Al Akhbar

Al Akhbar: Jurnal Pendidikan Bahasa Arab

Formative Assessment for Improvement and Self-Development in Arabic Language Learning

Angga Eka Firmansyah

Magister Pendidikan Bahasa Arab, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

Correspondence email: 250104220008@student.uin-malang.ac.id

Abstract

Assessment in Arabic language learning should serve not only to determine final grades but also to improve instruction and foster learner autonomy. Formative assessment involves teachers and learners gathering information about learning processes and outcomes and using it as feedback to adjust ongoing instruction. This article examines the concept and functions of formative assessment, its application in Arabic language learning, and practical steps for designing classroom assessments. It uses a narrative literature review of scholarly articles, policy documents, and relevant literature on formative assessment, feedback, self-assessment, and Arabic language learning evaluation. The sources were critically examined, thematically grouped, compared, and conceptually synthesized. The review indicates that an assessment is formative not because of the instrument used, but because evidence of learning is used to provide feedback and guide subsequent action. In Arabic language learning, this principle applies to listening (istima'), speaking (kalam), reading (qira'ah), and writing (kitabah), as well as vocabulary (mufradat) and grammar (qawa'id), through observation, performance tasks, reflection, self-assessment, peer assessment, and portfolios. The article proposes a design sequence connecting objectives, criteria, evidence, feedback, and follow-up.

Keywords: formative assessment; Arabic language learning; feedback; self-assessment; learner autonomy

A. Introduction

Learning Arabic as a foreign language requires learners to draw on linguistic knowledge to understand and convey meaning. This competence develops through four interrelated skills: *istima'* (listening), *kalam* (speaking), *qira'ah* (reading), and *kitabah* (writing). Vocabulary (*mufradat*) and grammar (*qawa'id*) support all four skills. Assessment should therefore elicit evidence that matches the intended skill. A listening task can ask learners to identify information in a spoken message, while a speaking task can show how they respond to an interlocutor. Reading and writing tasks can reveal how learners interpret a text and organise ideas in writing. Tasks can also give learners opportunities to apply their grammatical knowledge, as illustrated by the use of recitation in virtual *nahwu* instruction (Akmalia & Faizin, 2021).

Learners may progress at different rates across these skills. Some can understand short texts but hesitate in conversation, while others can express ideas orally and still need support in writing coherent sentences. A low score on a speaking task can reflect limited vocabulary, difficulty with pronunciation, or misunderstanding the instructions. Bamualim (2020) discusses how evaluation can establish whether teaching objectives have been achieved and guide improvements in Arabic instruction. Ridho (2018) emphasises selecting evaluation techniques that match the objectives and abilities being assessed. At the level of instructional development, Faizin (2020) describes how evaluation can identify strengths and weaknesses in an instructional system or product and inform revision. In classroom assessment, identifying the nature of a learner's difficulty helps the teacher decide what support to provide next.

Formative assessment connects evidence of learning with decisions about what teachers and learners do next. Black and Wiliam (1998, 2009) explain that evidence gathered during instruction serves a formative purpose when teachers and learners use it to adjust teaching and learning. A quiz, oral question, observation, discussion, or written task can serve this purpose when its results inform a subsequent action. For example, a teacher may use

learners' responses to organise groups for targeted practice, adjust the pace of a lesson, or provide extension tasks. Sadler (1989) adds that learners need to understand the expected standard, compare their work with it, and identify what they can do to improve.

Learner interest is one aspect of the classroom experience that has been examined in Arabic lessons at an Islamic junior secondary school (Munawaroh & Faizin, 2025). In formative assessment, learners participate by reviewing their work, interpreting feedback, and deciding what to improve. Nicol and Macfarlane-Dick (2006) explain how feedback can support self-regulation by helping learners understand objectives, monitor progress, and adjust their strategies. Self-assessment gives learners a role in judging their work against relevant criteria (Andrade, 2019). Carless and Boud (2018) describe feedback literacy as the capacity to make sense of feedback, make judgments, manage emotional responses, and take action. In an Arabic speaking lesson, learners might use a short set of criteria to review a dialogue, identify where their meaning became unclear, and decide what to practise before trying again.

The value of feedback depends on its content, focus, and delivery (Hattie & Timperley, 2007). In a speaking task, a teacher might point out that a learner conveyed the intended meaning clearly but used a verb form that did not agree with the subject pronoun. The teacher can then ask the learner to revise the utterance and repeat the exchange. In a writing task, feedback might focus on how ideas are connected across sentences, followed by an opportunity to revise the paragraph. These steps link the learning objective, the learner's current performance, and a specific action for improvement. They also require teachers to interpret evidence accurately and align feedback with the learning objective and the learner's stage of development (Schildkamp et al., 2020; Wolsey & Karkouti, 2024).

