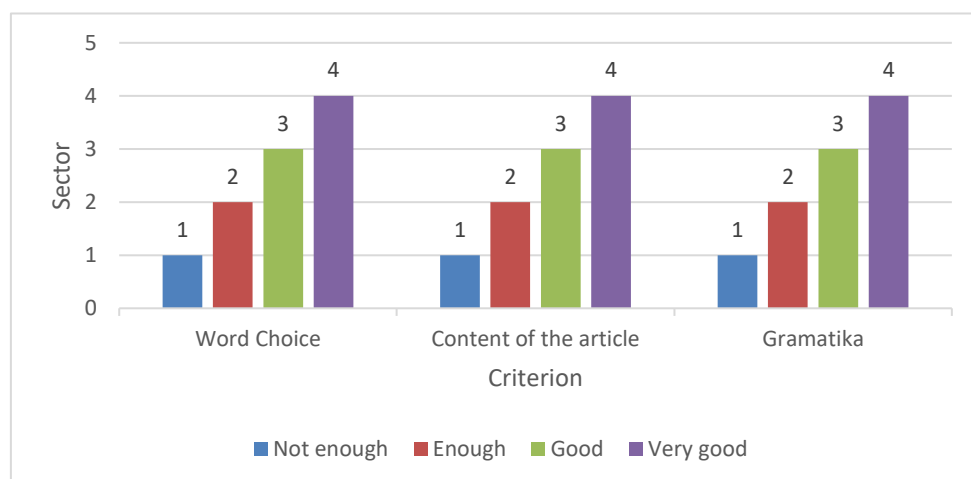


Figure 2. Portfolio Assessment Rubric

Assessment Instructions:

1. Give each criterion a score of 1–4 based on the description above.
2. Add up the scores from all three criteria to get the final score.
3. Maximum score: 12, minimum score: 3.

Calculation Steps

1. Add Up the Total Score
 Add scores from all three criteria (Word Choice, Writing Content, Grammar).
 Example: If a participant scores 4 on Word Choice, 3 on Written Content, and 4 on Grammar, then:
 $4+3+4=11$

2. Calculate the Percentage of Grades

Here's how to calculate the perfect score (on a scale of 100):

$$\text{Final score} = \left(\frac{\text{Total Score}}{\text{Maximum Score}} \right) \times 100$$

The maximum score is 12 (because of 3 criteria x maximum score of 4).

Example:

$$\text{Final score} = \left(\frac{11}{12} \right) \times 100 = 91.67$$

3. Rounding (Optional)

If the value needs to be rounded, then use the usual rounding rule (for example, 91.67 is rounded to 92).

The four main criteria that are the focus of the assessment in this study reflect a holistic approach to evaluating the teaching and learning process. *First*, Key Information Identification is a crucial first step. In this stage, students are tested on their ability to capture and understand the key points of a material or information. This ability helps them build a solid foundation of understanding before moving on to a more in-depth level of analysis. According to Robert Mills Gagne the theory of information processing learning, this theory considers learning as the process of information processing. As an illustration, cognitive psychology in this theory focuses on understanding the learning process through learning outcomes. However, what is more important is to understand information systems that play an important role in improving the learning process (Suryana et al., 2022).

Second, identifying Specific Details encourages students not only to understand the outline but also to be able to dive into information with a high level of accuracy. This criterion requires foresight in paying attention to nuances and important elements that may be missed at first glance (Nurul Fathimah Zahro, 2022). Identifying specific details is an important step in helping students understand information in depth. By focusing on the key aspects of the material studied, students can more easily relate the broader concepts to the detailed information that supports them. This strategy not only strengthens memory but also trains students' analytical skills to evaluate the relationships between different elements in a topic.

Third, analyzing and concluding Meaning is a stage where students are challenged to carry out a high-level thinking process. They do not just receive information but can process, interpret, and produce new understandings of the material studied. This stage develops critical thinking and analytical skills. According to Anderson and Krathwohl's theory, analyzing means breaking matter into its components and identifying the relationships between the parts. Based on the results of the research, the subject can be said to have reached the analysis stage, because it can break down the information into several parts (Syahri & Ahyana, 2021). Thus, an interactive portfolio is not only an evaluation tool, but also a means of developing students' critical and analytical thinking skills.

