



A CONCEPTUAL FRAMEWORK FOR EVALUATING ARABIC LANGUAGE INSTRUCTION IN THE DEVELOPMENT OF LANGUAGE PROFICIENCY SKILLS

Basmala^{1*}, Herdah²

^{1*2} IAIN Pare-Pare

*Email: malasajamo@gmail.com, Herdah@iainpare.ac.id

DOI: <https://doi.org/10.37081/jipdas.v6i3.4777>

Abstract

This study aims to formulate a conceptual framework for evaluating Arabic language learning oriented toward the development of language skills. Evaluation in Arabic language learning should not only measure learners' mastery of grammatical rules, but also assess their ability to use the language communicatively through listening, speaking, reading, and writing. Therefore, a comprehensive framework is needed to align learning objectives, assessment processes, and language competence development. This study employed a qualitative approach using a library research design. Data were collected from scholarly sources, including peer-reviewed journal articles and academic books related to Arabic language learning evaluation. The data were analyzed through content analysis to identify major concepts and recurring patterns in previous studies. The findings reveal that the proposed framework consists of four main components: fundamental concepts of evaluation, characteristics of Arabic language evaluation, assessment practices for the four language skills, and general principles of evaluation in Arabic language learning. This study contributes by integrating these components into a systematic framework. The framework is expected to guide the development of more comprehensive evaluation practices and support the improvement of learners' communicative competence in Arabic.

Keywords: Arabic Language Learning Evaluation, Arabic Language Skills, Conceptual Framework Of Evaluation.

1. INTRODUCTION

Arabic language learning across various educational institutions faces challenges in ensuring that evaluation processes are capable of comprehensively measuring the development of language skills. This issue arises because language learning requires not only the mastery of linguistic knowledge but also the ability to use language communicatively across diverse contexts. Therefore, the evaluation system employed in Arabic language learning must be capable of assessing learners' abilities in a holistic and continuous manner. However, in many educational settings, evaluation practices remain dominated by partial and skill-segmented assessment procedures, creating a mismatch between communicative learning objectives and assessment implementation. This condition indicates a broader conceptual problem in the design of Arabic language learning evaluation.

In practice, the evaluation of Arabic language learning in many educational institutions still tends to focus on written tests that assess grammatical aspects and vocabulary memorization. This condition often results in language skills such as speaking and listening being inadequately measured within the assessment process. Furthermore, advancements in educational technology and the adoption of communicative language teaching approaches necessitate evaluation systems that are more adaptive to the demands of modern language learning. Accordingly, the development of an evaluation framework capable of comprehensively measuring language skills has become an essential requirement in contemporary language education practices.

A number of studies have examined various evaluation approaches in Arabic language learning, particularly in relation to authentic assessment, competency-based evaluation, and the integration of



technology in assessment processes. These studies indicate that effective language evaluation should measure language proficiency contextually and should not be limited to structural aspects of the language. Several studies also emphasize the importance of performance-based assessment, which evaluates learners' ability to use language in real communicative situations. Research conducted by Moh. Ainin demonstrates that innovations in digital- and authentic-based evaluation can enhance the effectiveness of assessment in Arabic language learning. Another study by Fitriani and colleagues reveals that authentic assessment can provide a more comprehensive representation of learners' language proficiency through project-based tasks and portfolios. In addition, research by Mulyani and colleagues confirms that communicative task-based evaluation is capable of assessing language skills more contextually within modern Arabic language learning.

Despite these significant contributions, previous studies largely remain oriented toward specific techniques, instruments, or context-dependent strategies, resulting in a fragmented body of knowledge. Their emphasis has mainly been directed toward improving individual assessment approaches rather than integrating those approaches into a coherent conceptual structure. As a result, existing studies have generated practical innovations in Arabic language evaluation, yet they have not sufficiently produced an integrative theoretical framework that systematically connects evaluation principles, four-language-skill assessment, and communicative competence development. This unresolved limitation constitutes a significant research gap, as the absence of such a framework may hinder the alignment of evaluation practices with broader language learning objectives.

Although numerous studies have explored techniques and strategies for evaluating Arabic language learning, research that integrates multiple evaluation approaches into a unified conceptual framework remains relatively limited. This indicates that most studies still focus on specific methods of evaluation without linking them to a broader analytical structure. In fact, a conceptual framework for evaluation is crucial to ensure that assessment processes are aligned with language learning objectives as well as the development of learners' communicative competence. Therefore, the gap in previous studies is not merely methodological, but conceptual, namely the limited formulation of a systematic framework capable of synthesizing evaluation principles and practices holistically. This gap establishes the urgency of developing a more integrative conceptual framework for Arabic language learning evaluation.

Therefore, this study aims to develop a conceptual framework for evaluating Arabic language learning oriented toward the development of language skills. This framework is constructed through an analysis of the fundamental concepts of evaluation, the characteristics of language evaluation, the implementation of evaluation across the four language skills, and the general principles underlying Arabic language learning evaluation. Through this approach, the study seeks to integrate various concepts of educational evaluation and language learning into a systematic analytical framework. Consequently, this study is expected to provide a more comprehensive understanding of evaluation design that supports the development of language skills.

This study offers novelty by moving beyond partial discussions of evaluation techniques toward the formulation of an integrated conceptual framework that systematically relates evaluation concepts, language-skill assessment, and communicative competence development. Unlike previous studies that address evaluation through isolated approaches, this study synthesizes these dimensions into a unified analytical framework. The contribution of this study therefore lies not only in extending theoretical discussion on Arabic language learning evaluation, but also in providing a conceptual foundation for developing more applicable evaluation models in future research and practice.

