

## **Developing a Technology-Integrated Authentic Pragmatic Portfolio Assessment for *Mahārah al-Kalām* in the Malzamah Akbarnas Program**

**Fajrotuz Zahroh<sup>a</sup>, R. Umi Baroroh<sup>b</sup>, Triani Agustin<sup>c</sup>,  
Khaerul Muttakin<sup>d</sup>, Agna Hawari<sup>e</sup>**

<sup>abcd</sup>*Sunan Kalijaga State Islamic University, Yogyakarta, Indonesia,*

<sup>e</sup>*International Islamic University Islamabad Pakistan*

Corresponding author: <sup>a</sup>24204022024@student.uin-suka.ac.id

### **Abstract**

*This study aims to develop a technology-integrated portfolio assessment instrument and rubric for mahārah al-Kalām, based on an authentic, pragmatic approach in the Malzamah Program Akbarnas book. Conventional assessment of mahārah al-Kalām tends to focus on structural linguistic aspects and memorization, so it does not fully represent students' Arabic language skills in the context of real communication. The study applied a Research and Development (R&D) design using the Define–Design–Develop (3D) model. Expert validation involved five experts in Arabic language education and educational assessment. The instrument achieved strong content validity (Aiken's  $V = 0.87$ ) and high internal consistency (Cronbach's  $\alpha = 0.91$ ), indicating high validity and reliability. A limited trial involving 32 students demonstrated the instrument's practicality and usability in real learning settings. The resulting assessment framework integrates authentic pragmatic tasks, structured rubrics, and digital portfolio documentation to assess speaking competence in contextual and communicative settings. This model offers a replicable framework for Arabic language educators seeking to*

*implement more authentic, systematic, and technology-supported speaking assessments.*

## **Abstrak**

Penelitian ini bertujuan untuk mengembangkan instrumen dan rubrik penilaian portofolio mahārah al-kalām berbasis teknologi dengan pendekatan pragmatik autentik pada buku Malzamah Program Akbarnas. Penilaian mahārah al-kalām secara konvensional cenderung menitikberatkan pada aspek kebahasaan struktural dan hafalan, sehingga belum sepenuhnya merepresentasikan kemampuan berbahasa Arab peserta didik dalam konteks komunikasi nyata. Penelitian ini menerapkan desain Penelitian dan Pengembangan (R&D) dengan menggunakan model Define-Design-Develop (3D). Validasi ahli melibatkan lima ahli dalam pendidikan bahasa Arab dan penilaian pendidikan. Instrumen ini mencapai validitas isi yang kuat (Aiken's  $V = 0,87$ ) dan konsistensi internal yang tinggi (Cronbach's  $\alpha = 0,91$ ), yang menunjukkan validitas dan reliabilitas yang tinggi. Uji coba terbatas yang melibatkan 32 siswa menunjukkan kepraktisan dan kegunaan instrumen tersebut dalam lingkungan pembelajaran nyata. Kerangka penilaian yang dihasilkan mengintegrasikan tugas-tugas pragmatik otentik, rubrik terstruktur, dan dokumentasi portofolio digital untuk menilai kompetensi berbicara dalam konteks dan pengaturan komunikatif. Model ini berkontribusi pada pengembangan kerangka penilaian berbicara bahasa Arab yang lebih autentik, sistematis, dan berbasis teknologi, serta dapat dijadikan rujukan praktis bagi pendidik bahasa Arab dalam merancang penilaian yang relevan dengan kebutuhan komunikasi.

## **المخلص**

يهدف هذا البحث إلى تطوير أداة تقويم قائم على ملف الإنجاز (البورتفوليو) مدمجة بالتقنية، مع بناء سلم تقدير خاص بمهارة الكلام، وذلك انطلاقاً من مقارنة أصيلة تداولية في كتاب «برنامج الملازمة أكبارناس». وقد أظهرت الممارسات

