

## DESIGN OF COMMUNICATIVE ARABIC LISTENING SKILL ASSESSMENT TO MEASURE THE LANGUAGE SKILLS OF ISLAMIC JUNIOR HIGH SCHOOL STUDENTS

Andini<sup>1</sup>, Aswin Aan Efriyanto<sup>2</sup>, R. Umi Baroroh<sup>3</sup>,  
Adila Umniati Aufa<sup>4</sup>, Muhammad Zaqlul Rafif<sup>5</sup>

<sup>1,4,5</sup>Universitas Islam Negeri Sunan Kalijaga Yogyakarta; Indonesia

<sup>2</sup>Institut Agama Islam Negeri Syaikh Abdurrahman Siddik Bangka Belitung; Indonesia

Correspondence Email; 25204021025@student.uin-suka.ac.id

Submitted: 07/07/2026

Revised: 28/07/2026

Accepted: 10/08/2026

Published: 15/08/2026

### Abstract

This study aims to develop a comprehensive Arabic Listening Skill assessment instrument based on a communicative approach tailored for seventh-grade students at MTsN 3 Bogor. The research employs a structured Research and Development (R&D) methodology utilizing the ADDIE instructional design model, specifically limited to the Analyze, Design, and Develop stages. Primary data were obtained directly from subject-matter and evaluation experts, while secondary data were gathered from curriculum documents, syllabi, and relevant literature. Data collection techniques involved observation, interviews, validation questionnaires, and documentation. Furthermore, the collected data were analyzed using both qualitative descriptive analysis to interpret expert feedback and suggestions, and quantitative descriptive analysis to determine the validity and practicality of the developed assessment instrument. The findings indicate the successful development of a communicative Arabic Listening Skill assessment instrument comprising 10 items. These items are formulated through five alternative techniques: Oral Question-and-Answer, Role Play, Oral Retelling, Information Gap, and Task-Based Listening. Furthermore, the developed instrument demonstrated strong theoretical validity, confirming its effectiveness and relevance as a tool for measuring students' authentic Arabic listening proficiency in real-life communication contexts.

### Keywords

ADDIE; Arabic Listening Skill; Assessment Instrument; Communicative Approach; Innovation.



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

## INTRODUCTION

The philosophical and empirical foundation of this research is rooted in the crucial role of Arabic Listening Skill as the primary cognitive basis of the entire language acquisition process. Azhar (2022) asserts that the listening process is not a passive activity like a voice recorder but rather a complex psycholinguistic process that requires the simultaneous processing of phonological, syntactic, semantic, and pragmatic information. On an ideal level (ideality), modern curriculum regulations demand that learners possess functional communicative competence. However, the reality in *most Madrasah Tsanawiyah* (Islamic Junior High Schools/MTs) currently reveals a profound academic anxiety. The assessment of Arabic Listening Skill in the field is nothing more than a "functional illusion." Educators formally claim to be testing listening skills; yet, in practice, the assessment instruments provided to students are dominated by written reading texts and rigid grammatical demands without adequate audio stimuli (Rossi & Brunfaut, 2021).

The wide gap between curriculum ideals and the reality of conventional evaluation systematically traps students in the pitfall of memorizing literal vocabulary (*mufradat*). When listening evaluation instruments are reduced to merely disguised reading (*qira'ah*) tests, the essence of spontaneous meaning comprehension becomes paralyzed (Putri et al., 2026). Furthermore, this rigid evaluation approach psychologically increases language anxiety (and thickens the affective filter) among students, as it isolates language activities from their natural socio-cultural contexts (Ekawati et al., 2024). As a result of the persistence of this obsolete evaluation model, various empirical observations indicate that the mastery level of students' functional Arabic competence in madrasahs remains suboptimal (Maryani et al., 2024).

At MTsN 3 Bogor, the methodology for learning Arabic, specifically in the context of Arabic Listening Skill, still heavily relies on conventional teacher-centered approaches. Based on classroom observations conducted in Grade VII Arabic classes at MTsN 3 Bogor, listening instruction is generally delivered through a teacher-centered approach. Teachers predominantly explain vocabulary and translate textbook dialogues before asking students to answer comprehension questions. Audio recordings are rarely used, and listening activities are commonly replaced by reading aloud written texts. Interviews with two Arabic language teachers revealed that the assessment of *maharah istima'* is generally conducted using printed questions without authentic audio recordings. Documentation of semester examinations also indicates that listening tests primarily measure students' reading comprehension rather than their ability to comprehend spoken

Arabic.

The core problem in the learning process of Arabic Listening Skill at MTsN 3 Bogor is directly linked to the evaluation instruments utilized during assessments. Historically, assessment tools have lacked adequate audio, forcing students to rely on their reading skills rather than their listening comprehension. This discrepancy causes students to struggle to grasp implicit meanings or respond pragmatically in spoken Arabic. These findings demonstrate that the existing assessment instruments have not yet measured students' listening competence in authentic communicative situations. Consequently, the development of a communicative listening assessment instrument is necessary to provide a more valid measurement of students' listening comprehension, pragmatic interpretation, and sociolinguistic competence.