Finally, discussion and reflection provide a space for students to express their thoughts, share perspectives, and introspect on the learning process that has been passed. This criterion assesses not only intellectual ability but also social skills and metacognition ability (Dyah Ayu Dwi Astuti et al., 2023). Shows that the implementation of group discussions has a positive impact on students' ability to express their opinions, as well as encouraging the development of aspects such as cooperation, confidence, and communication skills. In addition, the discussion method has also been proven to significantly improve learning outcomes.

Simulation of the Application of the Interactive Portfolio Assessment Model *Mahārah al-Istimā'*

As a form of strengthening the innovation of the portfolio assessment model Interactive, the author simulates the application of this instrument to the learning of *mahārah al-istimā'* (listening skills) in class XI Madrasah Aliyah. This simulation illustrates how the assessment process is conducted systematically thru four main stages: identifying key information, identifying specific details, analysis and drawing meaning, and discussion and reflection.

The *first*, the teacher played a three-minute recording of an Arabic conversation on the theme of *al-bi'ah wa al-hifzh 'alayhā* (the environment and its preservation). The recording was taken from authentic sources relevant to students' daily lives, such as a radio interview about the importance of keeping the school clean. After listening twice, students were asked to complete the first stage of their portfolio assignment, which was to identify the main information. At this stage, students write down the core ideas, themes, and main messages of the conversation using their own sentences. The teacher then assessed the students' ability to grasp the main idea based on the criteria in the rubric: accuracy of ideas, completeness of information, and clarity of presentation.

The *second* stage is the identification of specific details, where students are asked to find more detailed supporting information, such as the names of characters, places, times, and factual data that appear in the listening text. This process demands high precision and focus. The teacher provided direct feedback by marking less accurate sections and offering improvement comments thru the learning journal in the students' portfolios.

The *third* stage focuses on analysis and drawing meaning. In this section, students are invited to interpret the implied messages and moral values contained in the conversation. For example, students can write an analysis that the topic reflects the value of *ihsan* in Islam toward the environment or the principle of social responsibility. The teacher assesses based on the students' demonstrated critical thinking skills and depth of interpretation, in accordance with the established rubric.

The *final* stage, discussion and reflection, is conducted thru group activities. Students discuss to compare their findings, discuss different meanings, and collectively draw conclusions about the lessons that can be learned from the text. After that, each student wrote a personal reflective note in their portfolio, containing their feelings, difficulties, and plans for improvement in the next listening session. This stage activates metacognitive abilities, as stated by Dyah Ayu Dwi Astuti et al. (2023), that reflection and group discussion can improve

critical thinking, communication, and cooperation among students.

The process of evaluating an interactive portfolio is continuing. Worksheets, teacher notes, and evaluation rubrics are gathered in a digital folder together with each student's assignment outcomes. Teachers help students understand their strengths and flaws by giving them descriptive comments in addition to scores. A portfolio conference, a quick meeting to assess the students' skill progress over time, is held between the teacher and students at the conclusion of the lesson unit. Through this approach, assessment functions as a tool for reflection and dialogic learning in addition to being an evaluation tool. Compared to traditional assessment methods that rely solely on objective tests like matching or true-false questions, this interactive portfolio model offers significant added value. This assessment emphasizes the process, accommodates differences in students' learning styles, and strengthens the relationship between teachers and students thru meaningful interaction and feedback. Additionally, this model aligns with the principles of competency-based learning and the independent curriculum, which place students as active subjects in the learning process.

Thus, this application simulation demonstrates that interactive portfolio assessment in *Mahārah al-istimā'* learning not only evaluates the final results but also develops students' analytical, reflective, and collaborative abilities. This model is expected to be an effective evaluative innovation in improving the quality of Arabic language learning in Madrasah Aliyah, while also strengthening the integration between cognitive, affective, and psychomotor aspects in a holistic learning process.

