



## Formative Assessment as a Learning Tool: Enhancing Arabic Vocabulary and Grammar through Interactive Evaluation in a Non-Formal Setting

\*Husnul Hamidah<sup>1,a\*</sup>, Mirdawati Razida<sup>2,b</sup>, Nur Qomari<sup>3,c</sup>

<sup>1,2,3</sup> State Islamic University of Maulana Malik Ibrahim Malang, East Java, Indonesia

<sup>a</sup>[husnulhamidah90@gmail.com](mailto:husnulhamidah90@gmail.com), <sup>b</sup>[mirdarasidah@gmail.com](mailto:mirdarasidah@gmail.com), <sup>c</sup>[qomari@uin-malang.ac.id](mailto:qomari@uin-malang.ac.id)

### ARTICLE INFORMATION

#### Article History:

Received : 30/06/2025

Revised : 28/09/2025

Accepted : 29/09/2025

Published : 30/09/2025

#### Keywords:

Arabic Language Pedagogy;

Formative Assessment; Non-

Formal Education;

Vocabulary Acquisition

#### DOI:

<https://doi.org/10.46963/asatiza.v6i3.3043>

#### \*Correspondence Author:

[husnulhamidah90@gmail.com](mailto:husnulhamidah90@gmail.com)  
[m](mailto:husnulhamidah90@gmail.com)

©Authors (2025). Licensed under [CC BY SA](https://creativecommons.org/licenses/by-sa/4.0/)

#### How to cite this article:

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>

### INTRODUCTION

Mastery of vocabulary (mufradat) and grammar (qawa'id) constitutes the foundational bedrock of Arabic language acquisition, serving as an indispensable prerequisite for developing comprehensive communicative competence across listening, speaking,

reading, and writing domains (Ainin, 2006; Bamualim, 2020). A learner's inability to build a robust lexicon directly impedes textual comprehension and the construction of coherent utterances, while a deficient grasp of complex Arabic grammatical structures, such as i'rab, hinders the analytical processes required

to decode and produce syntactically accurate sentences (Pramukti et al., 2021). Consequently, pedagogical efforts to enhance these core linguistic components remain a critical priority across diverse educational landscapes. However, prevalent instructional practices often prioritize content delivery and conventional exercises, relegating evaluation to a terminal, summative function aimed chiefly at measuring and grading student achievement (Atikah & Amelia, 2024; Ridho, 2018). This limited perspective overlooks the broader strategic potential of assessment as a diagnostic, formative, and reflective mechanism for both educators and learners, thereby neglecting its capacity to actively drive the learning process itself.

The theoretical underpinnings of educational assessment affirm its dual role as both a measurement tool and an integral component of pedagogy. Seminal work by Black and Wiliam (1998) established formative assessment as a powerful means of improving learning through continuous feedback and instructional adjustment, rather than merely auditing outcomes. This aligns with Scriven's (1967) foundational distinction, which underscores the complementary functions of formative and summative assessment in supporting educational development. Within language pedagogy, these principles demonstrate that well-integrated assessment can significantly enhance student engagement, foster learner autonomy, and deepen conceptual understanding (Hattie & Timperley, 2007; Lee, 2007). Effective formative evaluation enables instructors to identify specific

learning gaps and design more targeted, responsive instructional strategies (Iskandar & Rasmitadila, 2024; Wiliam, 2011).

Despite this established potential, the evaluation tools traditionally employed in Arabic language pedagogy often fail to support the holistic development of linguistic skills. As noted by Ainin (2006) and Ridho (2018), many instruments overemphasize rote memorization and deploy closed-ended questions, rendering them ineffective for measuring critical thinking and the authentic application of grammatical rules in communicative contexts. Furthermore, teachers may not consistently utilize evaluation data to inform and refine their instructional practices (Khofiah & Nasiruddin, 2024). This gap is particularly pronounced in non-formal educational settings, which cater to diverse learner objectives and often employ more flexible, communicative pedagogical approaches than traditional institutions (Rogers, 2014; Zaid et al., 2024). While existing research on Arabic language learning has extensively covered teaching strategies, instructional media, and classroom interaction, there has been comparatively limited scholarly attention devoted to how the design and implementation of evaluation tools themselves shape the acquisition of vocabulary and grammar. Research within dynamic non-formal settings, such as the LKP Kampung Arab Al-Azhar Pare, which emphasizes communicative competence, remains especially scarce and underexplored.