Accordingly, the novelty of this study lies in the formulation of a conceptual framework for evaluating Arabic language learning that integrates multiple evaluation approaches to support the comprehensive development of language skills. This integrative framework provides a new perspective on how Arabic language learning evaluation can be designed in a more systematic, contextual, and theoretically grounded manner.



2. METHOD

This study employed a qualitative approach with a library research design to develop a conceptual framework for evaluating Arabic language instruction in the development of language proficiency skills. The study focused on the critical analysis and synthesis of theories, models, and empirical findings related to Arabic language learning evaluation. A library research design was selected because it enables the systematic exploration of conceptual relationships among evaluation principles, language assessment practices, and language skill development within Arabic language education contexts.

The literature review process was conducted systematically to ensure the transparency and rigor of source selection. The researcher searched relevant literature through several academic databases, including Google Scholar, Scopus, ERIC and ScienceDirect. The search process utilized combinations of keywords such as “Arabic language evaluation,” “Arabic language assessment,” “language proficiency skills,” “evaluation framework,” “Arabic language learning,” and “language assessment principles.” Boolean operators such as AND, OR, and NOT were employed to refine the search results and improve the relevance of the identified literature. The literature search was limited to publications published between 2021 and 2026 in order to capture contemporary discussions and recent developments in language evaluation research.

The selection of literature sources followed explicit inclusion and exclusion criteria. The inclusion criteria comprised: (1) scholarly publications discussing language evaluation or assessment in Arabic language education; (2) peer-reviewed journal articles, academic books, and conference proceedings; (3) studies addressing evaluation of one or more language skills (listening, speaking, reading, and writing); and (4) sources published in English, Arabic, or Indonesian. Meanwhile, the exclusion criteria included: (1) publications unrelated to educational evaluation; (2) non-academic opinion articles or unpublished materials without academic review; (3) duplicated sources across databases; and (4) studies lacking substantial discussion on evaluation concepts or language proficiency assessment. These criteria were applied systematically to ensure the relevance and academic credibility of the selected literature.

Data collection was carried out through a multi-stage screening process. Initially, the researcher identified potentially relevant sources based on titles and abstracts. Subsequently, full-text readings were conducted to determine the suitability of each source according to the established inclusion criteria. During this stage, the researcher documented important information from each source, including theoretical perspectives, evaluation models, assessment procedures, and findings related to Arabic language proficiency development. This process resulted in the selection of literature considered representative for constructing the conceptual framework.

Data analysis employed qualitative content analysis techniques to identify patterns, conceptual categories, and theoretical relationships within the selected literature. The analysis began with intensive reading and coding of the collected sources. The identified concepts were then categorized into four major dimensions: (1) fundamental concepts of evaluation, (2) characteristics of language evaluation, (3) implementation of evaluation across language skills, and (4) principles of Arabic language learning evaluation. After categorization, the researcher conducted an analytical synthesis to examine the interrelationship among these four dimensions. Rather than merely describing each component separately, the analysis focused on explaining how evaluation principles influence implementation strategies, how evaluation characteristics shape assessment procedures, and how all components contribute operationally to the development of Arabic language proficiency skills. Through this integrative analysis, the study formulated a unified conceptual framework for Arabic language learning evaluation.

To enhance the trustworthiness and validity of the findings, this study applied several validation procedures commonly used in qualitative research. First, source triangulation was conducted by comparing concepts and findings from different types of academic sources, including journal articles, books, and conference proceedings. Second, peer debriefing was employed through academic discussions with colleagues familiar with language assessment and Arabic language education in



order to evaluate the consistency and coherence of the analytical interpretation. Third, theoretical cross-checking was applied to ensure that the synthesized framework remained aligned with established theories of educational evaluation and language assessment. These procedures were intended to strengthen the credibility, consistency, and analytical rigor of the study findings.

This study also adhered to the principles of academic integrity throughout the research process. All sources were appropriately cited and documented in the bibliography according to academic writing standards. The researcher ensured that all referenced materials were credible, relevant, and directly related to the research topic. Furthermore, the analysis was conducted objectively without manipulation or selective distortion of information from the reviewed literature. By maintaining transparency in literature selection, analytical procedures, and source utilization, this study sought to uphold responsible and rigorous academic research standards.

3. RESULT AND DISCUSSION

A. Fundamental Concepts of Evaluation in Arabic Language Learning

1. the nature of evaluation in language learning

The nature of evaluation in language learning is essentially associated with a systematic process aimed at determining the level of instructional effectiveness and assigning value to the learning outcomes achieved by students. From an educational perspective, evaluation is not merely concerned with assigning scores; rather, it constitutes an analytical process for examining the extent to which learning objectives have been attained. Ralph W. Tyler explains that evaluation is a process used to determine the extent to which educational objectives have been achieved; thus, it functions as a tool for assessing the alignment among objectives, instructional processes, and learning outcomes.

Meanwhile, Daniel L. Stufflebeam conceptualizes evaluation as a process of obtaining and providing useful information for making educational decisions. Accordingly, within the context of Arabic language learning, evaluation does not solely assess students' final performance but also provides essential information for teachers to refine instructional strategies, teaching methods, and the effectiveness of the curriculum employed.

The fundamental distinction among measurement, assessment, and evaluation lies in their scope and analytical function within the educational process. Suharsimi Arikunto explains that measurement is the process of comparing learning outcomes with predetermined quantitative standards, resulting in numerical data as objective evidence. Once such quantitative data are obtained, the subsequent stage is assessment, which involves interpreting or describing the measurement results to determine their meaning.

Djemari Mardapi emphasizes that assessment is the activity of interpreting measurement results so that they can be understood pedagogically. Thus, evaluation operates at a broader level, as it not only processes measurement and assessment data but also utilizes such information to determine the quality of instructional programs and to support more comprehensive educational decision-making.

