



Design of MABBAR (Mari Belajar Bahasa Arab): A Web-Based Evaluation System for Enhancing Arabic Listening Competency (Mahārah al-Istimā') in Junior High Schools

Fatchiatuzahro ^{a,1,*}, Muhammad Ainul Haq ^{b,2}, Ubay Abdillah ^{c,3}, Friendis Syani Amrullah ^{d,4} Abdul Fatah ^{d,5}

^{a)} Institut Ummul Quro Al-Islami Bogor, ^{b)} Universitas KH. Abdullah Faqih Gresik, ^{c)} Intsitut Ummul Quro Al-Islami Bogor, ^{d,4} Universitas KH. Abdullah Faqih Gresik, ^{d5} Intsitut Ummul Quro Al-Islami Bogor,

¹ fatchiatuzahro@iuqibogor.ac.id, ² ainulfelays77@gmail.com, ³ ubayabdillah@iuqibogor.ac.id, ⁴ friendissyani@gmail.com, ⁴ snowfatah24@gmail.com

Abstract

Arabic listening instruction at the junior high school level is often constrained by limited classroom exposure, non-interactive audio media, and conventional evaluation practices that do not provide immediate feedback. This study developed and evaluated MABBAR (Mari Belajar Bahasa Arab), a web-based evaluation system designed to support mahārah al-istimā' through structured audio input, Text-to-Speech (TTS) support, automated scoring, and teacher-oriented learning analytics. The study employed a Research and Development (R&D) design by adapting the five-stage MANTAP model and aligning it with Borg and Gall's development procedures. The empirical trial involved 39 eighth-grade students at SMP IT Yasalam Bogor, while product feasibility was assessed by material, media, and phonetics experts. Data were collected through expert validation rubrics, pre-test and post-test instruments, student response questionnaires, and field observations. The validation results indicate that MABBAR is highly feasible, with material validation reaching 5.00/5.00, media validation 4.87/5.00, and phonetic validation 92%. Descriptive statistics show an increase in mean listening scores from 74.64 to 84.26, with a lower standard deviation after the intervention. The paired-sample t-test further indicates a significant difference between pre-test and post-test scores. Students also gave positive responses to the system, particularly regarding ease of use, content clarity, and learning motivation. These findings suggest that MABBAR can serve as a practical digital evaluation model for Arabic listening instruction, although broader trials and offline access are recommended for future development.

Keywords: Arabic Linguistics, Arabic Teaching, Mahārah al-Istimā', MABBAR (Mari Belajar Bahasa Arab), Web-based Evaluation.

مستخلص

يعاني تعليم مهارة الاستماع باللغة العربية في المرحلة الثانوية من محدودية الأدوات الرقمية المتخصصة التي توفر تعرضًا سمعيًا منظمًا. وقد أكدت الدراسات الأولية في مدرسة SMP IT Yasalam أن الطرق التقليدية باستخدام الوسائط الصوتية لا تحقق تفاعلاً فعالاً لدى الطلاب. تهدف هذه الدراسة إلى تصميم وتقييم فعالية نظام MABBAR، وهو نظام تقييم قائم على الويب يدمج تقنية تحويل النص إلى كلام (TTS) بصوت ناطق أصلي مع التغذية الراجعة الآلية لتوفير مدخلات سمعية ذاتية التنظيم. تعتمد الدراسة على منهج البحث والتطوير (R&D) من خلال إعادة صياغة المراحل المنهجية لبورغ وغال (١٩٨٣) في خمس مراحل مقترحة من قبل سومارني (٢٠١٩). وقد تطور المشروع بدءًا من تحليل الاحتياجات الأولية في SMP Yasalam وصولاً إلى نشر المنتج النهائي، مع التركيز على دمج تحليلات التعلم وتحسين التصميم المعماري وواجهة منصة MABBAR. شملت العينة ٣٩ طالبًا من الصف الثامن في SMP IT Yasalam – بوغور. جُمعت البيانات باستخدام أدوات متعددة: استمارات التحقق من الخبراء، اختبارات قبلية وبعديّة، واستبيان قابلية الاستخدام المكوّن من ٢٠ بندًا وفق مقياس ليكرت الخماسي. أظهرت النتائج أن منصة MABBAR ذات جدوى عالية؛ إذ حصلت على درجة ٥,٠٠/٥,٠٠ من خبراء المادة، و ٥,٠٠/٤,٨٧ من خبراء الوسائط، و ٩٢٪ من خبراء الصوتيات فيما يخص تقنية TTS. كما أظهرت التجربة على ٣٩ طالبًا تحسنًا ملحوظًا (قيمة $t = 6,842$ ، $(p > 0,001)$)، حيث ارتفع متوسط الدرجات من ٧٤,٦٤ إلى ٨٤,٢٦، وانخفض الانحراف المعياري من ٨,٧٥ إلى ٦,٤٩. وقَدِّم الطلاب استجابة إيجابية بنسبة ٩٦,٦٤٪، مما يحقق معايير نموذج قبول التكنولوجيا (TAM) من حيث سهولة الاستخدام ووضوح المحتوى. بشكل عام، نجحت منصة MABBAR في تحويل تقييم الاستماع باللغة العربية إلى منظومة تعلم تكيفية ذاتية التنظيم، رغم بعض القيود مثل الاعتماد على الإنترنت وانخفاض ثقة الطلاب بنسبة ١٣٪.

الكلمات المفتاحية: اللغويات العربية، تعليم العربية، مهارة الاستماع، نظام MABBAR، التقييم القائم على الويب.