The four criteria identified in this study form a complex and interrelated learning system, rather than a simple linear sequence. The relationship between identifying key information, identifying specific details, analyzing and concluding meaning, and discussing and reflecting shows a hierarchical-cyclical pattern, in which each stage builds a foundation for the next stage while creating feedback that reinforces the abilities in the previous stage. This pattern reflects the integration of three fundamental dimensions of learning: the cognitive dimension, which operates in the first three criteria by processing information from basic to complex levels; the social dimension, which manifests itself in group discussions as the construction of knowledge through interaction; and the metacognitive dimension, which emerges in the reflection process as awareness of individual learning processes.

In a theoretical context, these findings present an interesting convergence between Gardner's multiple intelligence theory and Vygotsky's social constructivism principle, suggesting that effective learning requires the simultaneous activation of various types of intelligence through meaningful social interaction (Abraham & Adeniji, 2023) (Handayani & Darodjat, 2024). The Zone of Proximal Development proposed by Vygotsky does not only occur in the discussion phase, but actually operates in every transition between criteria, where students move from individual abilities to higher abilities with the help of peers and teacher facilitation. Furthermore, this research challenges the traditional hierarchy of Bloom's Taxonomy in the context of digital era learning, where the ability to identify key information is becoming increasingly critical due to the

phenomenon of information overload, while metacognitive reflection is actually the foundation for continuous learning, not just a supplement at the end of the process.

The practical implications of these findings are extensive, ranging from curriculum design that needs to integrate all four criteria into each learning unit with continuous formative assessment, to the development of teaching strategies that apply gradual scaffolding from the first to the fourth criteria with gradually decreasing teacher support. Differentiated instruction is crucial to accommodate variations in student abilities in each criterion, while the use of learning technology needs to be specifically designed to reinforce the achievement of each criterion. In a broader context, this research provides a framework for the development of flexible yet structured learning pathways, allowing students to progress at their own pace without losing learning coherence.

Various factors that influence the results of this study can be categorized into three interacting levels. At the internal level of students, cognitive factors such as working memory capacity determine the ability to process complex information in the second and third criteria, while prior knowledge influences the speed of identifying key information, and cognitive flexibility determines the ability to transition between criteria. Motivational factors also play a significant role, where intrinsic motivation encourages engagement in discussion and reflection, self-efficacy influences the courage to share perspectives, and a growth mindset determines the response to feedback in the reflection process (Wibowo, 2022). At the external level, instructional factors include the quality of learning materials that influence the ease of identifying information, task design that determines the level of engagement in analysis, and the quality of discussion facilitation that influences the depth of reflection (Basri, 2014). Environmental factors such as a supportive classroom climate, the quality of peer interactions, and adequate technological infrastructure also determine the successful implementation of the four criteria.

Cultural factors influence students' preferences for independent versus collaborative learning, local communication patterns determine the quality of discussions, and perceptions of authority influence students' courage in expressing their opinions (Ibnu Imam Al Ayyubi; Nurhikmah; Eko Prayetno, 2026). Systemic elements such as the relevant evaluation system, allocation of learning time, and availability of learning resources all have a substantial impact on the optimal achievement of each requirement. (Mursida, 2025). Interestingly, there are mediator and moderator variables that operate within this system, such as cognitive load, which mediates the relationship between material complexity and analytical ability; motivation, which mediates the relationship between learning design and student engagement; and social cohesion, which mediates the relationship between discussion structure and reflection quality.

The scientific contribution of this research lies in the development of an integrative learning model that proposes the Hierarchy-Cycle Learning Framework, challenging the linear view of learning progress by showing that reflection and discussion are not only the culmination of learning, but also catalysts for the next learning cycle. This research also proposes a Digital Age Learning Taxonomy that adapts traditional learning taxonomies to the characteristics of the

digital age, where information curation and metacognitive reflection skills are as important as higher-order analysis skills. These findings open up opportunities for further research that can explore longitudinal aspects to examine the development of each criterion in the long term, cross-cultural validation to test the generalization of findings, and neuroscientific approaches using neuroimaging to understand brain activation in each criterion.

The implications for education policy include the development of teacher training programs that focus on facilitating four-criteria learning, curriculum reform that integrates the four criteria into national standards, and assessment innovations that develop holistic assessment instruments to comprehensively measure the achievement of the four criteria. Thus, this research not only contributes new theoretical understanding of the learning process, but also provides a practical roadmap for more effective and relevant educational transformation in the 21st century, where the ability to process information, collaborate, and reflect are core competencies that students must master.