In Arabic language learning, evaluation plays a strategic role because the learning process requires the mastery of complex competencies, including listening, speaking, reading, and writing skills. Norman E. Gronlund and Kenneth D. Moore explain that evaluation is a systematic process of collecting, analyzing, and interpreting information to determine the extent to which learners achieve instructional objectives. This perspective indicates that evaluation is not solely outcome-oriented but also encompasses the learning process itself.

Properly designed evaluation enables Arabic language teachers to monitor the development of students' linguistic competence, identify learning difficulties, and design more effective pedagogical interventions. Therefore, evaluation functions as a feedback mechanism that connects instructional processes, language competency attainment, and the development of more adaptive teaching strategies.

Based on these perspectives, it can be concluded that the nature of evaluation in Arabic language learning constitutes a comprehensive process involving measurement and assessment activities to determine the overall quality of both instructional processes and learning outcomes.



Measurement generates quantitative data, assessment provides interpretation of such data, and evaluation utilizes the information obtained to determine instructional effectiveness and to make appropriate pedagogical decisions.

Thus, evaluation serves not only as a tool for measuring learning outcomes but also as an instrument for reflection and continuous improvement in enhancing the quality of Arabic language learning.

2. evaluation as an integral component of the instructional system

Evaluation is inherently inseparable from learning objectives within an educational system. Learning objectives function as the intended direction to be achieved, whereas evaluation serves as an instrument to determine the extent to which these objectives have been realized in the learning process. In other words, evaluation does not merely function to assign grades to learners; rather, it serves as a means of verifying the attainment of competencies that have been formulated during the initial stage of instructional planning. Through evaluation aligned with learning objectives, educators obtain empirical evidence regarding the level of instructional effectiveness and are able to assess whether the strategies employed have optimally supported the achievement of these objectives.

In practice, evaluation is also closely associated with both the learning process and learning outcomes. It is not only conducted at the final stage to determine students' overall achievement but also takes place throughout the instructional process. Through formative evaluation, educators are able to monitor students' engagement, identify the difficulties they encounter, and observe the gradual development of their competencies. Learning outcomes observed at the end of instruction essentially reflect the quality of the learning process experienced by students.

The integration of objectives, processes, and outcomes positions evaluation as a crucial foundation for instructional decision-making. Information derived from evaluation provides objective data that enable educators to determine appropriate follow-up actions. Based on evaluation results, teachers may decide whether students require remedial instruction, enrichment programs, or even modifications in instructional approaches and teaching methods.

Thus, evaluation functions not only as an assessment tool but also as a foundation for reflection and continuous improvement within the educational process, thereby enabling instruction to become more adaptive, effective, and responsive to learners' needs.

B. Characteristics of Evaluation in Arabic Language Learning

Evaluation in Arabic language learning should not be merely understood as the process of assigning grades; rather, it constitutes a systematic activity aimed at determining the extent to which learners' language competence and their understanding of language use have developed. Through evaluation, teachers can obtain a clear representation of students' level of proficiency in Arabic, both in terms of linguistic knowledge and their ability to use the language communicatively.

To ensure objectivity and effectiveness in measuring learners' linguistic competence, several fundamental criteria characterize evaluation in Arabic language learning, including:

1. evaluation is comprehensive

Evaluation does not solely focus on cognitive aspects such as vocabulary mastery, sentence structure, or grammatical rules; rather, it also encompasses the affective and psychomotor dimensions of learners. Within the cognitive domain, evaluation measures students' understanding of linguistic principles, including Arabic syntax and morphology, as well as their ability to comprehend meaning across various types of texts.

In the affective domain, assessment may reflect students' attitudes toward Arabic language learning, their interest in reading or studying Arabic texts, and their appreciation of the cultural values embedded within the language. Meanwhile, the psychomotor domain is manifested in practical skills such as articulating *hijaiyah* letters with accurate points of articulation, reading texts fluently, and speaking Arabic with appropriate intonation. With such a comprehensive scope, evaluation is capable of providing a more holistic representation of learners' language development.

This comprehensive orientation indicates that evaluation in Arabic language learning should be directed toward assessing learners as whole language users rather than merely measuring fragmented



linguistic components. The interrelationship among cognitive understanding, affective disposition, and psychomotor performance demonstrates that language competence develops through multiple interconnected dimensions. Therefore, a comprehensive evaluation framework allows assessment to capture not only academic achievement, but also learners' readiness, engagement, and practical ability in using Arabic as a means of communication.

2. evaluation oriented toward language skills

Arabic language learning is fundamentally directed toward the development of communicative competence; therefore, evaluation must measure the four language skills in an integrated manner, namely listening (*istima'*), speaking (*kalam*), reading (*qira'ah*), and writing (*kitabah*). Listening evaluation assesses learners' ability to comprehend information derived from conversations or spoken texts, whereas speaking evaluation measures their ability to express ideas orally with appropriate structure and pronunciation.

In reading skills, assessment is directed toward learners' ability to comprehend textual content, identify main ideas, and interpret implicit meanings. Meanwhile, in writing skills, evaluation examines the ability to construct sentences and paragraphs that are logical, well-structured, and consistent with Arabic linguistic conventions. This skills-oriented approach ensures that evaluation not only measures knowledge about the language but also assesses the ability to use the language in authentic contexts.

This characteristic further implies that Arabic language evaluation should not treat the four skills as isolated components, but as interrelated competencies that support communicative performance as a whole. Receptive skills such as listening and reading provide the foundation for productive skills such as speaking and writing, making integrated assessment essential. Through a language-skills orientation, evaluation is positioned not merely to determine achievement in separate skill areas, but also to represent learners' overall communicative development more systematically.