Studies of Arabic language teaching illustrate how assessment tasks can be matched to specific objectives. Wanti et al. (2022) discuss non-paper-based approaches to formative assessment. Zaharo et al. (2024) illustrate how word

puzzles can be used to monitor vocabulary learning, while Hamidah et al. (2025) describe interactive tasks for assessing vocabulary and grammar. For speaking, a short dialogue provides a sample of how learners express meaning and respond to others. For writing, a draft shows how learners organise ideas and express them in sentences. A second attempt at the dialogue or a revised draft allows teachers and learners to see how the performance changes after feedback.

This article explains how formative assessment can improve instruction and support learner self-regulation in Arabic language learning. It describes how teachers can design assessment tasks that align with learning objectives and presents sample designs for speaking and writing. The examples address four decisions teachers make during instruction: which competencies to observe, what evidence to collect, how to interpret that evidence, and what action teachers and learners can take in response.

B. Method

This article employs library research with a descriptive qualitative approach. This method was selected because the discussion focuses on examining concepts, research findings, and formative assessment practices in order to formulate improvement strategies and assessment-design procedures for Arabic language learning. The data sources comprise national scholarly journal articles relevant to formative assessment, language-learning evaluation, feedback, and Arabic language instruction. The literature was selected purposively by considering topical relevance, the clarity of publication details, and the contribution of each source to the focus of the review. Library research enables researchers to examine written sources systematically, connect relevant ideas, and develop an understanding of the issue under study (Sari, 2021). The review focused on sources addressing the conceptual foundations of formative assessment, Arabic language learning evaluation, feedback, self-regulation, and the assessment of language skills. Priority was given to sources providing definitions, models, principles, or

examples of implementation that could be related to the article's focus. Classical literature was retained when it served as a foundational reference for formative assessment, while more recent publications were used to explain developments in practice and the context of language learning.

Data were collected through documentation by reading, recording, and categorizing important information from the selected sources. The information gathered covered the definitions and purposes of formative assessment, forms of evidence of learning, principles of feedback, learner involvement, and application to the four Arabic language skills. The data were then analyzed descriptively and qualitatively by selecting information relevant to the article's focus, grouping it into themes, comparing ideas across sources, and developing a synthesis to address the review objectives. The analysis considered the context and meaning of the information rather than merely compiling quotations from various publications (Fadli, 2021). As this is a library-based study, the sample instruments and follow-up actions presented are conceptual designs based on the literature, not instruments that have been tested through field research. To maintain coherence between the sources and the discussion, each recorded item of information was assigned to a thematic group and then reviewed for its relevance to the review questions. The synthesis did not involve aggregating research results; instead, it identified similarities in ideas, differences in emphasis, and practical implications that could be drawn from the sources. Accordingly, the findings are presented as an argumentative discussion and sample designs explaining the relationship between objectives, evidence of learning, feedback, and follow-up action.

C. Results and Discussion

C.1 Formative Assessment as a Strategy for Improvement and Self-Development

In general, assessment is a systematic process of collecting, analyzing, and interpreting information about learners' achievement as a basis for

pedagogical decision-making. In learning, such information can be obtained through tests, observation, performance, assignments, and reflection. Assessment is an integrated part of the learning process, not merely an activity for measuring final outcomes (Kementerian Pendidikan Riset, dan Teknologi Republik Indonesia, 2022). More specifically, formative assessment refers to the activities through which teachers and learners obtain information about learning processes and outcomes and then use that information as feedback to adjust ongoing teaching and learning (Black & Wiliam, 1998). This definition emphasizes that the formative function is not determined by the form of the instrument but by the purpose for which its results are used. Therefore, written tests, oral questions, observation, discussion, reflection journals, and self-assessment can all form part of formative assessment when their results are interpreted and used to determine subsequent instructional steps. Conversely, an activity that produces only a score without feedback or follow-up does not fully perform a formative function.

Formative assessment is the process of collecting and using information during instruction to identify learners' progress and determine steps for improvement. (Black & Wiliam, 1998) emphasize that an assessment activity serves a formative function when the evidence of learning obtained is used as feedback to adjust teaching and learning. In practice, such evidence may come from oral responses, exercise results, observations of activities, or learners' work products. Teachers then compare the evidence with the objectives and success criteria to identify what has been mastered and what still needs reinforcement. Thus, quizzes, oral questions, observation, and written assignments do not automatically constitute formative assessment; their formative function emerges when the results are interpreted and acted upon.