The novelty of this research lies in its conceptualization of evaluation tools

not merely as instruments of measurement, but as active pedagogical strategies that directly contribute to students' linguistic development. Whereas previous studies, such as that by Khofiah and Nasiruddin (2024), have primarily focused on teaching methods or media, this study posits that the design, integration, and utilization of assessment are themselves powerful drivers of learning. Therefore, this study aims to provide an in-depth exploration of the role of evaluation tools in Arabic language learning, with a specific focus on vocabulary and grammar mastery within a non-formal educational context. It seeks not only to describe the forms and types of evaluation tools used but also to examine how these tools are integrated into the learning process, how they are perceived by teachers and students, and the extent of their impact on students' engagement and conceptual understanding. Using a descriptive qualitative approach, this study is expected to offer theoretical and practical contributions toward developing a more effective and humanistic Arabic language learning evaluation system.

The research is guided by the following questions: (1) What types of evaluation tools are used to assess vocabulary and grammar mastery at LKP Kampung Arab Al-Azhar Pare? (2) How do teachers and students perceive the role of these evaluation tools in the learning process? (3) What impact do these evaluation tools have on students' engagement and conceptual understanding?

## METHOD

This study employed a qualitative descriptive approach to provide a systematic and in-depth exploration of the role of evaluation tools in enhancing students' mastery of Arabic vocabulary and grammar within a non-formal educational setting. This methodology was selected for its capacity to facilitate a direct exploration of participants' lived experiences, perceptions, and practices in their natural context, prioritizing depth of understanding over quantification (Creswell & Poth, 2018). The research was conducted at LKP Kampung Arab Al-Azhar Pare in Pare, Kediri, East Java, Indonesia, a recognized non-formal institution specializing in intensive Arabic language acquisition. This site was purposefully selected as a critical case due to its explicit pedagogical focus on communicative competence and its reputed use of diverse and innovative pedagogical tools, making it an ideal context for investigating the integration of assessment into learning.

Participants were comprised of Arabic language teachers and students who were actively engaged in programs focusing on vocabulary and grammar development. A purposive sampling technique was utilized to select individuals who could provide rich, information-dense insights by virtue of their direct and sustained involvement with the evaluation tools under investigation (Palinkas et al., 2015). The final sample consisted of five teachers, with teaching experience at the institution ranging from three to ten years, and twenty students, who had been enrolled in the program for a minimum of

one intensive course cycle (approximately three months). The student cohort included 12 male and 8 female participants, aged 18 to 25, with diverse prior educational backgrounds. This sample size was determined to be sufficient to achieve data saturation, the point at which new data collection no longer yielded novel thematic insights or properties (Creswell & Poth, 2018).

Data collection was triangulated through three primary methods to ensure credibility and trustworthiness of the findings: semi-structured interviews, non-participant observation, and document analysis (Setyosari, 2010). The semi-structured interviews were conducted individually with all 25 participants using interview protocols tailored for teachers and students. These interviews, which lasted between 30 to 60 minutes, were designed to gather in-depth insights into their personal experiences with various evaluation tools, their perceptions of the tools' efficacy in supporting learning, and their interpretations of the tools' role in the instructional process. All interviews were audio-recorded and subsequently transcribed verbatim. Secondly, non-participant observations were carried out over 15 separate classroom sessions, totaling approximately 30 hours of observation. An observation protocol was used to record real-time practices related to the administration of assessments, student behavioral and cognitive engagement during evaluation activities, and the nature of teacher-student interactions following feedback. Finally, document analysis was performed on a range of relevant artifacts, including

institutional syllabi, detailed lesson plans, samples of student assignments, the design of various quizzes and tests, and institutional evaluation records. This analysis provided crucial corroborating evidence and a deeper contextual understanding of the formal and informal assessment ecosystem.