The global shift toward authentic assessment has been examined in several previous studies, yet it still leaves significant theoretical and practical gaps. A critical study by Husna & Bakar (2025) proves that the current assessment of Arabic Listening Skill in textbooks is highly monotonous and does not aim at Higher Order Thinking Skills (HOTS). Nevertheless, this study has limitations, as it focuses solely on the evaluative analysis of conventional instruments rather than formulating alternative instruments. On the other hand, a conceptual study by Khotimah et al. (2024) successfully outlines the effectiveness of the communicative approach, but it remains trapped at a normative-methodological level, leaving the specific parameters for developing its evaluation tools completely untouched. Meanwhile, research by Suharsono et al., (2023) attempting to introduce listening skills assessment innovation focuses more on technological mediums (such as Kahoot) and fails to provide a pedagogical framework for reconstructing the substance of the test items themselves. These limitations are also reflected in the broader field of language assessment. Furthermore, Park et al., (2022) argued that authentic, unscripted listening texts have rarely been adopted in listening tests because test developers tend to prioritize efficiency and standardized test specifications over authenticity, thereby limiting the representation of real-world communication contexts. Likewise, Lam (2021) argued that current language assessment practices rarely evaluate interactive listening, indicating that receptive listening skills remain underrepresented in authentic assessment.

In general, the fundamental difference between this research and the five aforementioned studies lies in their substantive focus and output. While previous studies have predominantly focused on criticizing existing conventional tools, shifting traditional tests to digital platforms, or

applying communicative approaches exclusively to speaking skills, this research fundamentally reconstructs the very substance of listening measurement tools. Assessment instruments play a crucial role as pedagogical gatekeepers, exerting a strong washback effect (Rahmawati et al., 2024). Without radical reform in evaluation tools, curriculum transformations including the progressive spirit of *Kurikulum Merdeka* and KMA 183 of 2019 will merely remain pedagogical slogans on paper without functional substance.

To break this epistemological deadlock, the primary objective of this research is to systematically design, develop, and validate a comprehensive Arabic listening skill assessment instrument based on a communicative approach, specifically tailored for seventh-grade students at MTsN 3 Bogor. Utilizing the ADDIE instructional design model specifically delimited to the Analyze, Design, and Develop stages this study aims to produce an innovative prototype consisting of 10 authentic assessment items. These items are operationalized through dynamic alternative evaluation techniques, including Task-Based Listening, Role Play, Oral Retelling, and Information Gap. Ultimately, this research aims not only to formulate an instrument that accurately measures students' sociolinguistic awareness and communicative problem-solving capacities in real-life contexts but also to provide a concrete, pragmatic pedagogical solution for educators. By doing so, this study aims to break the chain of students' habits of translating sentences into their mother tongue in their heads, proving that a well-developed, integrated oral-listening evaluation can effectively reduce internal barriers to language learning during this critical phase of academic transition.

## **METHOD**

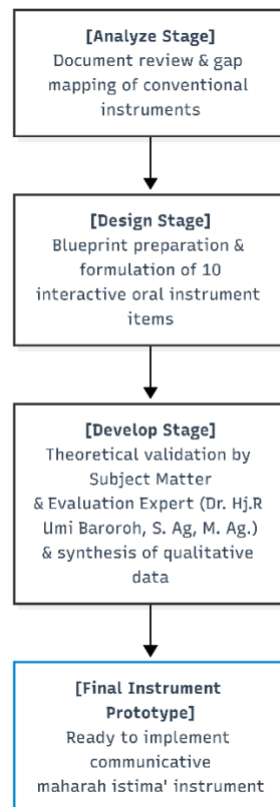
This study employs a Research and Development (R&D) methodology utilizing the ADDIE instructional design model. The implemented method is guided by the ADDIE instructional framework as proposed by Branch, (2009). In accordance with the primary objective of the study, which is to design and formulate a prototype of a theoretically valid communicative-based Arabic Listening Skill evaluation instrument, the scope of this research is specifically limited to the first three stages: Analyze, Design, and Develop. The instrument's development was deeply grounded in a comprehensive analysis of current learning needs, theoretical studies of Arabic language evaluation, of seventh-grade students at MTsN 3 Bogor. The development process was aligned with the learning objectives and essential materials contained in the Grade VII Arabic textbook used at

MTsN 3 Bogor. Subsequently, the drafted instrument underwent rigorous theoretical validation by language evaluation experts to ensure robust construct validity prior to any empirical implementation. Furthermore, the formulation of this design is also aligned with the applicable regulatory foundations, including the spirit of *Kurikulum Merdeka*, KMA 183 of 2019, and the Decree of the Director General of Islamic Education (SK Dirjen Pendis) Number 3302.

Data collection techniques in this study were conducted through two main methods: critical document review and expert validation sheets (expert judgment) (Setiawan et al., 2024). In the analysis stage, data collection focused on reviewing conventional evaluation instruments within the textbooks to map the gap between them and the need for an inclusive oral communication assessment. In the design stage, conceptual data were used to develop a blueprint for transforming the test into 10 interactive oral items, encompassing Oral Question and Answer, Role Play, Oral Retelling, Information Gap, and Task-Based Listening. Subsequently, during the development stage, data were obtained from comprehensive feedback from Arabic language evaluation experts who assessed the construct validity and the appropriateness of the language difficulty level relative to students' cognitive capacities. All collected data were analyzed using descriptive qualitative analysis. Data obtained from the critical document review and expert judgment were condensed, categorized, and synthesized to identify necessary revisions, which subsequently informed the refinement of the initial draft into a validated and implementation-ready prototype (Elangovan & Sundaravel, 2021).