3. continuous and systematic evaluation

Evaluation is not confined to the final stage of instruction but is implemented from the beginning to the end of the learning process. At the initial stage, evaluation may take the form of a pre-test to determine learners' prior knowledge and baseline competence. Subsequently, during the instructional process, formative evaluation is conducted to monitor learning progress and identify the difficulties encountered by students.

At the final stage, summative evaluation is administered to assess the overall achievement of competencies. The implementation of evaluation must be systematically designed through clearly defined steps, including the formulation of assessment indicators, the development of test blueprints, the construction of test items or tasks, and the analysis of evaluation results. Through a well-structured approach, evaluation is able to generate accurate information regarding the effectiveness of instruction.

This continuity reflects that evaluation functions not solely as a mechanism for measuring outcomes, but also as an ongoing process of monitoring, feedback, and instructional refinement. The systematic relationship among diagnostic, formative, and summative evaluation enables teachers to identify learning needs, improve instructional strategies, and ensure the alignment between learning objectives and assessment practices. Accordingly, continuous and systematic evaluation constitutes an essential characteristic of Arabic language learning assessment that supports both learner development and instructional effectiveness.

4. performance- and competency-based evaluation

Evaluation does not merely focus on what learners know; rather, it emphasizes what they are able to perform using the language they have learned. Accordingly, evaluation is often operationalized through performance-based tasks such as oral presentations, dialogues or conversations, expressive reading of texts, and the composition of simple essays in Arabic. Through these tasks, teachers are able to assess students' level of proficiency in using the language in practical contexts. A competency-based approach also requires clearly defined standards or achievement criteria, enabling evaluation results to indicate the extent to which learners have attained the expected level of language



proficiency.

Unlike approaches that primarily measure conceptual knowledge, performance- and competency-based evaluation emphasizes observable demonstrations of language ability. This characteristic reflects the view that language proficiency is evidenced through actual performance rather than solely through mastery of linguistic concepts. Through clearly defined competencies and authentic performance tasks, evaluation is able to determine not only whether learners possess knowledge of Arabic, but whether they can apply that knowledge meaningfully in communicative practice.

5. authentic evaluation in the context of language use

Evaluation in Arabic language learning also emphasizes the application of authentic assessment, which evaluates language competence within real-life contexts of use. Authentic evaluation is designed so that assessment tasks reflect actual communicative situations, thereby requiring learners to use the language functionally.

In practice, such evaluation may take the form of activities such as comprehending conversations, retelling the content of texts, composing simple messages or letters, and responding to specific communicative situations in Arabic. This approach enables multiple language skills to be employed in an integrated manner within a single assessment activity. Accordingly, evaluation does not merely measure discrete abilities but also assesses learners' capacity to use Arabic holistically across diverse communicative contexts.

While performance-based evaluation emphasizes the demonstration of ability, authentic evaluation further stresses the contextual realism in which that ability is assessed. This means that learners are evaluated not only on whether they can perform a language task, but also on whether they can use Arabic appropriately in situations resembling actual communication. Such contextual orientation strengthens the relevance of evaluation to real language use and reinforces the role of assessment in supporting communicative competence development.

C. The Implementation of Evaluation across the Four Arabic Language Skills

The implementation of evaluation in Arabic language learning must be aligned with the characteristics of the four core language skills listening, speaking, reading, and writing. Each skill involves distinct cognitive and performative processes; therefore, the evaluation techniques employed must be specifically designed to address these differences.

Appropriate evaluation provides a more accurate representation of learners' language proficiency while simultaneously enabling teachers to determine more effective instructional strategies. Consequently, the implementation of evaluation for each Arabic language skill must consider the forms of assessment, the instruments employed, and the criteria aligned with instructional objectives.

1. evaluation of listening skills (*maharah al-istima'*)

The evaluation of listening skills (*maharah al-istima'*) aims to assess learners' ability to comprehend messages delivered orally in Arabic. Forms of evaluation may include listening to dialogues, conversations, or spoken texts presented through audio recordings or direct teacher delivery. Following exposure to such materials, learners may be required to answer questions, complete missing information, identify main ideas, or summarize the content of the discourse. These evaluation techniques are designed to measure students' ability to capture meaning, understand conversational context, and identify key information conveyed orally.

The instruments used in listening evaluation may include multiple-choice items, short-answer questions, true–false items, or tasks requiring learners to note essential information from the text they hear. In addition, evaluation may be conducted through activities such as responding to oral instructions or retelling the content of the material that has been presented. These instruments must be carefully designed by considering the level of linguistic difficulty, clarity of audio delivery, and the relevance of the material to learners' proficiency levels. With appropriately designed instruments, listening evaluation can yield valid information regarding students' ability to comprehend Arabic through auditory input.



Beyond the design of instruments, effective listening evaluation also requires a balanced integration of formative and summative assessment, as both serve complementary functions in supporting the development of learners' communicative competence. Formative evaluation enables teachers to monitor learners' ongoing listening comprehension processes, identify difficulties in interpreting spoken input, and provide feedback that supports the gradual improvement of comprehension strategies. In contrast, summative evaluation serves to determine the extent to which learners have achieved expected listening competencies at the end of instruction.

Equally important, listening assessment should be grounded in contextualized language materials that reflect authentic communicative situations, rather than relying solely on decontextualized or memory-based tasks. In this sense, evaluation is not merely intended to test learners' recall of orally presented information, but to map the cognitive processes involved in interpreting meaning, processing contextual cues, and constructing understanding from auditory input. Through a balanced and context-sensitive approach, listening evaluation becomes more capable of representing authentic language comprehension and supporting the broader development of communicative competence.

2. evaluation of speaking skills (*maharah al-kalam*)

The evaluation of speaking skills (*maharah al-kalam*) is focused on assessing learners' oral performance in using Arabic as a means of communication. Common forms of evaluation include paired dialogues, oral presentations, storytelling, interviews, and simulated conversations in specific contexts. Through these activities, teachers are able to directly observe students' ability to express ideas, respond to questions, and appropriately employ vocabulary and grammatical structures in oral communication.