For data analysis, the study adopted the interactive model of qualitative data analysis proposed by Miles et al. (2014). This model involves three concurrent, iterative flows of activity: data reduction, data display, and conclusion drawing and verification. The process commenced with data reduction, whereby the voluminous raw data from interview transcripts, field notes, and documents were systematically coded and categorized to simplify and focus on emerging themes. This was followed by data display, which involved organizing the condensed information into structured narratives and thematic matrices to allow for the visual identification of patterns, relationships, and discrepancies. The final stage, conclusion drawing and verification, entailed interpreting the displayed data to generate meaningful and evidence-based conclusions. These conclusions were continually checked for validity through peer debriefing with fellow researchers and by conducting member checks, whereby preliminary findings were presented to a subset of participants to confirm the accuracy of the interpretations. This rigorous, cyclical process ensured that the final findings remained firmly grounded in the empirical evidence collected from the field.

Ethical considerations were paramount throughout the research process. Prior to data collection, ethical approval was sought and obtained from the institutional review board of the authors' affiliated university. Informed consent was secured from all participants, with a clear explanation of the study's purpose, their right to withdraw at any time without penalty, and the measures in place to ensure confidentiality. All identifying information, including names of individuals and the institution in published reports, was anonymized using pseudonyms to protect participant privacy. Data, including audio recordings and transcripts, were stored on a password-protected secure server accessible only to the core research team.

## RESULT AND DISCUSSION

### Result

#### *Typology and Application of Evaluation Tools*

The findings reveal a sophisticated evaluation ecosystem at LKP Kampung Arab Al-Azhar Pare, which strategically merges traditional classifications with digital innovation. The institution operationalizes a comprehensive typology encompassing both test and non-test instruments, as outlined in foundational frameworks (Ainin, 2006; Magdalena et al., 2021). Test-based evaluations are rigorously applied through multiple modalities. Written tests, administered both on paper and digitally via platforms like Google Forms, are used to assess discrete knowledge. For instance, objective formats target vocabulary recall, while subjective formats, such as essay prompts, demand the application of

grammatical rules in sentence construction. A teacher elucidated the diagnostic value of this approach, stating, *"The essay questions force them to actively assemble the language, not just recognize it. We can see their thought process and pinpoint the exact moment of grammatical breakdown."* Complementing this, oral tests form a cornerstone of assessment, with frequent spontaneous question-and-answer sessions allowing for real-time evaluation of pronunciation, fluency, and pragmatic language use.

Non-test instruments are deployed with equal intentionality to capture competencies in the affective and psychomotor domains. These include performance assessments like role-plays simulating real-world interactions and project-based evaluations such as creating and performing short dialogues. Analysis of student portfolios provided tangible evidence of growth, showcasing multiple drafts of writing assignments where feedback on vocabulary and grammar had been incrementally incorporated, thereby framing the portfolio as both a record of progress and a reflective assessment tool.

#### *Perceptions of Evaluation as a Learning Tool*

A central finding is the shared perception among teachers and students that evaluation is an integral, formative component of the learning journey, rather than a series of isolated judgment events. Teachers reported a conscious pedagogical shift towards using assessments formatively. One teacher emphasized, *"My goal is not to give a score, but to get a snapshot of their understanding at that moment. The quiz tells me what I need to*

*re-teach or practice more tomorrow.*" Students internalized this philosophy, perceiving evaluations as opportunities for growth. A student's reflection on gamified tools highlighted this: *"When we do a Kahoot game, it doesn't feel like a test. It feels like a challenge, and when I get a word wrong, I remember it better for next time because I want to win."* This perception was markedly stronger for interactive and collaborative evaluations, which were consistently described as less stressful and more beneficial than traditional, high-stakes tests.

#### *Empirical Impact on Vocabulary and Grammar Mastery*

Observation and interview data provide compelling evidence of the tangible impact of these strategies on language acquisition. In vocabulary development, the synergy between receptive tools (e.g., matching exercises) and productive tools (e.g., role-plays) effectively reinforced lexical knowledge. During a Wordwall activity, students demonstrated high engagement and rapid recall. More significantly, in subsequent, less structured activities, students were observed correctly deploying vocabulary that had been practiced in these gamified evaluations, with several instances of self-correction mid-dialogue, indicating the internalization of feedback.