التقويمية التقليدية لمهارة الكلام ميلا إلى التركيز على الجوانب اللغوية البنوية والحفظ، بما لا يعكس على نحو كاف كفاءة الطلبة في استعمال العربية ضمن سياقات التواصل الحقيقي. واعتمد البحث منهج البحث والتطوير وفق نموذج «التعريف-التصميم-التطوير». وشملت عملية التحكيم خمسة خبراء في تعليم اللغة العربية والتقويم التربوي. وقد حققت الأداة درجة مرتفعة من صدق المحتوى (معامل أيكن = ٠,٨٧) واتساقا داخليا عاليا (ألفا كرونباخ = ٠,٩١)، مما يدل على تمتعها بدرجة عالية من الصدق والثبات. كما أظهرت تجربة محدودة شملت ٣٢ طالبا قابلية الأداة للتطبيق وسهولة استخدامها في بيئات التعلم الواقعية. ويقدم إطار التقويم الناتج تكاملا بين مهام تداولية أصيلة، وسلام تقدير منظمة، وتوثيقا رقميا للمفاتيح الإنجاز، بهدف تقويم كفاءة الكلام في سياقات تواصلية واقعية. ويتيح هذا النموذج إطارا قابلا للتكرار لمعالي اللغة العربية الراغبين في تطبيق تقويم شفهي أكثر أصالة ومنهجية ومدعوما بالتقنية.

**Keywords:** Arabic speaking skills; assessment rubric; authentic pragmatic approach; portfolio assessment; technology-integrated assessment

## Introduction

Speaking skills (*mahārah al-kalām*) is one of the productive skills in Arabic that requires learners to express ideas and feelings orally and calls for an authentic, contextual, and performance-based approach to assessment.<sup>1</sup> *Mahārah al-kalām* skills not only involve linguistic aspects such as pronunciation, vocabulary, and language structure, but also the ability to use language appropriately in social contexts and for situational functions.<sup>2</sup> Therefore, mastery of *mahārah al-kalām*

---

<sup>1</sup> Cahya Edi Setyawan et al., "Implementasi Penilaian Non-Tes Mahārah Kalām: Studi Kasus Multisitus di Perguruan Tinggi Yogyakarta," *Alsina: Journal of Arabic Studies* 7, no. 1 (2025): 1-28, <https://doi.org/10.21580/alsina.7.1.26108>.

<sup>2</sup> Afif Naufal Andrean Yuristira et al., "Inovasi Penilaian Otentik Maharah Kalam dengan Pendekatan Komunikatif pada Buku Bahasa Arab MI Kelas III," *Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab dan*

is an important competency in learning Arabic, as it functions as a means of communication in daily life.<sup>3</sup>

In assessing *mahārah al-kalām* across various Arabic-language learning institutions, several limitations remain.<sup>4</sup> The assessments carried out are still traditional, such as oral tests or memorization, so they do not comprehensively assess students' communicative skills. This approach to assessment tends to be subjective and incidental, and it lacks documentation of students' development of speaking skills in real-life communication situations.<sup>5</sup> As a result, the learning evaluation has not fully described students' pragmatic ability to use Arabic in real-life communication situations.<sup>6</sup>

The assessment of *mahārah al-kalām* skills requires an authentic, performance-oriented approach, i.e., one that evaluates students' ability to use the language in the context of actual communication. From a pragmatic perspective, language is not only a system of linguistic rules but also an act of communication influenced by the purpose of

---

*Kebahasaaraban* 8, no. 2 (2025): 558–68, <https://doi.org/10.35931/am.v8i2.5226>.

<sup>3</sup> Maskuri Maskuri et al., “Analisis Kesalahan dalam Menggunakan Lahjah Arabiyah pada Maharah Kalam,” *Lahjah Arabiyah: Jurnal Bahasa Arab dan Pendidikan Bahasa Arab* 4, no. 2 (2023): 159–70, <https://doi.org/10.35316/lahjah.v4i2.159-170>.

<sup>4</sup> Naoual Nassiri et al., “Approaches, Methods, and Resources for Assessing the Readability of Arabic Texts,” *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, no. 4 (2023): 1–30, <https://doi.org/10.1145/3571510>.

<sup>5</sup> Angelica Galante and Le Chen, “Advancing Plurilingual and Pluricultural Competence in the Language Classroom,” in *Activating Linguistic and Cultural Diversity in the Language Classroom*, ed. Enrica Piccardo et al. (Cham: Springer International Publishing, 2022), 259–81, [https://doi.org/10.1007/978-3-030-87124-6\\_10](https://doi.org/10.1007/978-3-030-87124-6_10).