INTRODUCTION

In the context of Arabic language pedagogy in Indonesia, the primary challenge in teaching *mahārah al-istimā'* (listening skills) is closely linked to the lack of a supportive linguistic environment (*bi'ah lugawiyah*). At the junior high school level, students are typically exposed to Arabic phonology for only 2 to 4 hours per week within the classroom setting. This restricted duration hinders the process of phonological and semantic internalization, as learners lack consistent auditory input. Recent research indicates that without media provides repetitive and independent audio simulations, students

tend to experience stagnation in their receptive skills (Husna et al., 2022).¹ Operationally, this stagnation is manifested in the students' inability to distinguish similar phonemes (*makhārij al-hurūf*) and a failure to grasp global meaning from complex oral texts. Therefore, the integration of web technology is no longer merely an optional aid but an urgent necessity to extend the boundaries of conventional classrooms into a ubiquitous digital space.

The transition from conventional evaluation systems to web-based platforms is fundamentally grounded in the theory of Self-Regulated Learning (SRL). Within a web-based architecture, learners exercise full control over their learning pace, a concept known as self-paced learning, which allows for the repetitive access of audio materials until comprehensive understanding is achieved.² To facilitate this autonomy, the MABBAR system incorporates specialized features, including an adaptive audio question bank and an automated feedback interface. By providing these functionalities, the system serves as both an assessment tool and a medium for cognitive reinforcement, fostering student independence and learning motivation.

Technically, the selection of the Vue.js and Node.js frameworks in the development of MABBAR provides advantages in responsiveness and real-time data processing. Amidst the diversity of hardware owned by students-ranging from laboratory computers to personal smartphones a web-based system ensures inclusivity of access without requiring application installations that burden device memory. This technical efficiency is crucial in

¹ M. A. Husna et al., "Implementasi materi mahirah al-kalam dalam transformasi kurikulum bahasa Arab perspektif ACTFL pada pondok pesantren salaf," *Arabi: Journal of Arabic Studies* 7, no. 1 (2022): 38–50

² Elmabaredy, A., & Gencel, N. (2024). Exploring the integration of self-regulated learning into digital platforms to improve students' achievement and performance. *Discover Education*, 3(1), Article 262.

guaranteeing the continuity of the evaluation process, as technical obstacles often encountered in heavy applications are frequently a primary factor in decreasing student learning motivation.³ With a lightweight architecture, MABBAR is capable of delivering high-quality streaming audio without significant lag.

Furthermore, the developed evaluation system integrates simple gamification principles into the presentation of quizzes and feedback. Based on Self-Determination Theory (SDT), the provision of instant feedback in the form of scores and corrections immediately after completing tasks can trigger dopamine release, which enhances students' emotional engagement in learning.⁴ In the context of Arabic language learning, which is often perceived as 'rigid' and 'difficult' by junior high school students, the interactive approach through MABBAR is expected to transform such negative perceptions into an enjoyable and challenging learning experience. Evaluation is no longer viewed as a daunting burden, but rather as a challenge to achieve a higher level of proficiency.

Preliminary observations at SMP IT Yasalam revealed a critical gap in Arabic pedagogy: while students demonstrated adequate competence in reading, their listening scores consistently underperformed. The urgency of this research is underscored by documented evidence from the institution, which indicates that the

³ Mikita Piastou, "Comprehensive Performance and Scalability Assessment of Front-End Frameworks: React, Angular, and Vue.js," *World Journal of Advanced Engineering Technology and Sciences* 9, no. 2 (2023): 366–376

⁴ The analysis of the theoretical foundations of the interactive elements within MABBAR refers to the principles of Game-Based Learning, which aim to enhance both cognitive and emotional user engagement. An in-depth review regarding the integration of such theories into digital platforms can be found in Jeanine Krath, Linda Schürmann, and Harald F. O. von Korflesch, 'Revealing the Theoretical Basis of Gamification: A Systematic Review and Analysis of Theory in Research on Gamification, Serious Games and Game-Based Learning,' *Computers in Human Behavior* 125 (2021): 106963

average pre-test score for Arabic listening was only 74.64, with the lowest score reaching 60.⁵ This performance falls below the expected competency standards, confirming a significant pedagogical gap. This stagnation is primarily driven by the “echo-chamber” effect of conventional classroom settings, where students lack access to standardized phonetic models and experience high cognitive anxiety during one-off audio playbacks. These identified pedagogical barriers necessitate a shift from teacher-centered auditory drills to a more autonomous framework.

Therefore, the design of MABBAR (Mari Belajar Bahasa Arab) at SMP IT Yasalam represents a strategic step toward modernizing the local Arabic language curriculum and addressing these specific auditory learning deficits. By utilizing a systematic Research and Development (R&D) methodology, this study goes beyond the conceptual stage to include scientific validation of both material and media feasibility. The results are expected to provide practical contributions for Arabic language practitioners in creating adaptive evaluation systems, while offering theoretical insights to the CALL (Computer-Assisted Language Learning) literature, specifically concerning the optimization of listening skills through self-regulated digital platforms at the secondary education level.

The digital transformation in language learning over the last decade has driven significant changes in the way learners interact with materials, teachers, and evaluation systems. Educational technology has not only expanded access but also increased the effectiveness of the learning process through interactivity, personalization, and assessment automation.⁶ In the context of language learning, the integration of

⁵ Primary Research Data, “Results of Descriptive Statistical Analysis of Pre-test and Post-test Scores for Mahārah al-Istimā’ among Grade VIII Students of SMP IT Yasalam,” (Processed by researcher, 2025).