More comprehensively, the research framework that visualizes the logical flow, procedures, and systematic stages of this assessment development is illustrated in the following concept map (diagram).

**Figure 1.** ADDIE Research Framework (Limited Stages)



Primary data were obtained through structured expert validation sheets (expert judgment), whereas secondary data were derived from the Grade VII Arabic textbook, curriculum documents, and relevant literature (Jamhuri et al., 2024; Sani et al., 2026). As the present study focused solely on theoretical validation without empirical field testing, the study participants consisted exclusively of expert validators selected through purposive sampling based on their academic qualifications, expertise in Arabic language assessment, and professional experience in language education (Campbell et al., 2020).

During the Develop stage, the initial draft of the evaluation instrument was submitted to an expert in Arabic language education and assessment for validation. The expert evaluated the instrument for content validity, construct validity, language clarity, and practical applicability. The feedback obtained was used to revise and improve the instrument before field implementation. The validator highlighted the need for a more precise alignment between the vocabulary used in the Arabic listening audio prompts and the cognitive characteristics of Grade VII students at MTsN 3 Bogor. The validator that while the communicative framework was robust, the operational instructions for the Role Play and Information Gap tasks were initially too complex, potentially leading to cognitive overload during the assessment process. Furthermore, the validator

recommended refining the analytical scoring rubrics to ensure they accurately captured genuine oral communicative competence rather than mere theoretical comprehension.

In response to these expert suggestions, several substantial revisions were made to systematically refine the initial draft before finalizing the prototype. The most significant adjustments were made to the Task-Based Listening and Oral Retelling items, in which complex syntactic structures in the prompts were replaced with high-frequency vocabulary strictly aligned with the Grade VII syllabus. Additionally, the task instructions for all 10 interactive items were simplified into clear, direct steps to make them more student-friendly. Finally, the assessment rubric was recalibrated to include highly specific, observable indicators for communicative fluency and interactive responsiveness, directly addressing the validator's critique. These targeted modifications ensured that the final prototype was not only theoretically sound but also practically viable for authentic classroom implementation.

## FINDINGS AND DISCUSSION

### Findings

Based on the Research and Development (R&D) methodology utilizing the ADDIE framework limited to the *Develop* stage, the results of the development of this assessment instrument are described in the following three systematic stages:

### Analyze Stage

Before developing the Arabic language assessment innovation, it is essential to critique the weaknesses of the conventional instruments currently in use (Safitri & Baihaqi, 2026). A review of various madrasah modules and textbooks, including the projected 2024 Seventh-Grade Arabic Textbook, reveals an outdated testing pattern. The current system is overly centered on rote memorization and rigid translation, and it separates the language from its social communication context (Hakim, 2022). Field data indicate that textbook assessments are dominated by lower-order cognitive domains, such as vocabulary comprehension and sentence structure accuracy. This discrete test approach, which breaks the language down into small units, has proven ineffective and fails to adequately measure students' integrative communication skills (Liao et al., 2023). Based on this analysis, an absolute shift in evaluation orientation is required. Conventional instruments that tend to focus on mastering linguistic elements separately must be transformed into innovative instruments oriented towards the ability to understand and use the language in real communication

contexts (Alyahya, 2024; Hasrol et al., 2022).

Language assessment is currently undergoing a paradigm shift from traditional methods toward a more communicative approach. The following table presents a comparative analysis between text-centered conventional assessment instruments and adaptive innovative instruments. This comparison highlights how each instrument measures language skills, ranging from listening comprehension to problem-solving abilities within authentic social contexts. A conceptual comparison of this shift is presented in Table 1.

**Table 1.** Comparison of Conventional and Innovative Assessment Instruments in Language Learning

| No | Conventional Assessment Instrument                        | Innovative Assessment Instrument                                                       | Difference                                                                                                                                                    |
|----|-----------------------------------------------------------|----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Answering questions based on written reading texts.       | Answering questions based on spoken dialogues.                                         | The conventional instrument focuses on reading and writing skills, whereas the new instrument measures the ability to comprehend spoken information directly. |
| 2  | Selecting the correct answer based on the text.           | Matching audio information with appropriate images.                                    | The new instrument connects language comprehension with visual representation, making it more contextual.                                                     |
| 3  | Memorizing or practicing dialogues according to examples. | Selecting or providing appropriate responses based on communication situations.        | The new instrument assesses the ability to use language in accordance with its social context and communicative function.                                     |
| 4  | Writing down the main idea or translating the text.       | Retelling the content of the dialogue orally (oral retelling).                         | The new instrument emphasizes the ability to comprehend and communicate the message effectively.                                                              |
| 5  | Completing sentences or answering written questions.      | Filling out a table based on information obtained from the dialogue (information gap). | The new instrument measures the ability to use listened information for problem-solving.                                                                      |

*Source: Adapted by the authors based on Lam, (2021)*

As illustrated in **Table 1**, conventional assessment instruments primarily emphasize passive reading and writing skills, such as translating texts and memorizing dialogues. Conversely, innovative assessment instruments are designed to directly measure communicative competence. These new instruments facilitate the functional use of language through applied tasks, such as responding to specific communicative situations, retelling content orally, and solving problems based on audio information (e.g., information-gap tasks). This indicates that innovative assessments are significantly more effective in evaluating students' language proficiency in a contextual and comprehensive manner.