CONCLUSION

This study concludes that the Arabic language textbook for 11th grade of the Merdeka Curriculum for Madrasah Aliyah, published by Tiga Serangkai in 2024, still has limitations in the assessment instruments for *mahārah al-istimā'* (listening skills), as it still relies on objective test formats such as matching tests, true-false questions, and short objective essays. This assessment model tends to only measure surface-level understanding and has not yet been able to explore students' critical or reflective thinking abilities. To address this, this study develops an innovative portfolio assessment model that includes four main criteria: (1) identifying key information, (2) identifying specific details, (3) drawing conclusions and analyzing meaning, and (4) discussion and reflection. Each criterion is accompanied by a comprehensive scoring rubric that assesses aspects of word choice, content, and grammar on a scale of 1–4, aligning with the principles of the Merdeka Curriculum, which emphasizes authentic, in-depth, and project-based assessment. This innovation offers an alternative assessment model prototype in Arabic language learning that is not only results-oriented but also focuses on the learning process, fostering analytical, reflective, and communicative thinking skills. Thus, this research contributes conceptually to the development of authentic assessment instruments that encourage active student engagement and align with the spirit of differentiated learning in the Merdeka Curriculum.

Nonetheless, this study acknowledges several obstacles and limits. The examination is presently limited to listening and does not yet include other language abilities such as speaking, reading, and writing. Furthermore, the created portfolio evaluation model has not been empirically evaluated in a real-world learning setting, hence its effectiveness and dependability have not been directly verified. The practical implementation of this concept may confront challenges such as time constraints for teachers, differing levels of students' digital literacy in keeping digital portfolios, and the necessity for assessors to use rubrics consistently. To address these challenges, further research is needed thru field trials (pilot testing), teacher training, and empirical validation involving various

madrasah contexts. This effort is important to refine the instrument, improve the consistency of assessment, and measure its impact on student learning outcomes. Additionally, developing similar assessment models for other Arabic language skills will expand their usefulness and flexibility in supporting the more comprehensive implementation of the Merdeka Curriculum.

REFERENCES

- Abraham, M. O., & Adeniji, A. E. (2023). Integrating Gardner's Theory of Multiple Intelligences into Sustainable Development Education: Strategies and Challenges. *Education For Sustainable Development Introduction*, 1–23.
- Andayani, T., & Madani, F. (2023). Peran Penilaian Pembelajaran Dalam Meningkatkan Prestasi Siswa di Pendidikan Dasar. *Jurnal Educatio FKIP UNMA*, 9(2), 924–930. <https://doi.org/10.31949/educatio.v9i2.4402>
- Atikah, & Amelia. (2024). Strategi Penilaian dan Evaluasi Efektif Untuk Meningkatkan Kemampuan Siswa Serta Umpan Balik Dalam Pembelajaran Bahasa Indonesia. *Jurnal Pendidikan Dan Ilmu Bahasa*, 2(3), 76–84.
- Basri, H. (2014). Peningkatan Kualitas Pembelajaran Melalui Adopsi Dan Implementasi Teknologi Instruksional. *Jurnal Mudarrisuna*, 4(2), 228–248.
- Bok, H. G. J., de Jong, L. H., O'Neill, T., Maxey, C., & Hecker, K. G. (2018). Validity evidence for programmatic assessment in competency-based education. *Perspectives on Medical Education*, 7(6), 362–372. <https://doi.org/10.1007/s40037-018-0481-2>
- Destiana, D., Suchyadi, Y., & Anjaswuri, F. (2020). Pengembangan instrumen penilaian untuk meningkatkan kualitas pembelajaran produktif di sekolah dasar. *Jurnal Pendidikan Dan Pengajaran Guru Sekolah Dasar (JPPGuseda)*, 3(2), 119–123.
- Dyah Ayu Dwi Astuti, Sukanto, & Iin Purnamasari. (2023). Analisis Metode Diskusi Kelompok Terhadap Keterampilan Mengemukakan Pendapat Peserta Didik Sekolah Dasar. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(2), 4640–4651. <https://doi.org/10.36989/didaktik.v9i2.1150>
- Faizmailatus Sofa, R Umi Baroroh, Nurinadia, P., Wahidah, N., & Nurul Afifah, S. (2023). Innovation of Language Receptive Skills Based on Authentic Assessment. *Izdiyar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 6(1), 31–44. <https://doi.org/10.22219/jiz.v6i1.24029>
- Handayani, W., & Darodjat, D. (2024). Membangun Kecerdasan Sosial Melalui Pembelajaran Bahasa: Suatu Tinjauan Filsafat. *Journal of Knowledge and Collaboration*, 1(9), 377–388.
- Hasibuan, B., & Afrinaldi. (2023). Pengembangan E-Modul PAI Menggunakan Aplikasi Pageflip Professional Di SD Surya Kids Bukittinggi. *Jurnal Pendidikan Dan Konseling*, 5(2), 4416–4426.
- Ibnu Imam Al Ayyubi; Nurhikmah; Eko Prayetno, F. N. S. R. (2026). Maths In The Qur'an: Should The MI Curriculum Adopt An Islamisation Of Science Approach? *International Journal of Interdisciplinary Research*, 2(1), 1–12. <https://journal.as-salafiyah.id/index.php/ijir/index>
- Imawan, Y. (2023). Inovasi Bentuk Penilaian Tes Kompetensi Reseptif Berbahasa Arab dalam Buku Al-'Arabiyah Lin Nasyi'in Jilid 1. *Tsaqofiya: Jurnal*