⁶ A systematic review of how technology empowers the teaching and learning process in language education, along with applicative innovations relevant to

technology has proven to strengthen the practice of Technology-Enhanced Language Learning (TELL), particularly through the use of web-based platforms that provide rapid feedback and accurate learning analytics to support pedagogical decision-making.⁷

At the Junior High Schools, the need for adaptive, authentic, and data-driven learning evaluation has become increasingly relevant, aligning with national education policies that promote formative assessment and the integration of digital literacy into the curriculum. Digital-generation learners demonstrate a strong preference for evaluation media that are visual, interactive, and gamified; thus, web-based assessment systems have emerged as an approach capable of enhancing participation, motivation, and learning retention.⁸

Nevertheless, the implementation of Arabic language learning evaluation in various junior high school institutions is still dominated by conventional methods in the form of written tests, which lack the capacity to facilitate integrated language skills and fail to provide systematic developmental analysis. Conventional evaluation generally cannot provide question variations, automated grading, or learning analytics-based reporting, despite the increasing demand for these features in the digital learning era.⁹

the needs of contemporary learners, is discussed in depth by Nibedita Dash and Lalita Mohan Mohapatra in 'Technology Empowered Teaching and Learning in Language Education: A Systematic Review,' in *Technology Driven Language Learning: Innovations and Applications* (2025), 127–145.

⁷ Lee, J. (2022). *Web-based assessment tools for language learning: Pedagogical implications and learner engagement*. 35(4), 678–695.

⁸ Rahman, N. F., Masdar, M. I., Ilmiani, A. M., Habibie, F., & Samdouni, S. (2022). Transformation of Arabic assessment in Indonesia: Conventional assessment toward digital assessment. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya*, 10(2), 199–212.

⁹ Aladini, A., Mahmud, R., & Ali, A. A. H. (2024). The importance of needs satisfaction, teacher support, and L2 learning experience in Intelligent Computer-Assisted Language Assessment (ICALA): A probe into the state of willingness to communicate as well as academic motivation in EFL settings. *Language Testing in Asia*, 14(1), 58

Recent studies demonstrate that digital evaluation media significantly enhance the quality of language assessment through integrated pedagogical features. For instance, Fatchiatuzahro highlight that digital platforms such as Wordwall improve students' *mahārah kitābah* (writing skills) by utilizing interactive question formats and automated feedback. Although focused on writing, the success of such digital interventions underscores a critical pedagogical principle: that real-time reinforcement and adaptive engagement are essential for language acquisition. By extension, incorporating these same technological affordances into *istimā'* (listening) evaluations as seen in the MABBAR system can address the limitations of conventional auditory assessments. This correlation suggests that web-based media can transform evaluation at the junior high school level into a more effective, meaningful, and growth-oriented process across all linguistic competencies.¹⁰

Complementing this trajectory, Hermiwati et al. demonstrated that the integration of digital tools is equally vital for foundational linguistic elements; their development of a vocabulary learning application using the ADDIE model on the Learning Apps platform proved highly feasible (82.1%) and remarkably successful in elevating seventh-grade students' motivation and vocabulary mastery.¹¹

However, to date, the development of web-based evaluation systems specifically designed for *mahārat al-istimā'* (listening skills) in Arabic language learning at the Junior High Schools remains highly limited.¹² Most general platforms, such as Quizizz, Google Forms,

¹⁰ Fatchiatuzahro, Fahmi, N., Latifah, M., & Nur Azizah. (2025). Pemanfaatan Wordwall Sebagai Media Evaluasi Pembelajaran Maharah Kitabah Pada Mahasiswa PBA IUQI Bogor. *Mauriduna: Journal of Islamic Studies*, 6(1), 138–155.

¹¹ Hermiwati, H., Hasyib, M. T. J., Machmudah, U., & Wicaksono, M. A. (2024). Enhancing Arabic Vocabulary Acquisition: Development of Learning Apps Based Interactive Learning Media. *Lisanudhad*, 12(2), 30.

¹² In the global context of Arabic education, platforms like the Aljazeera Arabic learning

or school LMS, do not provide specialized Arabic audio processing features, item difficulty analysis for listening tasks, or reporting of listening results that align with the specific characteristics of the Arabic language. This condition indicates a significant space for innovation to introduce digital evaluation tools that are in harmony with the needs of the Arabic curriculum, technological developments, and the characteristics of junior high school students.

To address these needs, this study developed MABBAR, a web-based Arabic language evaluation system specifically engineered for *mahārah al-istimā'*. The primary technical distinction of MABBAR lies in its integration of high-fidelity native speaker audio libraries with a real-time diagnostic feedback loop, features that are absent in conventional platforms. The system incorporates an adaptive audio question bank, automated grading, and learning analytics to monitor student progress. Furthermore, it is structured around standardized listening indicators, delivered through an interactive interface designed to foster cognitive independence and motivation among junior high school students.

Substantively, the development of the MABBAR system is expected to enhance the objectivity and efficiency of evaluating *istimā'* skills, expand access to data-driven formative assessments, support teachers in mapping students' strengths and weaknesses, and present a digital evaluation model that can be replicated across

website provide high-quality, pre-determined resources. However, such platforms offer static content designed for general independent learners and do not allow instructors to customize materials. To address this limitation, MABBAR was developed as an adaptable, quiz-based learning platform similar to Quizizz where instructors have the flexibility to design and upload their own learning materials. While MABBAR is versatile enough to support various educational levels, this specific study utilizes it to provide repetitive and independent audio materials tailored for junior high school students. Furthermore, MABBAR integrates learning analytics based on student performance, directly supporting teachers' pedagogical decision-making in managing junior high school classrooms.

various schools. Thus, this research serves as a strategic effort to strengthen the quality of Arabic language learning evaluation, while simultaneously contributing to the enhancement of digital literacy and 21st-century competencies within the secondary education environment.