### Design Stage

As a realization of the paradigm shift in language evaluation, the following table outlines a comprehensive design for an innovative assessment instrument. This framework deliberately breaks

the dependency on written text modalities, transitioning toward an authentic listening skill (Arabic Listening Skill) evaluation utilizing acoustic wave stimuli and audio dialogues (Alfinnas et al., 2026). The assessment indicators have been significantly reoriented; rather than measuring passive grammatical memory, the items are designed to evaluate functional targets such as general and specific information comprehension, the interpretation of communication functions, the delivery of appropriate social responses, and communicative problem-solving (Afriati et al., 2026).

**Table 2.** Design Specifications and Assessed Aspects of the Innovative Listening Assessment

| Instrument |                                                                 |                                                                                                                       |                              |                                               |
|------------|-----------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|------------------------------|-----------------------------------------------|
| No         | New Assessment Instrument (Innovation)                          | Instrument Design                                                                                                     | Assessment Technique         | Assessed Aspect                               |
| 1          | استمع إلى الحوار ثم أجب شفهيًا عن الفكرة العامة للحوار          | Students listen to a dialogue and then orally state the main topic they understood.                                   | Non-Test: Listen and Answer  | OralGeneral information comprehension         |
| 2          | استمع إلى الحوار ثم أجب عن سؤال يتعلق بمعلومة محددة في الحوار   | Students listen to a dialogue and provide an oral response to questions related to specific information.              | Test: Oral Listen-and-Answer | Listen-Specific information comprehension     |
| 3          | استمع إلى الحوار ثم حدد هدف المتكلم من الحوار                   | Students listen to a dialogue and explain the communication goal the speaker wants to achieve.                        | Test: Listening              | ObjectiveCommunication function comprehension |
| 4          | استمع إلى الحوار ثم حدد وظيفة الكلام في الحوار                  | Students identify the function of speech in the dialogue and connect it with the appropriate communication context.   | Test: Listening              | ObjectiveInterpretation of communication goal |
| 5          | استمع إلى الحوار ثم اختر الرد المناسب للموقف                    | Students select the most appropriate response to continue the conversation.                                           | Test: Selection Listening    | ResponseSocial accuracy response              |
| 6          | استمع إلى الحوار ثم قم بدور أحد المتحاورين وأجب شفهيًا          | Students play the role of one of the characters in the dialogue and provide an oral response suitable to the context. | Test: Role Play              | Role PlayInteractional competence             |
| 7          | استمع إلى الحوار ثم لخص الفكرة الرئيسة للحوار شفهيًا            | Students orally retell the main content of the dialogue using simple Arabic expressions.                              | Non-Test: Retelling          | OralGlobal comprehension and synthesis        |
| 8          | استمع إلى الحوار ثم أخبر زميلك بأهم المعلومات الواردة في الحوار | Students convey important information obtained from the                                                               | Non-Test: Retelling          | OralOral message reconstruction               |

|    |                                                                        |                                                                                                                                                        |  |  |  |
|----|------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|
|    |                                                                        | dialogue to a peer.                                                                                                                                    |  |  |  |
| 9  | استمع إلى الحوار ثم أكمل الجدول بالمعلومات                             | Students use theTest: InformationDirected detailed listened to complete theGap comprehension required data.                                            |  |  |  |
| 10 | استمع إلى الحوار ثم اتخذ قرارا أو حل مشكلة بناء على المعلومات المسموعة | Students utilize theTest: Task-BasedCommunicative listened information toListening problem-solving make a decision or provide a solution to a problem. |  |  |  |

Source: Adapted by the authors based on Brown & Abeywickrama, (2019); Lam, (2021)