Assessment of speaking ability typically employs specific criteria that reflect essential aspects of oral communication. Commonly used criteria include fluency, accuracy of pronunciation (including *makhraj* and intonation), grammatical accuracy, adequacy of vocabulary, and the ability to convey ideas clearly and coherently. By applying clearly defined and measurable criteria, the evaluation of speaking skills can be conducted more objectively and can provide meaningful feedback for the development of learners' communicative competence.

Beyond the selection of evaluation tasks, the assessment of speaking skills should also take into account psycholinguistic factors that may influence learners' oral performance, particularly language anxiety and self-confidence. In many cases, students' speaking performance during assessment does not solely reflect their actual communicative ability, but may also be affected by nervousness, fear of making errors, or limited confidence when using Arabic in front of others. This indicates that effective speaking evaluation should not be restricted to eliciting language production, but should also create a supportive assessment environment in which learners feel encouraged to communicate with reduced psychological pressure.

Through supportive interaction, constructive feedback, and appropriately designed speaking tasks, evaluation can facilitate more authentic demonstrations of learners' oral ability. In this sense, the assessment of speaking should be understood not merely as measuring technical accuracy, but as supporting the holistic development of communicative competence by considering both linguistic performance and the psychological conditions that shape language use.

3. evaluation of reading skills (*maharah al-qira'ah*)

The evaluation of reading skills (*maharah al-qira'ah*) aims to assess learners' ability to comprehend written texts in Arabic. Forms of evaluation may include reading passages followed by comprehension questions, identifying the main ideas of paragraphs, locating specific information, and explaining the meaning of vocabulary within contextualized sentences. Through such evaluation, teachers are able to determine the extent to which students understand the content of texts and can relate the information presented.

Beyond functioning as a process of decoding written symbols, the evaluation of reading skills should also reflect the cognitive-interactive nature of reading, in which comprehension is constructed through an active interaction between the reader and the text. In this perspective, learners



are not merely expected to retrieve explicit information, but to engage in meaning-making processes involving inference, interpretation, and the integration of textual information with prior knowledge.

This cognitive interaction indicates that reading evaluation should be designed to capture not only technical comprehension, but also the extent to which learners actively negotiate meaning and respond critically to the text. Such an orientation is particularly relevant in modern Arabic language learning, where the development of Higher-Order Thinking Skills (HOTS) has become increasingly important. Accordingly, evaluation techniques such as identifying main ideas, interpreting contextual meaning, and analyzing textual information do not merely assess comprehension, but also function as pedagogical means to foster critical thinking and support the integrated development of communicative and cognitive competence.

Beyond assessing basic comprehension, reading evaluation may also measure higher-order analytical skills. Learners may be required to identify text structures, interpret implicit meanings, or draw conclusions from the information presented. This type of assessment contributes to the development of critical thinking skills while simultaneously strengthening learners' ability to engage with Arabic texts at a deeper level.

4. evaluation of writing skills (*maharah al-kitabah*)

The evaluation of writing skills (*maharah al-kitabah*) is directed toward assessing learners' ability to express ideas in written Arabic. Forms of assessment may include composing simple sentences, constructing paragraphs, writing descriptive texts, or producing short essays based on specific themes. Through these activities, teachers can evaluate students' ability to organize ideas, employ appropriate vocabulary, and apply grammatical rules accurately in written communication.

Beyond evaluating written products, the assessment of writing skills should also be understood through a process-oriented perspective in which writing is viewed as an iterative activity involving drafting, reviewing, and revising. In this regard, effective writing evaluation should not be limited to judging the final written output, but should also monitor learners' developmental progress throughout the writing process. Through formative assessment practices, teachers are able to provide scaffolding at different stages of writing development, support learners in refining their ideas and language use, and facilitate gradual improvement in written expression.

This perspective also promotes learner autonomy, as students are encouraged to reflect on feedback, revise their own writing, and develop greater responsibility for the quality of their written work. Accordingly, evaluation in *maharah al-kitabah* should capture not only the quality of the final product, but also the processes through which communicative competence in writing is progressively constructed.

In assessing writing skills, the use of scoring rubrics is essential to ensure that the evaluation process is both objective and systematic. Such rubrics typically encompass several aspects, including grammatical accuracy, vocabulary usage, coherence of ideas, clarity of content, and the correct use of spelling and punctuation in Arabic. The use of well-defined rubrics enables teachers to conduct consistent assessments while also providing constructive feedback to enhance learners' writing proficiency.

D. General Principles of Evaluation in Arabic Language Learning

A comparison of various scholarly perspectives indicates that evaluation principles have evolved from a predominantly technical approach toward a more communicative and contextual orientation, particularly in language learning. Within the general framework of educational evaluation, Ralph W. Tyler emphasizes that evaluation must be goal-oriented. According to this view, evaluation functions to examine the alignment between instructional objectives and the behavioral changes demonstrated by learners. This principle underscores that assessment must not be detached from the curriculum objectives that have been previously established.

Meanwhile, Suharsimi Arikunto highlights the importance of evaluation instrument quality through the principles of validity, reliability, and objectivity. These three principles constitute a scientific foundation that ensures assessment results accurately and fairly reflect learners' abilities. Accordingly, classical evaluation approaches emphasize both alignment with instructional objectives



and the precision of the measurement instruments employed in the assessment process.

Within the context of language evaluation, these principles have further developed by emphasizing meaning and communicative function. Mahmud Kamil An-Naqah asserts that language evaluation should assess the meaningfulness of messages conveyed by learners; thus, assessment should not merely focus on the accuracy of linguistic forms but also on the ability to communicate meaning effectively.