METHOD

This study adopts an integrated Research and Development (R&D) framework that synthesizes the ten systematic stages of Borg and Gall (1983)¹³ with the Model Penelitian dan Pengembangan Lima Tahap (MANTAP) developed by Sumarni (2019).¹⁴ By mapping Borg and Gall's comprehensive validation steps into the five macro-stages of the MANTAP model, this study ensures methodological rigor while maintaining operational clarity. The integrated R&D phases are structured and executed as follows:

Table 1. Integrated R&D Procedure Used in the Development of MABBAR

MANTAP Stage	Borg & Gall Alignment	Main Activities	Main Output
Stage I: <i>Preliminary Research</i>	<i>Research and information collecting</i>	<i>Needs analysis, problem identification, and analysis of listening evaluation practices at SMP IT Yasalam Bogor.</i>	<i>Problem map and user needs.</i>
Stage II: <i>Model Development</i>	<i>Planning and preliminary product development</i>	<i>Design of system architecture, content structure, audio evaluation flow, teacher dashboard, and student interface.</i>	<i>Initial MABBAR prototype.</i>

¹³ Borg, W. R., & Gall, M. D. (1983). Educational research: An introduction. *Longman*, (4).P. 775

¹⁴ Sumarni, S. (2019). *Model Penelitian dan Pengembangan Lima Tahap (MANTAP)*. UIN Sunan Kalijaga. P. 19-21

MANTAP Stage	Borg & Gall Alignment	Main Activities	Main Output
Stage III: Model Validation	Preliminary field testing and main product revision	Validation by material, media, and phonetics experts; revision based on qualitative feedback.	Validated prototype.
Stage IV: Model Testing	Main field testing, operational revision, and final revision	Student trial, usability response collection, pre test and post-test administration, and final technical refinement.	Finalized MABBAR platform.
Stage V: Dissemination	Dissemination and implementation	Implementation in the target school and academic dissemination through seminars and publication.	Disseminated product and research report.

Research Subjects and Setting

The empirical phase of this study was conducted at SMP Yasalam Bogor. To ensure comprehensive validation and testing of the MABBAR platform, three distinct groups of participants were strategically selected. First, multidisciplinary experts were engaged as validators, encompassing specialists in Arabic language pedagogy, digital learning media, and phonetics—the latter specifically tasked with evaluating the quality of the Text-to-Speech (TTS) audio components. Second, instructional viability was assessed by practitioners, consisting of active secondary-level Arabic language teachers. Finally, the target user trials involved 39 eighth-grade students selected via purposive sampling. This sampling strategy ensured that the student cohort possessed the specific intermediate language competency necessary to rigorously interact with and evaluate the system’s features.

Data Collection and Analysis Technique

Data were collected using expert validation rubrics, pre-test and post-test instruments for Arabic listening competence, a 20-item student usability questionnaire using a five-point Likert scale, and field observation notes. The validation rubrics assessed content relevance, curriculum alignment, interface design, navigation, readability, responsiveness, TTS clarity, pronunciation accuracy, and intonation consistency. The student questionnaire covered ease of use, content clarity, learning motivation, learning independence, and device compatibility.

Qualitative feedback from experts, teachers, and students was analyzed descriptively through data condensation, data display, and conclusion drawing, following Miles, Huberman (Matthew B et al., 2018). Quantitative validation data were converted into percentages and interpreted using feasibility categories. Student learning outcomes were analyzed using descriptive statistics and a paired-sample t-test to examine the difference between pre-test and post-test scores. The normality of score differences should be verified before final submission to ensure the appropriateness of the paired-sample t-test.

RESULT AND DISCUSSION

Product Development and Main Features of MABBAR

The result of this development study is a web-based Arabic language learning and assessment system named MABBAR that can be accessed at <https://mabbar.my.id/> for students and <https://mabbar.my.id/guru> for teachers. The system was designed to support the management of Arabic learning exercises, student data administration, and interactive evaluation, particularly for Mahārat al-Istimā' or Arabic listening skills. As an R&D product, MABBAR is not only presented as a digital platform, but also as an instructional

support system that integrates administrative functions and pedagogical activities in one interface. The product visualization shows that the system attempts to simplify the work of instructors in preparing learning materials and exercises while providing students with a more structured and interactive learning experience.

The first major component of MABBAR is the exercise management feature. As shown in Figure 1, the exercise management interface functions as the central page for instructors to organize and control the question bank. Through this page, instructors are able to view available exercises, identify exercise categories, and manage learning tasks according to the intended skill focus. The use of a card-based display allows each exercise category to be visually separated and easily recognized. This type of interface is important because the instructor does not need to search through a long list of unstructured items. Instead, the system presents the exercises in a more organized layout, making the process of locating, reviewing, and updating exercise materials more efficient¹⁵

In addition to displaying the exercise list, the system also provides operational buttons such as “Edit,” “Delete,” and “Add.” These functions indicate that MABBAR was developed not merely as a static repository of exercises, but as a dynamic content management system. These features are essential in digital learning development because Arabic learning materials, especially assessment items, require regular updating to maintain accuracy, relevance, and suitability to students’ proficiency levels.

¹⁵ Krug, S. (2013). Don’t Make Me Think Revisited: A Common Sense Approach.

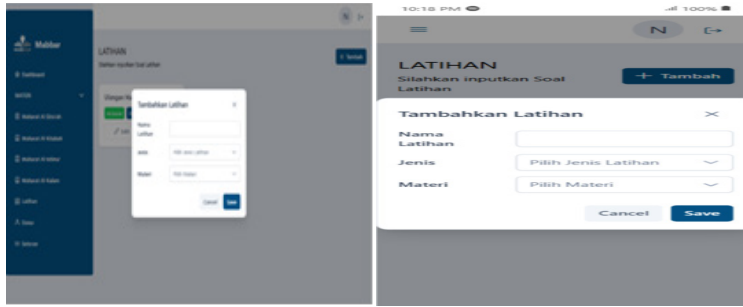


Figure 1. The exercise management interface for instructor

Figure 1 also shows that the exercise input process is supported by a popup modal. This feature allows instructors to enter or modify exercise data without leaving the main page. From a usability perspective, this design reduces unnecessary navigation and helps users remain focused on the task they are completing. Krug (2013) emphasizes that an effective interface should reduce users' cognitive burden by making actions clear and direct¹⁶. In the context of MABBAR, the popup modal supports this principle because the instructor can manage exercise data through a shorter workflow. The interface therefore contributes to administrative efficiency, especially when instructors need to prepare many exercise items within a limited time.¹⁷

The second component of MABBAR is the student management feature. As shown in Figure 2, this page is designed to help instructors or administrators manage student data in a structured format. The student management page displays student records in a systematic layout and includes pagination to divide the data into manageable sections. This feature is important because a digital learning and assessment system usually involves a large number of student users.