In the learning context, formative assessment can be understood as assessment conducted during the learning process to monitor progress, identify learners' needs and obstacles, and provide teachers and learners with

information for determining improvements. It enables teachers to adjust explanations, select reinforcement activities, or provide support according to classroom needs. Learners can use the information to recognize abilities they have developed, understand their mistakes, and set subsequent learning targets. Formative assessment therefore goes beyond assigning grades: it positions assessment results as feedback for continuous improvement in learning (Darwin et al., 2023) (Wanti et al., 2022)

Its implementation takes place in an interconnected cycle, beginning with the establishment of objectives and success criteria, followed by the collection of evidence of learning, identification of difficulties, provision of feedback, and implementation of follow-up action. Objectives and criteria need to be communicated in language that learners understand so that they know what abilities are expected. After evidence has been collected, teachers should do more than count correct and incorrect answers; they need to identify error patterns and their possible causes. This analysis provides a basis for deciding whether learners need further explanation, additional practice, guidance, or more challenging learning tasks. Unlike summative assessment, which summarizes achievement at the end of instruction, formative assessment is used while there is still an opportunity to improve the learning process and outcomes (Black & Wiliam, 2009).

Feedback becomes the core of an improvement strategy when it explains the intended objective, the learner's current level of achievement, and the next steps (Hattie & Timperley, 2007). Comments should address the task or process, be specific, and provide guidance that learners can act on. For example, in a kalam exercise, a teacher may point out that a learner has communicated the main idea but still needs to improve verb use or the pronunciation of certain vocabulary items. The learner is then given an opportunity to repeat the dialogue while taking the feedback into account. In addition to teacher feedback, reflection, self-assessment, and peer assessment can help learners identify their strengths, difficulties, and learning targets. The active use of feedback supports self-regulation and the development of

independent learning skills (Nicol & Macfarlane-Dick, 2006)(Andrade, 2019)(Carless & Boud, 2018).

The fundamental difference between formative and summative assessment lies in the function of the information produced, not merely in the timing of the assessment or the type of instrument used. A short test administered midway through instruction may serve a summative function if it is used only to assign grades and rank achievement. Conversely, a task conducted near the end of a unit may still serve a formative function if its results are used to identify gaps and provide opportunities for improvement before learning proceeds. Therefore, teachers should begin planning by asking what decisions will be made on the basis of the evidence, rather than merely deciding which type of test to administer.

The purposes of formative assessment include at least three interrelated objectives. First, it provides an overview of learners' developing understanding and abilities. Second, it supplies information that enables teachers to adjust instructional strategies, materials, and support. Third, it helps learners identify their current learning position and determine the steps they need to take. These three objectives show that formative assessment takes place through interaction. Teachers collect and interpret evidence, learners demonstrate their understanding through learning activities, and both use the information to determine subsequent action.

In classroom practice, evidence of learning does not always have to be obtained through formal instruments. Prompting questions, learners' responses during discussions, observation notes, short exercises, group work, recordings of speaking practice, and writing drafts can all serve as sources of information. Teachers can select the evidence most appropriate to the learning objectives and the time available. To avoid excessively subjective assessment, such evidence should be linked to clear indicators. A simple note such as "can identify the main idea but cannot yet summarize the text" is more informative than a general comment such as "does not understand the reading."

Establishing success criteria is also an important part of the formative cycle. Criteria provide a concrete description of the expected quality of work. In a speaking task, for example, the criteria may include message comprehensibility, vocabulary appropriateness, structural accuracy, pronunciation, and fluency. These criteria should be communicated before learners complete the task so that they know which aspects to consider. An excessive number of criteria can make assessment difficult; therefore, teachers should select a few priority indicators in accordance with the objective of each activity.

After collecting evidence, teachers need to interpret the information in context. Errors in verb forms, for example, may occur because learners do not yet understand changes according to the dhamir, have difficulty recalling vocabulary, or focus so heavily on fluency that they do not have time to check grammatical structures. Each cause requires a different follow-up action. Conjugation exercises may address morphological difficulties, whereas staged dialogue practice may help learners who know the rules but cannot yet use them spontaneously. Careful interpretation prevents teachers from assigning the same exercises to all learners without considering their individual needs.

Effective feedback needs to address learners' informational needs. Comments may explain the task objective, the aspects of the work that meet expectations, the aspects that do not yet meet the criteria, and concrete steps for improvement. Feedback consisting only of general praise or a mark indicating an error does not always help learners understand what they need to do. Conversely, comments that are too lengthy and contain too many corrections at once may make it difficult for learners to set priorities. Teachers can therefore select one or two of the most important aspects for improvement on each occasion and then provide an opportunity for practice or revision.

Learners' self-development occurs when they begin to use assessment information without always waiting for teachers' directions. Self-assessment

can begin with simple questions: which aspects have been mastered, which errors still occur, and what will be improved in the next task? Peer assessment can also be conducted using concise criteria. These activities are not intended to transfer responsibility for assessment to learners; rather, they train learners to observe the quality of work, provide comments respectfully, and accept suggestions as material for improvement.