In the domain of grammar, the impact was observed as a shift from abstract memorization to functional application. Analytical tasks, such as sentence transformation and peer error-correction, fostered a deeper, more critical engagement with grammatical structures. A vivid example was observed during a

group activity where students collaboratively corrected a paragraph riddled with errors. The ensuing discussion revealed students actively debating grammatical rules, articulating reasons for their corrections. Interview data confirmed this cognitive shift. One student explained, *"Before, grammar was just memorizing rules. Now, with the group work and having to find mistakes in others' writing, I understand why the rule exists and how to use it."* Across both domains, the provision of specific, explanatory feedback from teachers was repeatedly cited as the single most critical factor in deepening understanding and fostering a growth mindset.

#### **Discussion**

The findings from this study make several significant contributions to the discourse on language pedagogy and assessment. The identified typology confirms the value of a diverse assessment repertoire (Ainin, 2006), but its primary contribution lies in illustrating how this repertoire is animated by a pervasive formative philosophy within a non-formal context. The strategic digital integration, such as using Google Forms not just for efficiency but for its diagnostic analytics, exemplifies a pragmatic approach to leveraging technology for actionable pedagogical insights, aligning with contemporary explorations of EdTech in language learning (Rahayu & Purnawarman, 2019).

The most salient contribution is the empirical demonstration of how evaluation can be repositioned as a dynamic learning strategy. The high engagement with gamified tools supports

the core principles of formative assessment (Black & Wiliam, 1998), suggesting their success stems from creating a low-stakes environment for essential retrieval practice and delivering immediate feedback, which is crucial for memory consolidation and metacognitive development (Hattie & Timperley, 2007). Furthermore, the integration of evaluation into productive, communicative tasks underscores that assessment is most effective when it mirrors authentic language use, a cornerstone of task-based language teaching.

The findings from the grammar domain offer a powerful challenge to traditional, transmission-based approaches. The efficacy of collaborative tasks like peer correction highlights the vital role of social constructivist approaches. When students articulate and negotiate grammatical rules, they co-construct a deeper, procedural knowledge that moves beyond rote memorization, resonating with Vygotskian perspectives on social learning. The indispensable role of explanatory, task-focused teacher feedback, as consistently reported by participants, perfectly embodies the spirit of formative assessment and is a critical factor in building self-efficacy (Hattie & Timperley, 2007).

In conclusion, this study provides a coherent model from a non-formal setting for humanizing assessment. It demonstrates that when evaluation is varied, interactive, and embedded within a cycle of feedback and instructional adjustment, it transcends its traditional role as a benchmark. It becomes, instead, a powerful engine for vocabulary and

grammar acquisition, effectively bridging the artificial divide between teaching and testing. This approach fosters a cohesive learning environment that successfully addresses both the cognitive and affective dimensions of language learning.

## CONCLUSION

This study set out to investigate the types, perceptions, and impact of evaluation tools on Arabic vocabulary and grammar mastery within the non-formal educational context of LKP Kampung Arab Al-Azhar Pare. In direct answer to the research questions, the findings reveal a comprehensive typology of tools, encompassing both test and non-test instruments, delivered through traditional and digital means. Crucially, both teachers and students perceive these tools not as terminal judgments but as formative, integral components of the learning process. The empirical evidence demonstrates that this formative application has a tangible impact, enhancing student engagement, deepening conceptual understanding of grammar, and reinforcing the acquisition and active use of vocabulary.

The study's primary theoretical contribution is its elaboration of formative assessment theory within the specific domain of Arabic language pedagogy in a non-formal setting. It empirically validates the principles of Black and Wiliam (1998) and Hattie and Timperley (2007) by demonstrating how a diverse ecosystem of evaluations—from gamified quizzes to collaborative peer review—generates evidence for learning and provides the feedback necessary to close the gap between current and desired

understanding. Furthermore, it extends this theory by showing how such an approach can successfully demystify and make accessible a complex linguistic area like Arabic grammar, shifting learner engagement from rote memorization to analytical application through social constructivist practices.

The practical implications of these findings are significant for educators and curriculum developers in similar non-formal or communicative-focused language institutions. The research provides a replicable model where evaluation is seamlessly blended with instruction. Key recommendations include: moving beyond a narrow reliance on summative tests to incorporate a cycle of low-stakes, interactive assessments; leveraging digital tools not only for efficiency but for their diagnostic and motivational potential; and prioritizing the provision of specific, task-focused feedback that guides subsequent learning. This model proves that assessment can be a central driver of a positive and effective learning culture.