¹⁶ Krug, S. (2013). Don't Make Me Think Revisited: A Common Sense Approach.

¹⁷ Haagsman, M. E., Scager, K., Boonstra, J., & Koster, M. C. (2020). Pop-up questions within educational videos: Effects on students' learning. *Journal of Science Education and Technology*, 29(6), 713–724.

Without pagination, student data may appear too long, difficult to browse, and less efficient to manage. By dividing the data into pages, MABBAR allows administrators to access student information more easily and maintain better control over the database.

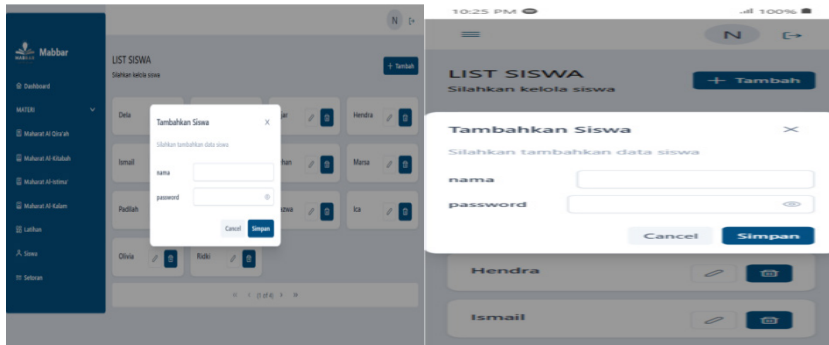


Figure 2. The student management interface

The student management interface also reflects the administrative function of MABBAR as a learning management support system. While the exercise management page focuses on instructional content, the student management page focuses on user data. These two components complement each other because effective digital assessment requires both well-prepared learning materials and well-organized learner records. Through this feature, MABBAR can support the practical needs of institutions or instructors who need to monitor student participation, organize accounts, and manage user access within the system. Therefore, the product is not limited to presenting learning content, but also provides the infrastructure needed for managing the learning process.

The most pedagogically significant component of MABBAR is the Mahārat al-Istimā' evaluation interface. As shown in Figure 3, this interface is structured into three main learning segments: preparation, listening activity, and evaluation. The first segment provides vocabulary support or *mufradāt*. This section introduces key words before students engage with the listening material. The

inclusion of vocabulary support is important because listening comprehension in Arabic is often influenced by students' familiarity with lexical items, pronunciation, and meaning. By presenting vocabulary before the listening activity, the system gives students initial linguistic support and prepares them to understand the audio input more effectively.

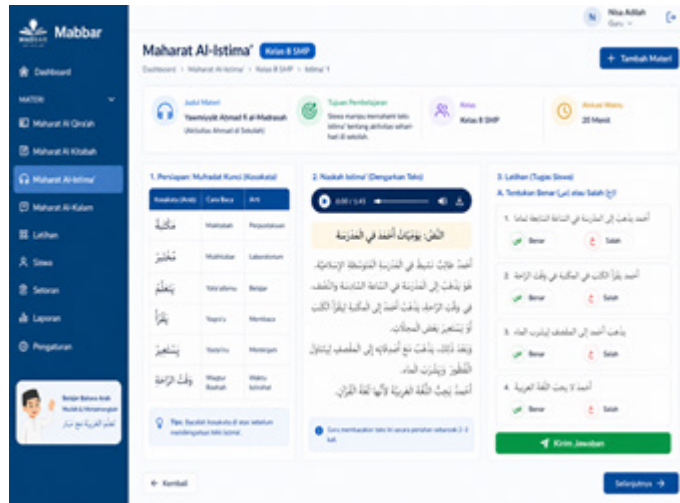


Figure 3. The Mahārat al-Istimā' evaluation interface

The second segment in Figure 3 is the active listening area. This part contains an audio player and script visualization. The integration of audio and text provides students with multimodal input, allowing them to listen to the Arabic material while also observing the written form of the language. This design is consistent with the principle of multimedia learning, which suggests that learning can be supported when verbal and visual information are presented in a coordinated way¹⁸. In MABBAR, the audio feature trains students' listening ability, while the script visualization helps them notice pronunciation, word forms, and sentence structure. This

¹⁸ Mayer, R. E. (2021). *Multimedia Learning* (3rd ed.). Cambridge University Press, pp. 67–85. Sweller, J. (2019). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 67(1), 1–16, pp. 5–12.

combination is particularly useful in Arabic listening instruction because learners often need support in connecting spoken forms with written Arabic script.

The third segment of the Mahārat al-Istimā' interface is the evaluation section. As shown in Figure 3, the evaluation is presented through interactive True/False items or Ṣaḥ/Khaṭa' exercises. After listening to the audio and reviewing the available support, students answer comprehension questions directly in the system and submit their responses through the provided button. This design transforms the listening activity from a passive audio exposure into an interactive assessment process. Students are not only asked to listen, but also to demonstrate comprehension through immediate task completion. The presence of a submission feature also indicates that the system is prepared to record or process student responses digitally.

Overall, the product visualization shows that MABBAR combines three important dimensions of a digital Arabic learning system: content management, learner management, and skill-based evaluation. The exercise management page supports instructors in preparing and revising learning materials. The student management page supports the organization of learner data. The Mahārat al-Istimā' interface supports students through a sequence of vocabulary preparation, audio-based learning, and interactive evaluation. This structure indicates that MABBAR was developed not only to digitize Arabic exercises, but also to organize the teaching, learning, and assessment process in a more systematic way.