However, self-assessment and peer assessment require gradual habituation. Learners need to understand the criteria and practise giving evidence-based comments rather than comments based on personal closeness or preference. Teachers can demonstrate how to provide specific feedback, for example by identifying a sentence component that is already accurate and another that needs improvement. Teachers also need to review peer-assessment results, particularly for language aspects that require accuracy. With such guidance, reflection can become a safe and constructive learning activity.

Overall, formative assessment can be understood as a recurring cycle rather than a one-time procedure. The results of follow-up action become new evidence for determining whether the improvement strategy has helped learners. If difficulties persist, teachers need to review their explanations, types of practice, or criteria. If learners have achieved the objectives, teachers can provide more complex tasks. This pattern makes instruction more responsive to classroom development and connects assessment with concrete pedagogical action.

C.2 Designing and Implementing Formative Assessment in Arabic Language Learning

The design of formative assessment begins with formulating learning objectives and observable indicators. Teachers need to ensure that the indicators represent the abilities they genuinely intend to develop, rather than merely aspects that are easiest to score. They then select tasks or activities that can elicit evidence of learning, establish success criteria, determine how results will be recorded, and plan follow-up action.

Objectives, activities, and instruments must be aligned. If the learning objective requires oral communication, for example, evidence should be obtained through dialogue or speaking performance, not solely through a written test on grammatical rules. Success criteria should be communicated before the task so that learners understand the aspects being assessed and can direct their efforts accordingly.

Assessment techniques need to be adapted to the characteristics of the four maharah because each skill produces different forms of evidence. In istima', teachers can assess the ability to identify the main idea, recognize detailed information, and understand the meaning of expressions in spoken language. In kalam, evidence can be obtained through dialogue, interviews, or presentations, using criteria such as message comprehensibility, vocabulary accuracy, pronunciation, and fluency. In qira'ah, assessment can focus on understanding the main idea and detailed information, as well as interpreting vocabulary meaning from context. In kitabah, teachers can examine the drafting process, structural accuracy, vocabulary choice, coherence of ideas, and revisions. Assessment therefore examines not only the final product but also the process learners go through.

Mufradat and qawa'id should be assessed through their use, not merely through memorization or recognition of forms. Vocabulary quizzes and structural exercises can provide an initial picture of learners' mastery, but their results need to be complemented by tasks requiring learners to use vocabulary and grammatical rules in sentences, dialogues, or paragraphs. Teachers can assess whether vocabulary is used appropriately in context and whether the selected structures support clarity of meaning. Rubrics should include key indicators aligned with the objectives and proficiency level, such as content accuracy, message comprehensibility, vocabulary use, and language structure. Language-proficiency descriptors may refer to the CEFR Companion Volume framework, adapted to the characteristics of Arabic language learning (Europe, 2020).

Assessment results are then analyzed to determine follow-up actions appropriate to the type of difficulty. Learners who cannot identify detailed information in istima' activities may be given practice in recognizing keywords or listening to spoken texts in stages. Speaking difficulties can be addressed by strengthening vocabulary, practising pronunciation, and using sentence patterns in simple dialogues. In qira'ah, teachers can help learners develop strategies for identifying main ideas and understanding vocabulary from context, while kitabah difficulties can be followed up through comments on drafts and opportunities to revise writing. Self-assessment and peer assessment can be used with simple criteria, while teachers retain responsibility for checking the accuracy of the assessment. Evidence from various activities should be considered proportionately so that instructional decisions do not depend on a single test result (Ridho, 2018)(Hamidah et al., 2025)(Zaharo et al., 2024).

At the planning stage, teachers need to ensure that learning objectives are formulated as observable abilities. The objective "understanding a text" remains too broad unless the evidence of understanding is specified. It can be made more precise by stating that learners should be able to identify the main idea, locate detailed information, or summarize a simple text. More measurable formulations help teachers select appropriate tasks and make learning targets easier for learners to understand. Indicators do not need to be numerous; a few indicators that genuinely represent the competency are more useful than a long list that is difficult to observe.

The selection of assessment techniques should consider alignment among competencies, context, and learner characteristics. Tests can be used to examine understanding of particular vocabulary items or structures, while observation and performance tasks are more appropriate for examining language use in communication. Portfolios can be used to trace the development of writing or project work over time. Questions and discussions can provide direct insight into learners' understanding. A proportionate combination of techniques helps teachers obtain a more complete picture, but

this does not mean that every activity must be assessed using numerous instruments.

In maharah istima', formative assessment can be conducted through listening to short texts, dialogues, or oral instructions. Teachers can ask learners to identify the topic, select relevant information, note keywords, or answer questions based on what they hear. The evidence of learning should reveal which aspects have not yet been understood; for example, a learner may identify the topic but fail to capture information about time and place. Follow-up action may include replaying a particular section, practising keyword recognition, or providing a transcript after the listening activity. Teachers should provide support gradually so that learners continue practising independent listening comprehension.