This perspective is consistent with the view of Rusydi Ahmad Thu'aimah, who emphasizes the principle of functionality, namely the ability to use language in authentic communicative contexts. From this standpoint, language evaluation must be capable of representing the extent to which learners can use language as a medium of social interaction. Consequently, evaluation is no longer confined to measuring linguistic structures but extends to assessing the successful use of language in meaningful and functional situations.

Recent research developments also highlight an increasing emphasis on ethical dimensions and accountability in educational evaluation. A study conducted by Syahrir Syahrir asserts that modern evaluation systems must incorporate the principle of transparency so that both the processes and outcomes of assessment can be clearly understood and publicly justified by teachers and educational institutions.

Based on the synthesis of these perspectives, this study identifies seven principal evaluation principles: meaningfulness, objectivity, validity, reliability, goal orientation, functionality, and transparency. The principles of meaningfulness and functionality emphasize that language evaluation must assess authentic communicative competence; objectivity, validity, and reliability ensure the accuracy and fairness of assessment outcomes; goal orientation guarantees alignment between evaluation and learning outcomes; and transparency underscores the importance of openness in the assessment process.

The integration of these seven principles is considered capable of establishing an evaluation system that is not only academically rigorous but also responsive to the demands of communicative and accountable language learning. These seven principal elements can be elaborated as follows:

1. the principle of meaningfulness

The principle of meaningfulness asserts that the evaluation of language learning must be capable of assessing learners' ability to convey and comprehend the meaning of messages effectively. Language, by its very nature, functions as a medium of communication; therefore, the success of learning is not solely measured by the accuracy of grammatical structures but also by learners' ability to communicate ideas, concepts, and information clearly to others.

Accordingly, meaningful evaluation is typically designed in the form of tasks that provide learners with opportunities to use language communicatively, such as dialogues, presentations, text composition, or the comprehension of discourse with specific communicative purposes. Through this approach, evaluation does not merely measure knowledge about the language but also assesses the ability to use language to convey meaning in authentic contexts.

2. the principle of objectivity

The principle of objectivity requires that the evaluation process be conducted fairly, impartially, and free from assessor subjectivity. Objective assessment must be based on clearly defined criteria and consistent standards so that evaluation results accurately reflect learners' abilities. To achieve objectivity, evaluation must employ structured instruments, detailed scoring rubrics, and systematic assessment procedures.

Thus, non-academic factors such as personal preferences, emotional proximity, or subjective perceptions should not influence assessment outcomes. Objectivity is also essential for maintaining learners' trust in the assessment system implemented within the instructional process.

3. the principle of validity

Validity refers to the degree to which an evaluation instrument accurately measures the competencies it is intended to assess. An evaluation is considered valid when its content, format, and procedures are aligned with instructional objectives and the competencies to be achieved by learners.



In language learning, validity implies that the form of evaluation must correspond to the skill being assessed. For example, speaking ability should be evaluated through speaking activities, whereas writing ability should be assessed through writing tasks. By ensuring the validity of instruments, evaluation results can provide an accurate representation of learners' abilities and serve as a reliable basis for decision-making in the instructional process.

4. the principle of reliability

Reliability refers to the consistency of evaluation results when applied under similar conditions or when assessed by different evaluators. A reliable evaluation produces relatively consistent results when repeated using the same procedures. In practice, reliability can be achieved through the development of clear instruments, the use of standardized scoring rubrics, and the implementation of systematic evaluation procedures. Reliability is essential because inconsistent assessment results may undermine confidence in the accuracy of evaluation. Therefore, effective evaluation must be designed in such a way as to yield stable and trustworthy results.

The principle of reliability also requires that evaluation results remain dependable across variations in assessment contexts, tasks, and evaluators. In this regard, reliability is closely related to the control of subjectivity, particularly in performance-based assessments such as speaking and writing, where scoring judgments may vary among raters. Therefore, efforts to strengthen reliability may include rater calibration, the application of explicit scoring criteria, and repeated assessment when necessary to verify the stability of results. Through such measures, reliability functions not only as a technical requirement of assessment, but also as a quality assurance principle that supports objective interpretation of learners' achievement and strengthens the credibility of evaluation outcomes.

5. the principle of goal orientation

The principle of goal orientation emphasizes that evaluation must always be aligned with predetermined learning objectives. Learning objectives serve as the primary reference in designing evaluation instruments, determining assessment indicators, and interpreting assessment results. Without alignment between objectives and evaluation, the assessment process may lose direction and fail to provide relevant information regarding learners' achievement.

Therefore, every evaluation activity must be designed based on the competencies to be attained, ensuring that assessment results accurately reflect the extent to which learners have achieved the intended learning objectives.

6. the principle of functionality

The principle of functionality asserts that language evaluation must assess learners' ability to use language practically across various communicative situations. Language is not merely learned as a set of linguistic rules but as a tool for interaction in everyday life. Accordingly, evaluation should be designed in the form of tasks that reflect authentic language use, such as engaging in discussions, composing messages, comprehending informational texts, or delivering presentations. Through this approach, evaluation can represent the extent to which learners are able to use language effectively and appropriately in real communicative contexts.

7. the principle of transparency

Transparency is a principle that emphasizes openness in the evaluation process, including procedures, assessment criteria, and the communication of evaluation results. Learners need to understand how they are assessed, which indicators are applied, and how final scores are determined.³ Such openness enables learners to recognize their strengths and weaknesses, thereby allowing them to improve their learning process in a more targeted manner.

Moreover, transparency enhances accountability and trust in the assessment system, as the entire evaluation process can be clearly understood and justified. With the implementation of transparency, evaluation functions not only as a tool for measuring learning outcomes but also as a reflective mechanism that promotes the continuous improvement of instructional quality.