- Pendidikan Bahasa Dan Sastra Arab*, 5(1), 19–36.
<https://doi.org/10.21154/tsaqofiya.v5i1.177>
- Khoerunnisa, S., Atiqoh, S., Lathifah, S., & Azis, A. (2025). Competency and Character Based Integrative Assessment Model in Islamic Religious Education (PAI) Learning. *Classroom: Journal of Islamic Education*, 2(1), 129–141.
- Kurnila, V. S., Badus, M., Jeramat, E., & Ningsi, G. P. (2022). Peningkatan Kemampuan Literasi Matematika Melalui Pendekatan Problem Based Learning Bermuatan Penilaian Portofolio. *Euler: Jurnal Ilmiah Matematika, Sains Dan Teknologi*, 10(1), 88–97. <https://doi.org/10.34312/euler.v10i1.13963>
- Lau, K. H., Lam, T., Kam, B. H., Nkhoma, M., Richardson, J., & Thomas, S. (2018). The role of textbook learning resources in e-learning: A taxonomic study. *Computers and Education*, 118, 10–24. <https://doi.org/10.1016/j.compedu.2017.11.005>
- Lee, N. (2006). Bridging the gap between theory and practice in integrated assessment. *Environmental Impact Assessment Review*, 26(1), 57–78.
- Lemu, G., Melese, W., & Debu, D. (2024). *An Analysis of Assessment Techniques and Learning Outcomes in Grade 10 History Curricular Materials*.
- Magdalena, I., Aciakatura, C., Putri, W. A., & Azzizah, F. N. (2023). Efektivitas Pembelajaran Menggunakan Penilaian Portofolio. *Tsaqofah*, 3(5), 802–815. <https://doi.org/10.58578/tsaqofah.v3i5.1383>
- Magdalena, I., Ramdhan, S. N., & Luftiyah, V. (2021). Analisis Penilaian Berbasis Fortofolio pada Pembelajaran Tematik Kelas 3 di Sekolah Dasar Negeri Batujaya Kecamatan Batu Ceper. *ARZUSIN*, 1(1). <https://doi.org/10.58578/arzusin.v1i1.111>
- Mamoon-Al-Bashir, M., Rezaul Kabir, M., & Rahman, I. (2016). The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalizing Teaching in Higher Education. *Journal of Education and Practice*, 7(16), 38–41.
- Martin, L., & Mahat, M. (2017). The Assessment of Learning Outcomes in Australia. *AERA Open*, 3(1), 125–131. <https://doi.org/10.1177/2332858416688904>
- Marzuki, I. (2023). Urgensi Penilaian Portofolio Dalam Evaluasi Pembelajaran Di Era Society 5.0. *Tadarus Tarbawiy: Jurnal Kajian Islam Dan Pendidikan*, 5(2), 171–179. <https://doi.org/10.31000/jkip.v5i2.10073>
- Munib, A. (2017). Penilaian Pembelajaran Bahasa Arab. In *Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta* (Issue May).
- Mursida, M. (2025). Pengaruh Kualitas Pengajaran, Fasilitas Belajar, Dan Sumber Belajar Terhadap Prestasi Belajar Siswa. *CENDEKIA: Jurnal Ilmu Pengetahuan*, 5(2), 579–591. <https://doi.org/10.51878/cendekia.v5i2.4727>
- Najmudin, D., & Ain, S. Q. (2024). Penilaian Portofolio Sebagai Instrumen Pengukuran Kompetensi Peserta Didik. *Celebes Journal of Elementary Education*, 2(1), 1–23.
- Nisrofah, S. (2024). Konsep Penilaian Portofolio Berbasis Multiple Intelligence Dalam Pembelajaran Pai. *Rabbani: Jurnal Pendidikan Agama Islam*, 5(1), 69–87. <https://doi.org/10.19105/rjpai.v5i1.11861>
- Nurhikmah, N. (2024). Character Education Islam From the Views of Imam Al-Ghazali. *Jurnal Al Burhan*, 4(1), 53–66. <https://doi.org/10.58988/jab.v4i1.300>