In maharah kalam, formative assessment is more meaningful when learners have opportunities to use language in situations with a communicative purpose. Pair dialogues, simple interviews, role-plays, and short presentations can be used according to proficiency level. Teachers can observe message comprehensibility, fluency, vocabulary accuracy, pronunciation, and structural use. Not all aspects need to be the focus of a single performance. If the activity aims to build confidence in conveying information, teachers can prioritize message comprehensibility and fluency, while structural accuracy can become the focus of a subsequent practice session.

In maharah qira'ah, evidence of learning can be obtained through reading texts and completing comprehension tasks. Questions may address the main idea, detailed information, relationships between sentences, contextual vocabulary meaning, or a summary of the text. Teachers need to distinguish learners who have difficulty recognizing vocabulary from those who understand the vocabulary but cannot yet connect information across the text. The first group may need vocabulary reinforcement and strategies for inferring meaning, whereas the second may need practice in identifying relationships between ideas and writing summaries.

In maharah kitabah, formative assessment needs to consider the writing process rather than only the final product. Teachers can collect outlines, initial drafts, feedback notes, and revised versions. This enables teachers to identify whether learners' difficulties lie in developing ideas, selecting vocabulary, constructing sentences, or organizing paragraphs. Feedback should prioritize the aspects that most affect the comprehensibility of the writing. Correcting all errors at once may cause learners to focus on correction marks without understanding the pattern of improvement.

Mufradat and qawa'id need to be integrated into language use. Vocabulary mastery cannot be measured solely by the ability to translate word lists; it should also be assessed through the ability to select and use words appropriately in context. Similarly, mastery of qawa'id should be assessed through the ability to apply patterns in sentences or communication. Short quizzes remain useful for identifying initial mastery but should be followed by productive tasks. For example, after learning family vocabulary and the mubtada'-khabar structure, learners can be asked to write several sentences introducing their family members.

Analytic rubrics can help teachers communicate criteria more transparently. A rubric does not need to contain numerous levels or indicators. Four performance levels can be used when their descriptions identify observable behaviours, such as "not yet demonstrating the ability," "beginning to demonstrate the ability with support," "meeting the criteria," and "exceeding the criteria." These descriptions need to be adapted to the learning objectives and learners' levels. Scores should not stand alone; brief notes about the observed evidence and suggestions for improvement make assessment results more meaningful.

The use of assessment results needs to be planned before a task is administered. Teachers can prepare several possible follow-up actions, such as brief reteaching, scaffolded practice, support groups, opportunities for revision, or enrichment tasks. This planning ensures that assessment does not stop at data collection. If most of the class has not understood the material,

teachers can review their explanations or select examples that are more closely related to learners' experiences. If only a few learners experience difficulties, small-group support may be more appropriate than reteaching the entire class.

Follow-up decisions also need to consider the workload of teachers and learners. Formative assessment conducted too frequently with complicated instruments can reduce learning time and overwhelm teachers with data processing. Teachers can therefore select important moments during instruction to collect evidence, such as after introducing a concept, after guided practice, and before learners complete independent tasks. Concise notes using priority indicators may be more effective than lengthy forms that are not used for decision-making.

Table 1. Summary of formative assessment focuses across the four Arabic language skills

Language skill	**Evidence of learning observed**	**Sample instruments/techniques**	**Direction for follow-up**
Istima' (listening)	Identifies the topic, main idea, detailed information, and keywords in spoken language.	Listening to short dialogues/texts, comprehension questions, keyword note sheets.	Replay difficult sections, practise identifying keywords, reinforce vocabulary, and develop step-by-step listening strategies.
Kalam (speaking)	Message comprehensibility, fluency, pronunciation, vocabulary accuracy, and context-appropriate use of structures.	Pair dialogues, interviews, role-plays, short presentations, observation rubrics.	Pronunciation practice, reinforcement of sentence patterns, repeated dialogue practice, and enrichment.

Language skill	Evidence of learning observed	Sample instruments/techniques	Direction for follow-up
Qira'ah (reading)	Understanding the main idea and detailed information, relationships between ideas, and contextual vocabulary meaning.	Short texts, literal and inferential questions, idea maps, brief summaries.	Strategies for identifying main ideas, contextual vocabulary reinforcement, guided reading, and summarizing practice.
Kitabah (writing)	Development of ideas, vocabulary and structural accuracy, paragraph organization, and ability to revise.	Writing outlines, drafts, analytic rubrics, teacher/peer comments, portfolios.	Targeted revision, sentence-construction practice, structural reinforcement, and comparison of drafts with final writing.