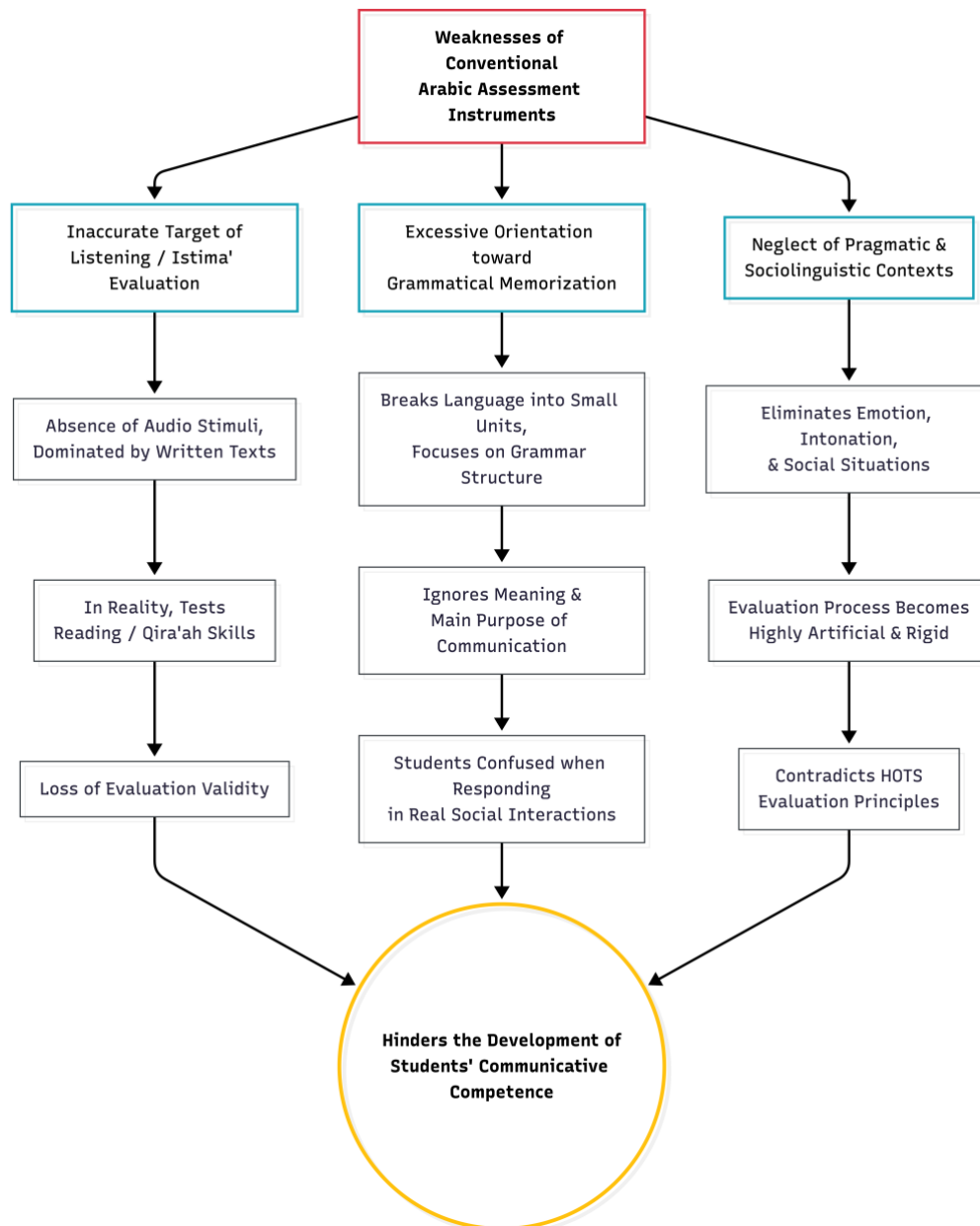
The detailed specifications in the table highlight a critical transformation designed to resolve the fundamental flaws of conventional Arabic assessment instruments. Traditional evaluations often fail to accurately assess listening skills by relying heavily on written texts, inadvertently testing reading, and excessively focusing on grammatical structures stripped of their communicative context. In contrast, this innovative design mandates the use of authentic audio stimuli to ensure high construct validity. By emphasizing meaning, functional interaction, and Higher Order Thinking Skills (HOTS), this transformative assessment model ensures that students are accurately evaluated on their ability to comprehend and actively solve communication problems in real-world social interactions.

Based on the analysis in the table above, the conventional Arabic assessment instrument evidently has fundamental flaws that hinder the development of students' communicative competence. The most striking weakness is the inaccuracy of the listening evaluation target Arabic Listening Skill. Most instruments used in madrasahs claim to test this skill, but they do not include any audio stimuli. The dominant use of written texts undermines this instrument's validity, because in practice the system tests reading skills rather than listening. This aligns with the findings of (Salsabila et al., 2026), which emphasize that the Arabic Listening Skill assessment must be based on authentic audio stimuli.

In addition to losing its evaluation target, this conventional assessment also shows an excessive orientation toward grammatical memorization. This approach breaks the language into small units, focusing solely on grammatical accuracy while ignoring the meaning and primary purpose of communication. Consequently, students become confused when required to respond to interlocutors in real social interactions. This condition directly contradicts the principles of Higher Order Thinking Skills (HOTS) evaluation, which requires students to be able to solve communication problems in accordance with everyday reality (Emira et al., 2025).

Therefore, a transformation toward an innovative instrument design grounded in a communicative approach is an academic imperative. This step is absolutely necessary to accurately and functionally train and measure students' active communication skills.

**Figure 2.** Innovation of Assessment Instruments Based on the Communicative Approach



Referring to the previous diagram highlighting the various weaknesses of the conventional evaluation system, innovating the Arabic Listening Skill assessment is imperative. This design shifts the evaluation paradigm from merely testing rigid grammatical memorization to testing the ability to comprehend functional meaning. Listening is no longer viewed as a passive process, but as an active cognitive activity where students analyze sounds while utilizing their social context knowledge to comprehend discourse.

To realize this design, the study proposes five variants of communicative test models:

**Figure 3.** Five Variants of Communicative Test Models for Assessing Listening Skills



1. **Comprehension of Dialogic Contexts** This test plays dialogues from familiar everyday situations. Students then answer questions that implicitly explore essential information (5W+1H). This model trains students' skills to guess meanings from context.
2. **Visual-Audio Cognitive Mapping** Students listen to purely oral descriptions and are asked to select the most accurate illustrative picture. This technique breaks students' habit of translating sentences into their mother tongue in their heads.
3. **Pragmatic Response Selection** In this test, students listen to a short utterance and select the most socially and culturally appropriate text response option. The assessment focuses on emotional intelligence and sociolinguistic awareness.
4. **Extraction of Macro-Ideas (HOTS-Based)** Students listen to a longer monologue/dialogue (1–2 minutes) and are asked to synthesize and write down its main idea using simple Arabic (Sugiara et al., 2025).
5. **Authentic Information Gap** Students are asked to listen to a dialogue while filling in a gap-filled table or form. This model demands high concentration as students must quickly and accurately filter specific data (Rahmadani et al., 2023).