Despite these contributions, this study has several limitations. As a qualitative case study, its findings are context-specific and not statistically generalizable. The research was conducted in a single institution renowned for its innovative practices, and thus the transferability of the model to more traditional or resource-constrained settings requires further investigation. The study also relied on self-reported data from participants, which, while rich, is subject to perception biases.

Future research should, therefore, seek to validate these findings through mixed-methods or quantitative studies across a broader range of non-formal institutions. Longitudinal research tracking the long-term retention of vocabulary and grammar knowledge acquired through such formative systems would be highly valuable. Additionally, investigating the specific cognitive load and learning mechanisms at play in gamified Arabic grammar evaluation could provide a deeper theoretical understanding of why these tools are effective. Ultimately, this study underscores that in the pursuit of Arabic language proficiency, the strategic integration of evaluation into the fabric of pedagogy is not an optional enhancement but a fundamental imperative for fostering deep, functional, and humanistic learning.

## REFERENCES

- Ainin. (2006). *Evaluasi dalam Pembelajaran Bahasa Arab (Teori dan Praktek)*. Malang: Misykat.
- Atikah, & Amelia. (2024). Strategi penilaian dan evaluasi efektif untuk meningkatkan kemampuan siswa serta umpan balik dalam pembelajaran bahasa Indonesia. *Jurnal Pendidikan dan Ilmu Bahasa*, 2(3), 76–84. <https://doi.org/10.59059/perspektif.v2i3.1454>
- Bamualim, M. (2020). Kedudukan Dan Tujuan Evaluasi Pembelajaran Bahasa Arab. *Jurnal Al-Fawa'id: Jurnal Agama dan Bahasa*, 10(2), 1–10. <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>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Iskandar, N. M., & Rasmitadila. (2024). Peningkatan Kualitas Pembelajaran melalui Evaluasi yang Efektif: Tinjauan Terhadap Praktik dan Metode Evaluasi. *Karimah Tauhid*, 3(2), 2270–2287. <https://doi.org/10.30997/karimahtauhid.v3i2.11945>
- Khofiah, & Nasiruddin. (2024). *Interactive Learning of Arabic Grammar at Al-Azhar Course*, 9(November), 305–317. <https://doi.org/10.21462/ijefl.v9i2.813>
- Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing*, 12(3), 180–198. <https://doi.org/10.1016/j.asw.2008.02.003>
- Magdalena, I., Utami, W. D., Rahma, S. B., Anggraini, I. A., & Wijaya, R. R. (2021). *Instrumen Tes dan Non Tes dalam Evaluasi Pembelajaran*. 8, 30461–30472.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Pramukti, R. H., Ngatman, N., & Chamdani, M. (2021). Pengaruh penguasaan kosakata terhadap pemahaman isi wacana pada siswa kelas v SD se-Kecamatan Banyumas tahun ajaran 2020/2021. *Kalam Cendekia: Jurnal Ilmiah Kependidikan*, 9(3). <https://doi.org/10.20961/jkc.v9i3.53892>
- Rahayu, I. S. D., & Purnawarman, P. (2019, June). The use of quizizz in improving students' grammar understanding through self-assessment. In *Eleventh Conference on Applied Linguistics (CONAPLIN 2018)* (pp. 102-106). Atlantis Press. <https://doi.org/10.2991/conaplin-18.2019.235>
- Ridho, U. (2018). Evaluasi dalam pembelajaran bahasa Arab. *An Nabighoh Jurnal Pendidikan Dan Pembelajaran Bahasa Arab*, 20(01), 19. <https://doi.org/10.32332/an-nabighoh.v20i01.1124>
- Rogers, A. (2014). *The base of the iceberg: Informal learning and its impact on formal and non-formal learning*. Verlag Barbara Budrich.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum*

- evaluation* (pp. 39–83). Rand McNally.
- Setyosari, P. (2010). *Metodologi Penelitian Pendidikan dan Pengembangan*. Jakarta: Kencana.
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Zaid, A. H. bin, Widyanti, L. N., Ismail, M., & Jannah, D. A. M. (2024). Implementasi pendekatan komunikatif (Communication Approach) dalam meningkatkan keterampilan berbahasa Arab. *Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab dan Kebahasaaraban*, 7(2), 682. <https://doi.org/10.35931/am.v7i2.3769>