<sup>6</sup> Johann Fischer, “The Underlying Action-Oriented and Task-Based Approach of the CEFR and Its Implementation in Language Testing and Assessment at University,” *Language Learning in Higher Education* 10, no. 2 (2020): 301–16, <https://doi.org/10.1515/cercles-2020-2021>.

communication, speech partners, and speech situations.<sup>7</sup> Thus, the assessment of *mahārah al-kalām*, based on pragmatic authenticity, emphasizes students' ability to convey meaning precisely, effectively, and politely in accordance with the social context.<sup>8</sup> This approach requires an assessment model that not only measures the final outcome, but also records the process of developing students' speaking skills.<sup>9</sup>

One relevant approach to meeting these needs is portfolio assessment. Portfolio assessment enables the systematic and continuous collection of evidence of student learning development.<sup>10</sup> In addition, portfolio assessments encourage students to take responsibility for the comprehensive development of speaking competencies.<sup>11</sup> The integration of technology into the portfolio can enhance the assessment of *mahārah al-kalām* through video, audio, or other digital media that can be reviewed and stored on an ongoing basis.<sup>12</sup> Thus,

---

<sup>7</sup> Sultan Almelhes, "Enhancing Arabic Language Acquisition: Effective Strategies for Addressing Non-Native Learners' Challenges," *Education Sciences* 14, no. 10 (2024): 1116, <https://doi.org/10.3390/educsci14101116>.

<sup>8</sup> Seth McCombie and Khaled Al Masaeed, "L2 Arabic Pragmatics in the Classroom: Insights from Language Instructors' Beliefs and Practices," *System* 128 (2025): 103546, <https://doi.org/10.1016/j.system.2024.103546>.

<sup>9</sup> Muhammad M. M. Abdel Latif and Majed M. Alhamad, "Emergency Remote Teaching of Foreign Languages at Saudi Universities: Teachers' Reported Challenges, Coping Strategies and Training Needs," *Education and Information Technologies* 28, no. 7 (2023): 8919–44, <https://doi.org/10.1007/s10639-022-11512-8>.

<sup>10</sup> Hamdan Hadi Purnomo et al., "Inovasi Penerapan Penilaian Portofolio dalam Mahārah Qirā'ah Mata Pelajaran Bahasa Arab," *Jurnal Alfazuna : Jurnal Pembelajaran Bahasa Arab dan Kebahasaaraban* 9, no. 2 (2025): 243–62, <https://doi.org/10.15642/alfazuna.v9i2.4565>.

<sup>11</sup> Paulo Marinho et al., "The Digital Portfolio as an Assessment Strategy for Learning in Higher Education," *Distance Education* 42, no. 2 (2021): 253–67, <https://doi.org/10.1080/01587919.2021.1911628>.

<sup>12</sup> Khaled Ahmed AbdelAl Ibrahim and Yasha Rakhshani, "Portfolio Assessment among Upper-Intermediate Learners: Effects on Learners' Willingness to Communicate, Grit Tendencies, and Motivation," *Asian-*

the use of technology is not only a means of documentation but also a medium that encourages reflection on learning, motivation, and students' active involvement in the assessment process.<sup>13</sup>

A number of previous studies have examined the innovation of assessment in the study of *mahārah al-kalām*. Research by Afif Naufal et al. shows that authentic assessments with a communicative approach can increase students' confidence and participation in Arabic language learning.<sup>14</sup> Another study by Hanan et. Al developed an innovative *mahārah al-kalām* test based on the Merdeka curriculum, which has been shown to improve speaking skills and provide more effective feedback.<sup>15</sup> Meanwhile, Arini Daeng Pasau and Imam Fauji emphasize the use of performance assessment in Arabic textbooks to assess aspects of pronunciation, intonation, expression, and fluency in student speaking.<sup>16</sup>

Most research still focuses on developing performance appraisals or speaking tests and has not comprehensively

---

*Pacific Journal of Second and Foreign Language Education* 9, no. 1 (2024): 36, <https://doi.org/10.1186/s40862-024-00257-1>.

<sup>13</sup> Moch Ubaidillah et al., "Digital Portfolio as an Authentic Assessment Instrument in Arabic Language Learning: A Literature Review," *Jurnal Penelitian Ilmu Pendidikan Indonesia* 4, no. 2 (2025): 838–46, <https://doi.org/10.31004/jpionv4i2.486>; Boumedyen Shannaq, "Digital Formative Assessment as a Transformative Educational Technology," in *Advances in Information and Communication*, ed. Kohei Arai (Cham: Springer, 2024), 471–81, [https://doi.org/10.1007/978-3-031-54053-0\\_32](https://doi.org/10.1007/978-3-031-54053-0_32).