The following rubric is an example of a tool for translating indicators across the four language skills into observable performance descriptors. Teachers can adapt the indicators and weighting according to learning objectives, proficiency levels, and learner characteristics. A four-level scale is used to show differences in achievement progressively, with each level providing guidance for follow-up action rather than merely a numerical score.

Table 2. Sample analytic rubric for formative assessment across the four Arabic language skills

Language skill	4 – Excellent	3 – Good	2 – Fair	1 – Needs guidance
Istima' (listening)	Understands the spoken text and accurately identifies the main information.	Understands most of the spoken text, with only a few errors.	Understands the general topic but still has difficulty identifying important information.	Does not yet understand the spoken text and requires assistance.

Language skill	4 – Excellent	3 – Good	2 – Fair	1 – Needs guidance
Kalam (speaking)	Conveys messages clearly and fluently, using appropriate vocabulary and structures.	The message is understandable, although there are minor vocabulary or structural errors.	The message is not yet fluent, and some parts are difficult to understand.	Is not yet able to convey a clear message and requires guidance.
Qira'ah (reading)	Understands the text, identifies important information, and draws accurate conclusions.	Understands most of the text, with only a few errors.	Understands the topic of the text but still has difficulty identifying information and drawing conclusions.	Is not yet able to understand the text and requires assistance.
Kitabah (writing)	Writes coherent and comprehensible sentences or paragraphs using appropriate vocabulary and structures.	The writing is understandable and reasonably coherent, with a few errors.	The writing is not yet coherent and contains several errors that interfere with comprehension.	Is not yet able to produce comprehensible writing and requires guidance.

In formative use, rubric scores should not be treated as the ultimate goal. Teachers need to examine achievement patterns across each aspect to determine follow-up action. For example, learners who receive low scores for message comprehensibility need practice in organizing ideas and guided dialogue; learners who communicate well but remain weak in grammatical structure can be given targeted practice with sentence patterns. Learners can also use the same rubric for self-reflection before and after revision, making changes in performance visible.

Scoring guidelines: select one score for each language skill based on the descriptor that best matches the learner's work. If all four skills are assessed, the maximum score is 16. The grade can be calculated using the formula

(score obtained ÷ 16) × 100. For example, scores of 3 for istima', 3 for kalam, 2 for qira'ah, and 4 for kitabah produce a grade of 12 ÷ 16 × 100 = 75. Teachers should still add a brief note describing the abilities that have been demonstrated and the aspects requiring improvement.

Usage note: the kalam rubric should be used to provide feedback on priority indicators. Teachers do not need to score every aspect in a single performance; the assessment focus may change according to the objective of the lesson.

C.3 Sample Assessment Designs and Follow-up Actions

The first design is intended for Grade XI students at an Islamic senior high school (Madrasah Aliyah), using the topic الحوار عن الهوايات (dialogue about hobbies) with a focus on the use of fi'il mudhari' (the present-tense verb). The learning objective is for learners to conduct a simple dialogue about hobbies using fi'il mudhari' forms that agree with the dhamir (pronoun). Before the activity, the teacher communicates the objective and success criteria: the use of at least five relevant hobby-related vocabulary items, accurate verb forms for first-person and second-person pronouns, and a comprehensible dialogue. Learners then prepare dialogues in pairs and perform them in front of the class. During the activity, the teacher uses a simple observation sheet to record vocabulary accuracy, grammatical structure, pronunciation, and fluency. After the performance, the teacher identifies two aspects that were done well and one priority area for improvement. Learners are given time to revise their dialogues and perform them again so that the use of feedback can be observed in their subsequent performance.

Table 3. Operational design of formative assessment for maharah kalam

Aspect	Formative assessment design for maharah kalam
Learning objective	Learners are able to conduct a simple dialogue about hobbies using at least five relevant vocabulary items and fi'il mudhari' forms appropriate to the dhamir.

Aspect	Formative assessment design for maharah kalam
Task and procedure	Learners prepare a dialogue in pairs lasting 3–5 minutes. The teacher communicates the objectives and criteria before the practice. The teacher observes the dialogue using a rubric; learners complete a self-reflection checklist.
Evidence of learning	Dialogue performance, vocabulary accuracy, use of fi'il mudhari', pronunciation, fluency, and ability to respond to an interlocutor.
Success criteria	Learners use vocabulary relevant to the topic, select verb forms appropriate to the dhamir, convey an understandable message, and respond to their interlocutor. Performance levels: needs guidance, developing, meets expectations, and excellent.
Sample feedback	"Your hobby-related vocabulary is appropriate and your message is understandable. Pay attention to the fi'il mudhari' form for the pronoun أنتَ, which you still confuse with أَنَا. Practise the أنتَ and أَنَا forms using three verbs, then perform the dialogue again."
Follow-up action	The teacher groups errors by type. Learners who need support complete conjugation exercises and guided dialogue practice; learners who have met the criteria extend their dialogue. The repeated performance is compared with the initial performance.