4. CONCLUSION

This study yielded four principal findings concerning evaluation in Arabic language learning. First, the fundamental concept of evaluation in Arabic language learning indicates that evaluation constitutes a comprehensive process involving both measurement and assessment activities to determine the attainment of learning objectives and to assess the quality of instructional processes and learning outcomes. Second, the characteristics of evaluation in Arabic language learning emphasize that evaluation must be comprehensive, oriented toward language skills, implemented in a continuous and systematic manner, and focused on performance-based assessment as well as the use of language in authentic contexts.

Third, the implementation of evaluation across the four Arabic language skills listening, speaking, reading, and writing demonstrates that each skill requires distinct forms of assessment, instruments, and criteria in accordance with its specific characteristics. Fourth, the general principles of evaluation in Arabic language learning highlight the importance of meaningfulness, objectivity, validity, reliability, goal orientation, functionality, and transparency as foundational elements in designing and implementing evaluation that is both accurate and aligned with instructional objectives.

The strength of this study lies in its ability to construct a conceptual synthesis that bridges theories of educational evaluation with the specific needs of Arabic language learning. It also proposes a structured evaluation framework encompassing fundamental evaluation concepts, the characteristics of language evaluation, their application across the four language skills, and the underlying principles guiding evaluation practices. This systematic organization provides a clearer understanding of the interrelationship among learning objectives, evaluation processes, and the development of learners' language competencies.

Nevertheless, this study has certain limitations, as it primarily focuses on conceptual aspects and therefore does not comprehensively represent the diversity of evaluation practices across different Arabic language learning contexts. In addition, the study does not yet address the integration of technology in the evaluation of Arabic language learning, which has become increasingly significant in modern education. Therefore, future research is recommended to extend this inquiry by developing more contextual and adaptive evaluation models that respond to technological advancements and the evolving demands of Arabic language learning in the digital education era.

5. REFERENCES

- Abdullah, Soran, Mustafa Altun, and Farhad Majeed Hama. "The Effect of Language Learning Anxiety on English Oral Performance." *International Journal of Social Sciences & Educational Studies* 9, no. 3 (2022): 337–47. <https://ijsses.tiu.edu.iq/index.php/ijsses/article/view/200>.
- Ainin, Moh. "Evaluation Models and Digital Assessment Innovation in Arabic Language Learning: Towards Authentic and Competency-Based Evaluation." *Mudalla: Proceeding International Conference On Arabic Language* 1, no. 1 (2023): 35–42. <https://proceeding.arab-um.com/index.php/mudalla/article/view/6>.
- al-Nāqah, Mahmūd Kāmil. *Ta'lim Al-Lughah Al-'Arabiyyah Li Al-Nāṭiqīn Bi Lugāt Ukhrā: Asāsuhu Wa Madākhiluhu Wa Ṭuruq Tadrīsihā*. Makkah: Jāmi'ah Umm al-Qurā, 1985.
- Alfinnas, Hafilah, Moh Ali, and A Syathori. "Instrumen Penilaian Autentik Pada Pembelajaran Bahasa Arab Di Madrasah Aliyah." *Aphorisme: Journal of Arabic Language, Literature, and Education* 7, no. 1 (2026): 193–210. <https://www.researchgate.net/publication/401259295>.
- Arikunto, Suharsimi. *Dasar-Dasar Evaluasi Pendidikan*. 2nd ed. Jakarta: Bumi Aksara, 2018.
- Baihaqi, M., Mahmudah Mahmudah, Ahmad Jundi al Mubarak, Nabilah Rabbaniyah, and Dwi Susanto. "Studi Validasi Dan Reliabilitas Butir Soal Tes Sumatif Dalam Pendidikan Bahasa Arab." *Cerdika: Jurnal Ilmiah Indonesia* 5, no. 4 (2025): 1530–42. <https://garuda.kemdiktisaintek.go.id/documents/detail/5098444>.
- Dewi, Azzaria Mida Mustika, and R. Umi Baroroh. "Maharah Qira'ah Assessment Innovation with Receptive Active Language Competency Test." *Kalamuna: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban* 4, no. 2 (2023): 235–44. <https://doi.org/10.52593/klm.05.2.08>.