### Develop Stage (Development and Validation)

In ensuring the feasibility of the instrument, this research took a distinct approach. While (Pranata, 2022) strongly emphasized mathematical statistical validity testing, this study focused on theoretical validation. The instrument's feasibility was evaluated comprehensively in terms of its alignment with language acquisition theories and prior research findings. This decision was taken to ensure that the suitability of the communicative construct truly became a solid main foundation before the instrument was empirically implemented.

### Discussion

The transformation of the instrument design from the conventional model to this innovative design has crucial pedagogical significance. First, this design fundamentally restores the construct validity of Arabic Listening Skill. Previously, the conventional instrument was proven to have a fundamental flaw due to the absence of audio stimuli, which practically resulted in the system only testing reading skills (Hamid, 2022; Rahmadani et al., 2023; Syafitri et al., 2026). Moreover, the old instrument broke the language into small units purely focused on memorizing theoretical rules, ignoring pragmatic contexts and emotional/intonation fluctuations. Consequently, students were proven confused when required to respond to interlocutors (Fitria et al., 2025).

The restoration of the instrument through pure acoustic wave-based testing is in line with the principles of Communicative Language Teaching (CLT) (Triani, 2026). Canale & Swain, (1980) emphasize that the main goal of language learning is to achieve communicative competence, which includes sociolinguistic competence. The proposed instruments, such as Pragmatic Response Selection and Authentic Information Gap, are specifically designed to facilitate these demands by bringing real-world interaction simulations into the classroom. Furthermore, the use of evaluation models based on Information Gaps and the extraction of macro-ideas forces students to engage in Higher Order Thinking Skills (HOTS). Listening is no longer seen as a passive process but as an active cognitive process of meaning-making. Bachman & Palmer, (1996) assert that ideal language evaluation must reflect authentic language use.

The design of these five innovative instrument variants simultaneously positions this research to firmly fill the gaps in various previous Systematic Literature Reviews (SLR). A study by (Subqi & Husen, 2024) has shown that innovative approaches increase student motivation, but these findings were limited to classroom learning strategies. This study goes beyond those boundaries by designing the physical form of its evaluation instruments. Additionally, studies by (Adhistia et al.,

2024) and (Wisman & Kurniawan, 2020), which proved the effectiveness of digital evaluation, have not touched the depth of material design. This study aims to reconstruct the substance of the test items communicatively, aligning with Momang's ( findings on the feasibility of authentic modules. When compared to (Pahlefi, 2022) research, which discusses instruments in general theoretical terms, this study possesses a comparative advantage by explicitly integrating the communicative approach as the main operational framework.

Thus, the communicative approach in developing this instrument functions not only as a measurement tool but also as a pedagogical stimulus. This design fosters students' independence in inferring meaning from context and has strategic implications for Arabic language teaching practices at the MTs level, particularly in proportionally accommodating Differentiated Learning within the *Kurikulum Merdeka*.

## CONCLUSION

This study successfully developed a prototype for a communicative approach-based Arabic Listening Skill assessment instrument for at MTsN 3 Bogor through the Analyze, Design, and Develop (ADDIE) stages. This development radically deconstructs the conventional evaluation paradigm that has thus far reduced listening skills to written memorization tests, producing 10 innovative instrument items through five variants of oral tests (Oral Listen-and-Answer, Visual-Audio Mapping, Role Play, Oral Retelling, and Information Gap). Through theoretical validation, this prototype has been conceptually proven capable of restoring the construct validity of Arabic Listening Skill. This design breaks the habit of literal translation by returning evaluation to meaning-based processing using authentic audio stimuli, while simultaneously encouraging Higher Order Thinking Skills (HOTS) that align with Communicative Language Teaching (CLT) principles and the *Kurikulum Merdeka*. Given that the methodological scope of this research is limited to the construct development and theoretical validation stages, its main limitation is the absence of field trials. Therefore, it is highly recommended for future researchers to continue this study to the Implement and Evaluate stages directly in the classroom. This subsequent step is essential for testing the level of empirical validity, statistical reliability, and the practicality of using this communicative instrument in the actual educational ecosystem of Madrasah Tsanawiyah.