The final prototype of MABBAR therefore represents a functional product for Arabic language learning and assessment. Its interface design emphasizes simplicity, structured navigation, and task-oriented features. Its listening evaluation module reflects a pedagogical sequence that begins with preparation, continues with multimodal listening exposure, and ends with comprehension

assessment. In line with cognitive load theory, such a structured interface may help reduce unnecessary processing demands by guiding users through clear stages of activity¹⁹. Thus, MABBAR can be described as a digital learning product that integrates usability, administrative efficiency, and pedagogical support for Arabic listening instruction.

Expert Validation and Product Feasibility

Expert validation was conducted to ensure that MABBAR met content, media, and phonetic standards. The results show that the system is highly feasible for use in Arabic listening evaluation. Material experts confirmed that the learning content was aligned with the junior high school Arabic curriculum, while media experts highlighted the clarity of navigation, visual consistency, and cross-device responsiveness. The phonetics expert confirmed that the TTS output was sufficiently clear and consistent for listening evaluation. The following table summarizes this result:

Table 2. Summary of Expert Validation Results

Validator	Experts	Aspects Assessed	Score	Category	Revision Made
Material experts	3	Curriculum alignment; vocabulary and grammar accuracy; contextual relevance.	5.00/5.00	Highly feasible	Contextual examples were added.
Media experts	3	Navigation, UI design, readability, responsiveness, and interface consistency.	4.87/5.00	Highly feasible	Color contrast and help icon were improved.

¹⁹ Mayer, R. E. (2021). Multimedia Learning (3rd ed.). Cambridge University Press, pp. 67–85. Sweller, J. (2019). Cognitive load theory and educational technology. Educational Technology Research and Development, 67(1), 1–16, pp. 5–12.

Validator	Experts	Aspects Assessed	Score	Category	Revision Made
Phonetics expert	1	TTS clarity, pronunciation, intonation, and listening-task suitability.	92%	Highly feasible	Audio quality was confirmed; presentation flow was refined.

The high feasibility scores indicate that MABBAR is appropriate not only as a digital test platform but also as a structured learning support tool. The material validation confirms content validity, while the media and phonetic validation strengthen the technical and auditory reliability of the platform.

Student Perception and Usability

Student perception was measured after the field trial using a usability questionnaire. The results indicate positive responses across all measured indicators, especially content clarity and ease of use.

Table 3. Student Perception and Usability Responses (N = 39)

Indicator	Agree/Strongly Agree (n)	Percentage	Category
Ease of use (navigation and interface)	37	%94.8	Very high
Clarity of learning content and quizzes	38	%97.4	Very high
Improvement in learning motivation	35	%89.7	High
Independence in learning	32	%82.1	High

The usability data suggest that students were able to operate the system with minimal difficulty and perceived the learning content as clear. The relatively lower score for learning independence indicates that some students still require teacher guidance when transitioning from conventional listening activities to autonomous digital practice. This finding is important because digital platforms do not automatically produce independent learning; they must be accompanied by scaffolding, orientation, and continuous classroom support.

Effectiveness of MABBAR on Improving Listening Competence

The effectiveness of MABBAR was examined by comparing students’ pre-test and post-test scores. The descriptive statistics are presented in Table 4.

Table 4. Descriptive Statistics of Pre-test and Post-test Scores

Statistical Description	Pre-test	Post-test
Number of participants (N)	39	39
Mean score	74.64	84.26
Standard deviation (SD)	8.75	6.49
Median	75.00	81.00
Minimum score	60	75
Maximum score	89	95

Table 4 shows that the mean score increased from 74.64 to 84.26 after students used MABBAR, indicating an average gain of 9.62 points. The standard deviation decreased from 8.75 to 6.49, suggesting that student achievement became more consistent after the intervention. The improvement in the minimum score from 60 to 75 also indicates that lower-performing students benefited from the structured audio input and immediate feedback provided by the system.

The paired-sample t-test, as shown in figure 4, indicates a statistically significant difference between pre-test and post-test results. This finding supports the conclusion that MABBAR contributed positively to students’ Arabic listening performance. Pedagogically, the improvement can be attributed to three integrated features: repeated access to audio input, structured vocabulary scaffolding before listening, and immediate feedback after evaluation. These features help students monitor their performance and reduce uncertainty during listening tasks.

Paired Samples Test									
Paired Differences									
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Pretest - Posttest	-9.237	3.781	.613	-10.480	-7.994	-15.061	37	<.001

Figure 4. Paired Sample T-Test Results

Based on the table above, the results of the t-test yield a significance value (p-value) of 0.001, which is lower than the alpha level of 0.05 (alpha = 0.05). The calculation also shows a t-statistic (t count) of 6.842, which is significantly higher than the t-table (t table) for df = 38. Consequently, these results indicate that the *istimā'* evaluation system within the MABBAR application significantly influences the difference between pre-test results (evaluation conducted without MABBAR) and post-test results (evaluation using MABBAR). Therefore, it can be concluded that MABBAR is an effective *mahārah al-istimā'* evaluation system for junior high school students (MTs/SMP).

Consequently, this indicates that the *istimā'* evaluation system within the MABBAR application significantly influences the difference between pre-test results and post-test results. Therefore, it can be concluded that MABBAR is an effective *mahārat al-istimā'* evaluation system for junior high school students (MTs/SMP). The system's effectiveness is further underlined by a balanced analysis of its functional attributes, which were identified through direct user feedback and technical performance during the field trial.