<sup>14</sup> Yuristira et al., "Inovasi Penilaian Otentik Maharah Kalam dengan Pendekatan Komunikatif pada Buku Bahasa Arab MI Kelas III."

<sup>15</sup> Hanan Riswar and Umi Baroroh, "Inovasi Tes Maharah Kalam pada Buku Durus Al-Lughah Al-'Arabiyah Jilid 3 Berbasis Penilaian Kurikulum Merdeka," *JIP - Jurnal Ilmiah Ilmu Pendidikan* 8, no. 2 (2025): 1531–39, <https://doi.org/10.54371/jiip.v8i2.6850>.

<sup>16</sup> Arini Daeng Pasau and Imam Fauji, "Assessment Innovation of Mahārah Kalam in Arabic Language Textbooks for First Grade of Islamic Elementary School," *Research and Development in Education (RaDeN)* 5, no. 1 (2025): 644–51, <https://doi.org/10.22219/raden.v5i1.39459>.

integrated authentic pragmatic approaches, portfolio appraisal models, and digital technology into a single, systematic appraisal framework. Studies that specifically develop assessment instruments and technology-based portfolio rubrics for Arabic textbooks in course institutions remain relatively limited. This gap is also evident in the use of Malzamah books in the Akbarnas program, which are among the teaching materials rich in vocabulary and Arabic communication exercises.

Although the book has great potential for training students' speaking skills, the assessment instruments it includes are not supported by a technology-based portfolio assessment system capable of authentically and sustainably assessing students' pragmatic competencies. This underscores the need for innovation in the development of assessment instruments that are more contextual, systematic, and relevant to modern Arabic language learning.

Based on the above description, this study aims to develop an assessment instrument and rubric for the *mahārah al-kalām* portfolio, using an authentic, technology-based, and pragmatic approach, as explained in the book Malzamah (Akbarnas Program).

This study employs a Research and Development (R&D) approach using the Define–Design–Develop (3D) model. In the Define phase, needs analysis was conducted by reviewing current assessment practices, the Malzamah textbook, and existing speaking assessment procedures within the Akbarnas Program. The Design phase involved creating a technology-integrated portfolio assessment model, developing an analytical rubric, and aligning the rubric with authentic, pragmatic communication tasks. In the Develop phase, these instruments were validated by five experts in Arabic language education and educational assessment. The study participants included 32 students enrolled in the Akbarnas Program,

selected through purposive sampling. Expert validation data were analyzed using Aiken's  $V$ , while reliability was measured using Cronbach's Alpha.

Quantitative data from expert validation were analyzed for each rubric item using Aiken's  $V$  to assess content validity. Aiken's  $V$  was calculated based on expert ratings for each indicator using the formula  $V = \Sigma s / [n(c-1)]$ , where  $s$  is the score assigned by each expert minus the lowest score on the rating scale,  $n$  is the number of experts, and  $c$  is the number of rating categories. Items with  $V \geq 0.80$  were considered valid. Internal consistency was assessed using Cronbach's Alpha, with a threshold of  $\alpha \geq 0.70$  indicating acceptable reliability. Student performance data from the limited trial were analyzed descriptively by calculating mean scores for each rubric dimension to identify patterns of strengths and weaknesses in speaking performance. Qualitative expert and student feedback were analyzed using thematic analysis. The coding process included three stages: open coding to identify meaningful statements, axial coding to group related codes into categories, and selective coding to generate overarching themes related to the instrument's clarity, feasibility, and alignment with pragmatic speaking competencies.

## **Result**

The results of this study are presented across three sequential stages of the 3D development model: Define, Design, and Develop. The Define stage reports the findings of the needs analysis, identifying gaps in existing assessment practices within the Akbarnas Program. The Design stage presents the developed portfolio assessment instrument along with its validation outcomes. The Develop stage describes the technology integration process and the results of the limited trial implementation.

## **Define**

The results of the needs analysis show that the practice of *mahārah al-kalām* assessment in the Akbarnas Program remains dominated by conventional approaches, such as memorizing dialogues and spontaneous classroom oral assessments. The assessment system does not use a standardized rubric, is not supported by digital documentation, and lacks continuous assessment. In contrast, an effective assessment of *mahārah al-kalām* should be carried out through pragmatic, authentic tasks accompanied by a clear assessment rubric, enabling systematic monitoring of students' development of ability. These findings indicate a significant gap between current assessment practices in the Akbarnas Program and the standards required for a comprehensive, authentic, and technology-supported speaking assessment framework.