## REFERENCES

- Adhistia, S. K., Hikmah, & Masrun. (2024). Analisis Instrumen Evaluasi Keterampilan Menyimak Bahasa Arab pada Platform Alef. *Mumtaza*, 4(1), 47–55.
- Afriati, I., Rambe, R. P., Izzah, N., & Amanda, R. (2026). Pengembangan Model Asesmen Kinerja Menyimak (Istima') Bahasa Arab Berbasis Teks Fungsional Kelas X SMA Babul Magfirah. *Shaut Al Arabiyyah*, 14(1), 36–60. <https://doi.org/10.24252/Saa.V14i1.64293>
- Alfinnas, H., Ali, M., & A.Syathor. (2026). Instrumen Penilaian Autentik pada Pembelajaran Bahasa Arab di Madrasah Aliyah. *Aphorisme*, 7(1), 193–210. <https://doi.org/10.37680/Aphorisme.V7i1.9131>
- Alyahya, R. S. W. (2024). The Development of a Novel, Standardized, Norm-Referenced Arabic Discourse Assessment Tool (Adat), Including an Examination of Psychometric Properties of Discourse Measures in Aphasia. *International Journal of Language & Communication Disorders*, 59(5), 2103–2117. <https://doi.org/10.1111/1460-6984.13083>
- Azhar, M., Wahyudi, H., & Promadi, P. (2022). Arabic Language Learning with Communicative Method and Factors Affecting Student's Speaking Ability. *Takuana: Jurnal Pendidikan, Sains, dan Humaniora*, 1(2), 92–101. <https://doi.org/10.56113/Takuana.V1i2.33>
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press.
- Branch. (2009). *Instructional Design: The Addie Approach*. <https://doi.org/10.1007/978-0-387-09506-6>
- Brown, H. D., & Abeywickrama, P. (2019). *Language Assessment: Principles and Classroom Practices* (3rd Ed.). Pearson.
- Campbell, S., Greenwood, M., Pior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive Sampling: Complex or Simple? Research Case Examples. *Journal of Research in Nursing*, 25(8). <https://doi.org/10.1177/1744987120927206>
- Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/Apllin/I.1.1>
- Ekawati, D., Hidayat, H. D., & Rosyada, D. (2024). Enhancing Vocabulary, Dialogue, and Writing Skills in Arabic Through Communicative Language Teaching: an Experimental Study. *International Journal of Religion*, 5(9), 896–907. <https://doi.org/10.61707/5880sz57>
- Elangovan, N., & Sundaravel, E. (2021). Method of Preparing a Document for Survey Instrument Validation By Experts. *Methodsx*, 8(101326). <https://doi.org/10.1016/J.Mex.2021.101326>
- Emira, A. E., Fatimah, S. N., Ridlo, U., & Raswan, R. (2025). Instrumen Evaluasi Maharah Kalam dalam Pembelajaran Bahasa Arab. *Al-Tadris: Jurnal Pendidikan Bahasa Arab*, 13(2), 615–641. <https://doi.org/10.21274/Tadris.2025.13.2.615-641>
- Fitria, H. S., Baroroh, R. U., Kamaliah, N., & Kamila, F. (2025). Innovation of Pragmatic Approach Through TestiIn Maharah Istima' In Arabic Language Book Class V Kemenag 2020. 8(4). <https://doi.org/10.31764/Pendekar.V8i4.34731>
- Hakim, I. (2022). Exploring the Challenges of Communicative Language Teaching in Indonesian Context from a Constructivist Perspective. *Els Journal on Interdisciplinary Studies in Humanities*, 5(1), 1–11. <https://doi.org/10.34050/Elsjish.V5i1.19163>
- Hamid, M. A. (2022). The Development of an Evaluation Instrument for the Implementation of the Arabic Language Curriculum in Islamic High School. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 14(1), 242–257. <https://doi.org/10.24042/Albayan.V14i1.10303>
- Hasrol, S. B., Zakaria, A., & Aryadoust, V. (2022). A Systematic Review of Authenticity in Second Language Assessment. *Research Methods in Applied Linguistics*, 1(3), 100023.