- Nurul Fathimah Zahro. (2022). Analisis Kemampuan Pemecahan Masalah Peserta Didik Dalam Menyelesaikan Soal PISA. *Didactical Mathematics*, 4(1), 148–155. <https://doi.org/10.31949/dm.v4i1.2073>
- Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicologia Educativa*, 20(2), 65–77. <https://doi.org/10.1016/j.jpse.2014.11.002>
- Rahmawati, D. N., Nisa, A. F., Astuti, D., Fajariyani, F., & Suliyanti, S. (2022). Pemanfaatan Aplikasi Quizizz sebagai Media Penilaian Pembelajaran Ilmu Pengetahuan Alam. *Dawuh Guru: Jurnal Pendidikan MI/SD*, 2(1), 55–66. <https://doi.org/10.35878/guru.v2i1.335>
- Sumardi, S. (2017). Pengembangan Model Penilaian Portofolio dalam Pembelajaran Matematika Berbasis Lesson Study di SMP. *Jurnal VARIDIKA*, 28(2), 170–175. <https://doi.org/10.23917/varidika.v28i2.3031>
- Suryana, E., Lestari, A., & Harto, K. (2022). Teori Pemrosesan Informasi Dan Implikasi Dalam Pembelajaran. *Jurnal Ilmiah Mandala Education*, 8(3), 1853–1862. <https://doi.org/10.58258/jime.v8i3.3498>
- Syahri, A. A., & Ahyana, N. (2021). Analisis Kemampuan Berpikir Tingkat Tinggi Menurut Teori Anderson Dan Krathwohl. *Jurnal Riset Dan Inovasi Pembelajaran*, 1(1), 41–52. <https://doi.org/10.51574/jrip.v1i1.16>
- Tamphu, S., Suyitno, I., Susanto, G., Budiana, N., Salim, M. R., Nurhikmah, & Purnawati, W. (2024). Building bridges to the future of learning: Exploring artificial intelligence research using R-Studio assisted bibliometrics. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2417623>
- Utami, R. W. T. L., R. Umi Baroroh, & Nur Yusra Binti Yacob. (2024). Innovation of Cognitive Diagnostic Assessment in Mahārah istimā' in Class XII Arabic Textbook. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 7(2), 249–264. <https://doi.org/10.22219/jiz.v7i2.31330>
- Wibowo, A. (2020). The use of online mathematics game for portfolio assessment of elementary school students. *Math Didactic: Jurnal Pendidikan Matematika*, 6(1), 50–64. <https://doi.org/10.33654/math.v6i1.909>
- Wibowo, A. (2022). Strategi Peningkatan Learning Orientation Melalui Intrinsic Motivation, Transformational Leadership dan Growth Mindset. In *Journal of Economic Perspectives* (Vol. 2, Issue 1). Universitas Islam Sultan Agung (Indonesia).



© 2026 by the authors. Submitted for possible open-access publication under the terms and conditions of the Creative Commons Attribution (CC-BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).