Pedagogical and Technical Interpretation

The pedagogical significance of MABBAR lies primarily in its capacity to transform Arabic listening assessment from a terminal, audio-only activity into a guided learning process. Its strongest contribution is not simply the digitization of exercises, but the

arrangement of learning stages that connect vocabulary preparation, controlled audio exposure, and interactive comprehension checking. In this sense, the product positions *mahārah al-istimā'* evaluation as a formative activity. Students are first supported through key vocabulary, then allowed to control the listening input, and finally asked to demonstrate comprehension through digital tasks. This sequence is pedagogically meaningful because it reduces the abruptness of conventional listening tests and provides temporary support before learners are expected to perform independently.²⁰

Another strength of MABBAR is the integration of technical responsiveness with pedagogical usability. A listening platform cannot be separated from the quality of its media delivery because audio delay, interruption, or interface lag may distract learners and weaken comprehension. The use of modern web technology in MABBAR is therefore pedagogically consequential: a responsive system supports the continuity of attention, while the structured interface helps learners focus on the task rather than on navigation problems. This point is consistent with research emphasizing the importance of low-latency media delivery and the negative effect of auditory distraction on comprehension-related performance. Thus, technical stability should be read as part of the learning design, not merely as a programming achievement.²¹

The use of native-speaker Text-to-Speech also strengthens the pedagogical value of the product. The phonetic validation result should not be repeated merely as a numerical score; rather,

²⁰Husna, M. A., Inayah, I., Mubarak, F., Taufiqurrahman, T., & Qomariyah, L. (2022). Implementasi materi mahirah al-kalim dalam transformasi kurikulum bahasa Arab perspektif ACTFL pada pondok pesantren salaf. *Arabi: Journal of Arabic Studies*, 7(1), 38–50.

²¹ Bentaleb, A., Lim, M., Akcay, M. N., Begen, A. C., Hammoudi, S., & Zimmermann, R. (2025). Toward one-second latency: Evolution of live media streaming. *IEEE Communications Surveys & Tutorials*, 28, 2418–2456.

it indicates that the listening input provided by the system is sufficiently standardized for instructional and evaluative purposes. A stable auditory model is important in Arabic learning because inconsistent pronunciation may lead students to misidentify sounds, words, or sentence boundaries. Recent work on speech synthesis suggests that synthetic voices have become increasingly capable of providing clear and pedagogically usable audio models, while studies on technology-mediated language learning show that digital environments can expand students' access to language input beyond the classroom. For this reason, MABBAR may help create a more consistent listening environment, although teacher guidance remains necessary to contextualize pronunciation, meaning, and communicative use.²²

MABBAR also has implications for teacher decision-making. The system's automatic scoring and learning analytics can help teachers identify patterns of student difficulty more efficiently than manual checking alone. In a conventional listening test, the teacher often receives only the final score, whereas a digital system can potentially show which items, sounds, vocabulary areas, or comprehension types are most problematic. This supports a diagnostic orientation to teaching: teachers can use student-response data to revise materials, provide remedial practice, or design targeted follow-up instruction. Such a function is aligned with research on teachers' use of data from digital learning platforms, which shows that platform-generated data can inform instructional design when teachers are able to interpret it meaningfully.²³

²² Perrotin, O., Stephenson, B., Gerber, S., Bailly, G., & King, S. (2025). Refining the evaluation of speech synthesis: A summary of the Blizzard Challenge 2023. *Computer Speech & Language*, 90, Article 101747.

²³ Hase, A., & Kuhl, P. (2024). Teachers' use of data from digital learning platforms for instructional design: A systematic review. *Educational Technology Research and Development*, 72(4), 1925–1945.

From the perspective of technology acceptance, the positive response to MABBAR's interface and device compatibility implies that the product has met two important conditions for classroom adoption: perceived usefulness and perceived ease of use. When students and teachers can access the system through laptops, tablets, or smartphones without excessive technical difficulty, the probability of sustained use becomes stronger. This finding is consistent with the Technology Acceptance Model and subsequent meta-analytic evidence showing that ease and usefulness remain central predictors of educational technology adoption. However, acceptance should not be interpreted as complete pedagogical success. It only indicates that users are willing and able to use the system; the deeper educational impact still depends on how teachers integrate it into instruction, feedback, and assessment cycles.²⁴

The main limitation of MABBAR is its dependence on continuous internet access. This limitation has both pedagogical and equity implications. Students in areas with unstable connectivity may experience interrupted learning, reduced access, or unequal opportunities to complete listening practice. In addition, a fully online system remains vulnerable to infrastructure disruption and cyber threats, including attacks that may affect access to educational services. These risks reflect a broader concern in the literature on the digital divide and network security. Consequently, the next development stage should prioritize offline or low-bandwidth modes, local caching of audio materials, and more resilient access mechanisms. Without such features, MABBAR may be less inclusive for learners outside stable digital environments.²⁵

²⁴ Fred D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* 13, no. 3 (1989): 319–40.

²⁵ Lythreathis, S., Singh, S. K., & El-Kassar, A. N. (2022). The digital divide: A review and future research agenda. *Technological Forecasting and Social Change*, 175, 121359.

A second limitation concerns learner confidence. The finding that a portion of students still lacked confidence after using the application suggests that digital tools cannot automatically produce autonomous learning. Self-efficacy develops through repeated successful experiences, feedback, modeling, and social support. If learners are asked to work independently before they feel capable, the digital environment may become intimidating rather than empowering. This issue can be interpreted through self-efficacy theory and the concept of scaffolding: learners need temporary supports that help them move gradually from assisted performance to independent mastery. Therefore, MABBAR should be further developed by adding adaptive hints, progressive difficulty levels, model answers, guided practice, teacher feedback features, and possibly peer-supported learning spaces.²⁶

Overall, the pedagogical implication of MABBAR is that Arabic listening assessment can be redesigned as an integrated learning ecosystem. The product's strength lies in its combination of structured input, controlled listening exposure, interactive response, automatic scoring, and teacher-oriented data. At the same time, its limitations show that technological feasibility must be accompanied by inclusive access, learner support, and pedagogical integration. Future development should therefore not only add technical features such as offline mode, richer media, mobile applications, and grading-system integration, but also strengthen the learning supports that help students become confident and independent Arabic listeners.