- Fauzi, Ahmad, and Siti Rahmawati. "Evaluation of Speaking Competence in Arabic Language Learning: A Performance Assessment Approach." *Arabic Studies: Arabic Teaching Innovation and Arabic Literature* 2, no. 1 (2025): 12–22. <https://ojs.bustanilmu.com/index.php/IJAS/article/view/73>.
- Firdaus, Muhammad, Syahrudin Usman, Kamaluddin Abunawas, Misykat Malik Ibrahim, and Haniah. "Evaluation of the Arabic Language Learning Implementation and Challenges Faced in the Foreign Language Intensification Program (PIBA)." *Arabiyatuna: Jurnal Bahasa Arab* 9, no. 1 (2025): 307–26. <https://doi.org/10.29240/jba.v9i1.12505>.
- Fitriani, and Herdah. "Analisis Praktik Evaluasi Pembelajaran Bahasa Arab Berbasis Pengetahuan, Keterampilan, Dan Sikap Terhadap Guru Di Sekolah." *Al Ibrah: Journal of Arabic Language Education* 8, no. 1 (2025): 34–40. <https://doi.org/10.24256/jale.v8i1.5779>.
- Fitriani, Namiyah, Badrus Zaman, and Muhammad Imron. "Pengembangan Evaluasi Pembelajaran Bahasa Arab Berbasis Penilaian Autentik Di Madrasah Tsanawiyah Miftahul Huda Madura." *Tarunaedu: Journal of Education and Learning* 03, no. 02 (2025): 78–89. <https://journal.staitaruna.ac.id/index.php/jel/article/view/629>.
- Gronlund, Norman Edwin, and Robert Leslie Linn. *Measurement and Assessment in Teaching*. 11th ed. Upper Saddle River, New Jersey (NJ), USA: Pearson Education, 2009.
- Hakim, Tirtu Luqmanul, Abdul Haris, and Mokhammad Miftakhul Huda. "Assessment of Speaking Skills (Maharah Kalam) in Arabic Language Lesson at Madrasah Ibtidaiyah Riyadlus Sholihien Jember." *Journal of Science and Education (JSE)* 5, no. 2 (2025): 599–605. <https://jse.rezkimedia.org/index.php/jse/article/view/485>.
- Listiani, Fittrizah, and Nasrullah. "An Analysis of Formative and Summative Assessment by English Foreign Language (EFL) Teacher." *JELLT (Journal of English Language and Literature Teaching)* 8, no. 2 (2023): 517–26. <https://ejournal.unima.ac.id/index.php/jellt/article/view/7834>.
- Liu, Suli. "A Research on the Blended Evaluation Mode in College English Writing Course." *Journal of Language Teaching and Research* 13, no. 4 (2022): 763–71. <https://doi.org/10.17507/jltr.1304.09>.
- Manafe, Yetursance Y., and Louis F. Boesday. "Measurement Design and Evaluation Analysis of Learning Achievements To Support Accreditation in the Electrical Engineering Education Study Program, Nusa Cendana University." *Journal of Language, Culture and Education Studies* 2, no. 1 (2025): 1–9. <https://doi.org/10.61784/jlces3005>.
- Mancar, Sinem Arslan, and Deniz H. Gulleroglu. "Comparison of Inter-Rater Reliability Techniques in Performance-Based Assessment." *International Journal of Assessment Tools in Education* 9, no. 2 (2022): 515–33. <https://doi.org/10.21449/ijate.993805>.
- Mardapi, Djemari. *Pengukuran, Penilaian, Dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika, 2012.
- Maureen, Jessica Febrianti, and Vera Syamsi. "Evaluating Uptake Towards Written Corrective Feedback To Improve Writing Skills." *IJEE (Indonesian Journal of English Education)* 11, no. 2 (2024): 247–58. <https://doi.org/10.15408/ijee.v11i2.40309>.
- Mulyani, Rinda Eka, Ikhwanul Habib, and Nur Hadi. "Implementing Authentic Assessment to Enhance Language Skill in Arabic Language Learning." *Abjadia: International Journal of Education* 10, no. 2 (2025): 338–49. <https://ejournal.uin-malang.ac.id/index.php/abjadia/article/view/32783>.
- Nurfauzan, Muhammad Fakhri, Maizar Tri al Iqram, Raswan, and Ubaid Ridlo. "Evaluasi Pembelajaran Bahasa Arab Sebagai Bahasa Asing: Konsep, Tantangan, Dan Strategi." *Jejak Digital: Jurnal Ilmiah Multidisiplin* 2, no. 1 (2026): 564–73. <https://indojournal.com/index.php/jejakdigital/article/view/1649>.
- Nurzahira, Fadhila, Irwan M. Jayadi, and Ubaid Ridlo. "Konsep Evaluasi Pembelajaran Bahasa Arab." *Ihya Al-Arabiyah; Jurnal Pendidikan Bahasa Dan Sastra Arab* 7, no. 1 (2023): 33–50. <https://jurnal.uinsu.ac.id/index.php/ihya/article/view/26379>.
- Pahlefi, M. Riza. "Kajian Teoritis Pengembangan Instrumen Penilaian Keterampilan Menyimak



- (Mahārah Al-Istima') Dalam Pembelajaran Bahasa Arab M." *Uktub: Journal of Arabic Studies* 2, no. 2 (2022): 68–84. <https://ojs.uktub-journal.com/index.php/uktub/article/view/6458>.
- Stufflebeam, Daniel Leroy, and Christopher Lee Smith Coryn. *Evaluation Theory, Models, and Applications*. 2nd ed. San Francisco: Jossey-Bass, 2014.
- Syafiqoh, Nada Nabilah, R. Umi Baroroh, Tasha Ayu Az-Zahra, and Moh Paris Fauzi Sawedi. "Innovation in the Assessment of Maharah Kitabah Based on Akm Arabic Books Kemenag." *Lahjah Arabiyah: Jurnal Bahasa Arab Dan Pendidikan Bahasa Arab* 4, no. 1 (2023): 20–33. <https://journal.ibrahimy.ac.id/index.php/Lahjah/article/view/2504>.
- Syahrir, and A. Fandir. "An Educational Evaluation Model to Strengthen Institutional Accountability and Quality Through a Systematic Literature Review Approach." *Panthera: Jurnal Ilmiah Pendidikan Sains Dan Terapan* 5, no. 4 (2025): 1541–65. <https://doi.org/10.36312/panthera.v5i4.760>.
- Ṭu'aymah, Rushdī Aḥmad. *Ta'lim Al-'Arabiyah Li Ghayr Al-Nāṭiqīn Bihā: Manāhijuhu Wa Asālībuhu*. Rabat: ISESCO (Islamic Educational, Scientific and Cultural Organization), 1989.
- Tyler, Ralph Winfred. *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press, 1949.
- Wahana Roza, Risma, Suparmanto Suparmanto, Yazid Yazid, Galang Candra Pramana, and Eka Nurcahaya. "Evaluasi Pembelajaran Bahasa Arab Berbasis Computer Based Test Di Madrasah Aliyah Nurul Iman Ujan Rintis." *Lughati: Jurnal Pendidikan Bahasa Arab* 03, no. 01 (2025): 76–89. <https://ejournal.unkafa.ac.id/index.php/lughati/article/view/1547?utm>.