<https://doi.org/10.1016/J.Rmal.2022.100023>

- Husna, N., & Bakar, M. Y. A. (2025). Studi Kritis : Penilaian Pembelajaran Bahasa Arab dalam Maharoh Istima ' Berbasis Hots pada Buku Bahasa Arab Mi Kelas VI. *Khatulistiwa: Jurnal Pendidikan dan Sosial Humaniora*, 5(2), 442–461. <https://doi.org/10.55606/Khatulistiwa.V5i3.6966>
- Jamhuri, W., Widia, B., & Kasih, N. (2024). Development of Quizizz Application as an Instrument for Evaluation in Arabic Language Learning at Madrasah Aliyah. *Ij-atl (International Journal of Arabic Teaching and Learning)*, 08(02), 93–109. <https://doi.org/10.33650/Ijatl.V8i2.9696>
- Khotimah, M. H., Aqila, Indriati, S., & Sahkholid. (2024). Pendekatan Komunikatif dalam Pembelajaran Bahasa Arab: Maharoh Istima' dan Kalam. *Innovative: Journal of Social Science Research*, 4(6), 5051–5068. <https://doi.org/10.31004/Innovative.V4i6.17084>
- Lam, D. M. K. (2021). Don't Turn a Deaf Ear : A Case for Assessing Interactive Listening. *Applied Linguistics*, 42(4), 740–764. <https://doi.org/10.1093/Applin/Amaa064>
- Liao, L., Ye, S. X., & Yang, J. (2023). A Mini Review of Communicative Language Testing. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/Fpsyg.2023.1058411>
- Maryani, N., Syafei, I., & Kosim, A. (2024). Improving Arabic Speaking Proficiency (Muhādatsah) Using Interactive-Communicative Instruction. *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, 8(1), 18–33. <https://doi.org/10.15575/Jpba.V8i1.34894>
- Momang, H. D. (2021). Pengembangan Modul Buku Ajar Digital Keterampilan Menyimak Berdasarkan Pendekatan Autentik. *Kembara: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, 7(1), 71–93.
- Pahlefi, M. R. (2022). Kajian Teoritis Pengembangan Instrumen Penilaian Keterampilan Menyimak ( Mahārah Al - Istima ' ) dalam Pembelajaran Bahasa Arab. *Uktub : Journal of Arabic Studies*, 2(2), 68–84.
- Park, Y., Lee, S., & Shin, S.-Y. (2022). Developing a Local Academic English Listening Test Using Authentic Unscripted Audio-Visual Texts. *Language Testing*, 39(3). <https://doi.org/10.1177/02655322221076024>
- Pranata, A. (2022). *Pengembangan Instrumen Tes Keterampilan Menyimak dalam Pembelajaran Bahasa Arab bagi Peserta Didik*. 205–212.
- Putri, A. U., Abdurrahman, M., & Ali, M. (2026). Balancing Language Skills in Arabic Assessment: Challenges of Curriculum Implementation in an Islamic Integrated School. *Ijalt*, 1(1), 69–87. <https://doi.org/10.32332/Ijalt.V8i01.13326>
- Rahmadani, F. N., Ramadhan, M. A. F., Agustin, P. O., Albany, A. Z., Septyasari, I. S., Aziz, M. F. T., Muhammad, R., Fitriana, R. D., & Sholihatin, E. (2023). Analisis Sociolinguistik Ragam Bahasa dalam Komunikasi di Media Sosial oleh Generasi Milenial Mahasiswa UPN "Veteran" Jawa Timur. *Pendidikan Bahasa Indonesia dan Sastra (Pendistra)*, 6(2), 253–263. <https://doi.org/10.54367/Pendistra.V6i2.3227>
- Rahmawati, Zakiyah, Z., & Sitompul, J. (2024). Washback Effect Assessment of Mahārah Istimā' Using Google Forms. *Ta ' Lim Al - ' Arabiyyah : Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 8(1), 73–84. <https://doi.org/10.15575/Jpba.V8i1.34358>
- Rossi, O., & Brunfaut, T. (2021). Text Authenticity in Listening Assessment: Can Item Writers Be Trained to Produce Authentic-Sounding Texts? *Language Assessment Quarterly*, 18(4), 398–418. <https://doi.org/10.1080/15434303.2021.1895162>
- Safitri, Z., & Baihaqi, M. (2026). Quality Profile of Arabic Final Semester Assessment Items : A Psychometric Analysis. *Al-Lisan: Jurnal Bahasa*, 11(1), 87–102. <https://doi.org/10.30603/Al.V11i1.7322>
- Salsabila, S., Baroroh, R. U., Ibrahim, A. H. N., & Rahma, A. (2026). Inovasi Instrumen Penilaian

- Autentik Maharah Istima' dengan Pendekatan Pragmatik dalam Buku Bahasa Arab Kelas VIII. *Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab dan Kebahasaaraban*, 9(1), 318–327. <https://doi.org/10.35931/Am.V9i1.5720>
- Sani, N., Rahayu, D., Jh, V. A., & Rahman, Z. (2026). Innovating Assessment for Grade VII Arabic Curriculum: A Design – Development Study of Ministry Textbooks. *Lingua: Jurnal Keilmuan dan Kependidikan Bahasa Arab*, 12(1), 12–22. <https://doi.org/10.32678/Lingua.V12i1.12647>
- Setiawan, R., Wagiran, & Alsamiri, Y. (2024). Construction of an Instrument for Evaluating the Teaching Process in Higher Education: Content and Construct Validity. *Reid (Research and Evaluation in Education)*, 10(1), 50–63. <https://doi.org/10.21831/Reid.V10i1.63483>
- Subqi, A., & Husen, M. H. A. (2024). Penggunaan Pendekatan Inovatif dalam Evaluasi Maharatul Istima' di Madrasah Aliyah Al-Hamadiyah Cipancur Tasikmalaya. *Edupesantren*, 3(2), 48–64.
- Sugiara, L. R., Baroroh, R. U., Siregar, A. F., Muttakin, K., & Wahid, I. D. (2025). Musthofa Sha'ban's Perspective on Arabic Listening Assessment for Beginners. *Journal of Arabic Language Teaching*, 5(2), 249–262. <https://doi.org/10.35719/Arkhas.V5i2.2373>
- Suharsono, Shodiq, M. J., & Muhtarom, Y. (2023). Inovasi Penilaian Pembelajaran Bahasa Arab Berbasis Games Kahoot untuk Meningkatkan Keterampilan Menyimak di Ma'had Iit Rabbani Bengkulu. *Mahira: Journal of Arabic Studies and Teaching Student Research*.
- Syafitri, N. S., Fahma, N., & Maftuhah, L. H. (2026). Improving Listening Skills Through the Use of Mondly and Busuu Apps Subject: Comparative Study Using the Mondly and Busuu Apps. *Jurnal Pembelajaran Bahasa dan Sastra*, 5(1). <https://doi.org/10.55909/Jpbs.V5i1.1078>
- Triani. (2026). Sustainable Istima' (Listening) Learning: Learning Innovation Based on the Communicative Language Teaching (CLT) Approach. *Walada: Journal of Primary Education*, 5(1), 191–199. <https://doi.org/10.61798/Wjpe.V5i1.464>
- Wisman, & Kurniawan, R. (2020). Pengembangan Media Pembelajaran Keterampilan Menyimak Berbasis Online Menggunakan Google Form dan Google Classroom. *Silampari Bisa*, 3(2), 289–309. <https://doi.org/10.31540/Silamparibisa.V3i2>