Usage note: a portfolio shows the revision process, not merely the score for the final piece of writing. Teachers can ask learners to mark the changes they made after receiving feedback so that the learning process becomes more visible.

The second design is intended for Grade X students at an Islamic senior high school (Madrasah Aliyah), using the topic مفردات وكتابة فقرة قصيرة عن الأسرة (vocabulary and writing a short paragraph about the family). The learning objective is for learners to write a paragraph of four to six sentences introducing family members using simple vocabulary and structures. The success indicators include the use of at least six family-related vocabulary items, accurate use of the mubtada'-khabar structure, and paragraph coherence. Learners first prepare a draft using the vocabulary they have studied. The teacher then provides feedback on the aspects that most affect the comprehensibility of the writing, such as vocabulary choice, structural

agreement, and the sequence of information. Learners use the comments to revise their drafts before submitting the final text. The draft and revised version can be kept in a simple portfolio to demonstrate the development of writing ability.

Table 4. Operational design of formative assessment for maharah kitabah and mufradat

Aspect	Formative assessment design for kitabah and mufradat
Learning objective	Learners are able to write a paragraph of 4–6 sentences about their family using at least six relevant vocabulary items and a simple mubtada'-khabar structure.
Task and procedure	Learners prepare an outline, write an initial draft, receive feedback, and then revise their writing. At the end of the activity, learners complete an exit ticket identifying the vocabulary they have mastered and the aspects they still find difficult.
Evidence of learning	Outline, initial draft, feedback notes, revised text, use of family-related vocabulary, structural accuracy, and paragraph coherence.
Success criteria	The text contains 4–6 relevant sentences, uses at least six family-related vocabulary items accurately, applies the structures studied, and presents information coherently.
Sample feedback	"Your paragraph already includes information about your family. Check whether the demonstrative هٰذِهِ agrees with feminine nouns such as أُخْتِي. Revise that sentence and add one piece of information to clarify the relationship between family members."
Follow-up action	Learners revise their writing. The teacher compares the draft and revised version in the portfolio, identifies recurring errors, and provides small-group practice to reinforce grammatical structures.

The two designs show that follow-up decisions need to be derived from the evidence collected. In the kalam task, evidence from dialogue performance is used to determine subsequent oral practice, while the opportunity to repeat the task helps teachers determine whether learners have used the feedback. In the kitabah task, evidence from drafts and revised versions is used to monitor writing development and identify recurring errors. A simple

portfolio can help teachers and learners observe changes in work quality over time. The frequency of portfolio collection needs to be adjusted to the duration of the learning unit and the class workload so that documentation remains useful without creating excessive administrative demands.

Assessment designs should be regarded as adaptable examples rather than fixed formats for every class. Teachers need to consider class size, instructional time, media availability, proficiency level, and unit objectives. In a class with many learners, dialogues can be conducted in pairs or small groups using an observation sheet containing only priority indicators. Teachers can assess several pairs in rotation while the other pairs conduct self-assessment using a checklist. This arrangement helps make assessment manageable without sacrificing learners' opportunities to practise.

To maintain consistency, teachers can use concise evidence records. These records do not need to contain lengthy descriptions for every learner; they need only document the indicators achieved, the main difficulty, and the follow-up action. For example, simple codes can be used to indicate a need for reinforcement in vocabulary, structure, pronunciation, or organization of ideas. The codes must be accompanied by explanations that are understandable to teachers and, where necessary, to learners. The data can then be reviewed in the next activity to determine whether the difficulty has decreased or requires a different strategy.

Assessment implementation also needs to consider affective aspects. Learners who are afraid of making mistakes may remain silent when asked to speak, preventing teachers from obtaining an accurate picture of their actual abilities. A supportive classroom atmosphere, opportunities to practise in pairs, and feedback focused on aspects that can be improved can help reduce this pressure. Teachers still need to maintain achievement standards, but the way assessment results are communicated should encourage learners to take risks and show willingness to improve their work.

The main implication for teachers is that formative assessment needs to be designed alongside learning objectives rather than added after activities have

been completed. Teachers should select several priority indicators, use simple instruments, record patterns of difficulty, and provide opportunities for improvement. Too many instruments or excessively detailed criteria may make it difficult for teachers to provide prompt and targeted feedback. Assessment results do not always need to be converted into numerical grades; descriptive notes indicating demonstrated abilities and next steps are often more useful for guiding learning. In this way, assessment can become a natural part of classroom interaction.