²⁶ Albert Bandura, "Self-efficacy: Toward a Unifying Theory of Behavioral Change," *Psychological Review* 84, no. 2 (1977): 191–215.

CONCLUSION

The development of MABBAR suggests that Arabic listening assessment can be improved when digital technology is designed not only as a testing tool, but also as a structured support system for learning. Through the 5-stage MANTAP model, MABBAR provides a web-based environment that combines vocabulary preparation, audio-based listening practice, interactive evaluation, and automated feedback for *mahārah al-istimā'*. These features indicate that the platform has the potential to make listening evaluation more organized, accessible, and responsive to students' learning needs. Experts validation, student testing, and user response data show that MABBAR is a feasible and promising medium for supporting Arabic listening instruction at the junior high school level. Its main contribution lies in the integration of pedagogical scaffolding and digital assessment: students are not only assessed after listening, but are also guided before and during the listening process. This finding is important because *mahārah al-istimā'* is often evaluated through conventional audio tasks that provide limited preparation, limited feedback, and limited opportunities for independent repetition.

Nevertheless, the findings should be interpreted within the boundaries of the study. The implementation was conducted with a limited sample in one school context, and the system still depends on stable internet access. In addition, a small proportion of students still reported low confidence, indicating that digital access alone does not automatically ensure learner autonomy. Future research should therefore examine MABBAR in broader institutional settings using stronger experimental designs, while future development should prioritize offline accessibility, mobile application support, teacher-oriented analytics, and more adaptive feedback for listening and pronunciation practice.

REFERENCES

- Albantani, A. M., Madkur, A., & Rozak, A. (2022). Student Self Regulated Learning Strategy In Online Arabic Learning. *Ijaz Arabi Journal of Arabic Learning*, 5(1).
- Aloraini, N., & Cardoso, W. (2022). Social media in language learning: A mixed-methods investigation of students' perceptions. *Computer Assisted Language Learning*, 35(8), 1707–1730. <https://doi.org/10.1080/09588221.2020.1830804>
- Anggraeni, P., Sunendar, D., Maftuh, B., & Sopandi, W. (2023). How Do 6Cs Appearance at Elementary Schools Learning? *DWIJA CENDEKIA: Jurnal Riset Pedagogik*, 7(2).
- Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction*. Longman, (4).
- Fatchiatuzahro, Fahmi, N., Latifah, M., & Nur Azizah. (2025). Pemanfaatan Wordwall Sebagai Media Evaluasi Pembelajaran Maharah Kitabah Pada Mahasiswa PBA IUQI Bogor. *Mauriduna: Journal of Islamic Studies*, 6(1), 138–155. <https://doi.org/10.37274/mauriduna.v6i1.1332>
- Husna, M. A., Inayah, I., Mubarak, F., Taufiqurrahman, T., & Qomariyah, L. (2022). Implementasi materi mahirah al-kalim dalam transformasi kurikulum bahasa Arab perspektif ACTFL pada pondok pesantren salaf. *Arabi: Journal of Arabic Studies*, 7(1), 38–50.
- Krug, S. (2013). *Don't Make Me Think Revisited: A Common Sense Approach*.
- Matthew B, M., A, M. H., & Saldana, J. (2018). *Qualitative Data Analysis: A Methods Sourcebook (The Fourth Edition)*. SAGE Publications.
- Mikita Piastou, "Comprehensive Performance and Scalability Assessment of Front-End Frameworks: React, Angular, and Vue.js," *World Journal of Advanced Engineering Technology*

- and Sciences 9, no. 2 (2023): 366–376, <https://doi.org/10.30574/wjaets.2023.9.2.0153>.
- Krath, J., Schürmann, L., & von Korflesch, H. F. O. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, 125, 106963. <https://doi.org/10.1016/j.chb.2021.106963>
- Dash, N., & Mohapatra, L. M. (2025). Technology empowered teaching and learning in language education: A systematic review. In *Technology driven language learning: Innovations and applications* (pp. 127–145). Springer.
- Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13–35. <https://doi.org/10.1016/j.compedu.2018.09.009>
- Fred D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* 13, no. 3 (1989): 319–40.
- Lythreathis, S., Singh, S. K., & El-Kassar, A. N. (2022). The digital divide: A review and future research agenda. *Technological Forecasting and Social Change*, 175, 121359. <https://doi.org/10.1016/j.techfore.2021.121359>
- De Neira, A. B., Kantarci, B., & Nogueira, M. (2023). Distributed denial of service attack prediction: Challenges, open issues and opportunities. *Computer Networks*, 222, Article 109553. <https://doi.org/10.1016/j.comnet.2022.109553>
- Albert Bandura, "Self-efficacy: Toward a Unifying Theory of Behavioral Change," *Psychological Review* 84, no. 2 (1977): 191–215.
- L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*, ed. Michael Cole et al. (Cambridge,

- MA: Harvard University Press, 1978), 79–91.
- Usher, E. L., & Morris, D. B. (2023). Self-efficacy. In Encyclopedia of mental health (3rd ed., pp. 117–124). Academic Press. <https://doi.org/10.1016/B978-0-323-91497-0.00085-0>
- Haagsman, M. E., Scager, K., Boonstra, J., & Koster, M. C. (2020). Pop-up questions within educational videos: Effects on students' learning. *Journal of Science Education and Technology*, 29(6), 713–724. <https://doi.org/10.1007/s10956-020-09847-3>
- Perrotin, O., Stephenson, B., Gerber, S., Bailly, G., & King, S. (2025). Refining the evaluation of speech synthesis: A summary of the Blizzard Challenge 2023. *Computer Speech & Language*, 90, Article 101747. <https://doi.org/10.1016/j.csl.2024.101747>
- Hase, A., & Kuhl, P. (2024). Teachers' use of data from digital learning platforms for instructional design: A systematic review. *Educational Technology Research and Development*, 72(4), 1925–1945. <https://doi.org/10.1007/s11423-024-10356-y>