At the learner level, the habit of receiving and using feedback needs to be developed gradually. Teachers can begin with a simple checklist, brief reflection questions, and opportunities to revise tasks. Once learners understand the criteria, their involvement in self-assessment and peer assessment can be expanded. Teachers need to create a classroom atmosphere in which learners view errors as information for learning rather than merely as reasons for receiving low grades. This approach positions learners as active participants who can identify their needs, monitor their progress, and take steps towards improvement more independently.

D. Conclusion

Formative assessment in Arabic language learning is a cycle of using evidence of learning to improve instruction and help learners develop independent learning. Its formative nature is not determined by the type of test or instrument but by the use of information to identify gaps, provide specific feedback, and determine follow-up action. Its implementation needs to be adapted to the characteristics of istima', kalam, qira'ah, kitabah, mufradat, and qawa'id. The assessment designs based on dialogue and paragraph writing show that feedback becomes more meaningful when learners are given genuine opportunities to repeat, revise, or extend their performance. Teachers therefore need to integrate objectives, criteria, evidence collection, feedback, and follow-up action into a single instructional sequence. Since this article is a literature review presenting conceptual

designs, the effectiveness of the sample instruments still needs to be examined through empirical research in different Arabic language learning contexts.

Acknowledgments

The author would like to thank the lecturer of the Arabic Language Learning Evaluation Development course for the academic guidance provided during the preparation of the paper on which this article is based.

References

Akmalia, F., & Faizin, K. (2021). Recitation: An alternative method for virtual nahwu learning. *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban, 5*(1), 97–110. https://doi.org/10.15575/jpba.v5i1.12435

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*, Article 87. https://doi.org/10.3389/feduc.2019.00087

Bamualim, M. (2020). Kedudukan dan tujuan evaluasi pembelajaran bahasa Arab. *Jurnal Al-Fawa'id: Jurnal Agama dan Bahasa, 10*(2), 1–9. https://doi.org/10.54214/alfawaid.Vol10.Iss2.141

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Darwin, D., Boeriswati, E., & Murtadho, F. (2023). Asesmen pembelajaran bahasa dalam Kurikulum Merdeka Belajar pada siswa SMA. *Lingua Rima: Jurnal Pendidikan Bahasa dan Sastra Indonesia, 12*(2), 25–36. https://doi.org/10.31000/lgrm.v12i2.8639

Faizin, K. (2020). Analisis penggunaan metode penelitian evaluasi pada penelitian bahasa Arab model pengembangan. *Tabyin: Jurnal Pendidikan Islam, 2*(1), 39–53. https://doi.org/10.52166/tabyin.v2i1.27

Hamidah, H., Razida, M., & Qomari, N. (2025). Formative assessment as a learning tool: Enhancing Arabic vocabulary and grammar through interactive evaluation in a non-formal setting. *Asatiza: Jurnal Pendidikan, 6*(3), 344–353. https://doi.org/10.46963/asatiza.v6i3.3043

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2022). *Panduan pembelajaran dan asesmen: Pendidikan anak usia dini, pendidikan dasar, dan pendidikan menengah*. Pusat Kurikulum dan Pembelajaran. https://kurikulum.kemdikbud.go.id/file/1720050633_manage_file.pdf

Munawaroh, B., & Faizin, K. (2025). Analysis of student learning interest in Arabic language subjects at MTs Manba'ul Huda Sekaran-Balen. *Assuthur: Jurnal Pendidikan Bahasa Arab, 4*(1), 36–44. https://doi.org/10.58194/as.v4i1.1763

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Ridho, U. (2018). Evaluasi dalam pembelajaran bahasa Arab. *An-Nabighoh: Jurnal Pendidikan dan Pembelajaran Bahasa Arab, 20*(1), 19–26. https://doi.org/10.32332/an-nabighoh.v20i01.1124

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714

Sari, R. K. (2021). Penelitian kepustakaan dalam penelitian pengembangan pendidikan bahasa Indonesia. *Jurnal Borneo Humaniora, 4*(2), 60–69. http://jurnal.borneo.ac.id/index.php/borneo_humaniora/article/view/2249

Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research, 103*, Article 101602. https://doi.org/10.1016/j.ijer.2020.101602

Wanti, A. I., A'yuni, Q., & Chamidah, D. (2022). Model dan praktik: Asesmen formatif non paper-based dalam pembelajaran bahasa Arab. *Arabi: Journal of Arabic Studies, 7*(1), 76–92. https://doi.org/10.52593/ARJAS.2022.7.1.06

Wolsey, T. D., & Karkouti, I. M. (2024). Formative assessment and feedback for language learners. In *The TESOL encyclopedia of English language teaching*. https://doi.org/10.1002/9781118784235.eelt1061

Zaharo, Chamidah, D., & Jundi, M. (2024). Word puzzle-based formative assessments track student progress in Arabic language learning. *Kilmatuna: Journal of Arabic Education, 4*(1), 45–60. https://doi.org/10.55352/pba.v4i1.871