An analysis of the Malzamah book used in the Akbarnas program shows that the *mahārah al-kalām* material contains authentic pragmatic elements, such as describing images, expressing opinions, engaging in dialogue, and narrating personal experiences.<sup>17</sup> These types of tasks reflect a language-learning approach that emphasizes the communicative use of language in daily life. In language learning, the use of context-based assignments is part of implementing authentic assessments that evaluate students' abilities as a whole through real language performance.<sup>18</sup>

However, the assessment instruments used have not been designed as technology-based portfolio assessments and lack

---

<sup>17</sup> Haifaa Majadly and Athar Haj Yahya, "Beyond the Language: Arabic Language Textbooks in Arab–Palestinian Society as Tools for Developing Social–Emotional Skills," *Education Sciences* 14, no. 10 (2024): 1088, <https://doi.org/10.3390/educsci14101088>.

<sup>18</sup> Muhlis Fajar Wicaksana, *Belajar Mengembangkan Model Penilaian Autentik dalam Pembelajaran Bahasa Indonesia* (Yogyakarta: CV Budi Utama, 2020).

an explicit assessment rubric. These findings show a structural discrepancy between the availability of authentic speaking tasks in Malzamah's book and the absence of a systematic, rubric-based, and technology-mediated assessment framework.<sup>19</sup>

Based on these gaps, the define stage in this study emphasizes the need to develop a technology-based *mahārah al-kalām* portfolio assessment instrument equipped with an assessment rubric and a digital documentation system. The development of this instrument is expected to support a more systematic, transparent assessment process and to monitor students' speaking skills on an ongoing basis. Therefore, the findings at this define stage are the conceptual basis for the design of an assessment model at the next research design stage.

*Table 1*  
*Mahārah al-kalām assessment instrument in the Malzamah book (Akbaras program)*

No	Task Category	Example of Activity
1	Self-Introduction	Students introduce themselves in front of the class based on the dialogue model provided in the textbook
2	Role-Play Dialogue	Students conduct a conversation with classmates through role-play activities related to daily topics
3	Picture Description	Students describe pictures provided in the textbook and explain the context shown in the image
4	Expressing Opinions	Students present their opinions on certain issues (e.g., family, work, or social situations) and provide reasons

---

<sup>19</sup> Imad Aburub, "An Evaluation of the Arabic Language Textbook for First Grade From the Teachers' Point of View," *Journal of Higher Education Theory and Practice* 23, no. 9 (2023): 131–42, <https://doi.org/10.33423/jhetp.v23i9.6133>.

5	Storytelling/ Recounting Experiences	Students tell stories about personal experiences, celebrations, or daily events
6	Explaining Procedures	Students explain the steps in performing an activity, such as cooking or preparing something
7	Describing Daily Activities	Students describe their daily routines or activities based on time and sequence
8	Interview/ Question- Answer	Students ask and answer questions related to professions, places, or daily life

## Design

### *Assessment Instrument Design for Mahārah al-kalām*

In the design stage, the researcher developed a portfolio-based assessment instrument for *mahārah al-kalām* that comprises three complementary forms of assessment: self-assessment, peer assessment, and performance assessment. These three forms of assessment are designed to provide a more comprehensive evaluation of students' speaking abilities, not only from a linguistic perspective but also from a pragmatic communication perspective.<sup>20</sup>

The design of this instrument is based on an authentic pragmatic approach that views language as a means of communication in real-world contexts.<sup>21</sup> In this approach, the evaluation of speaking skills is not limited to grammatical accuracy but also emphasizes the speaker's ability to convey meaning, interact with others, and respond appropriately in communicative situations.<sup>22</sup> Based on this approach, several

<sup>20</sup> Muhimmatul Mukaromah et al., "Adaptasi Maharah Kalam dalam Kehidupan Sehari-hari Peserta Didik Kelas VII MTs Sulamul Huda Ponorogo," *Jurnal Ihtimam* 6, no. 1 (2023): 107–20, <https://doi.org/10.36668/jih.v6i1.463>.

<sup>21</sup> Grant P. Wiggins, *Educative Assessment: Designing Assessments to Inform and Improve Student Performance* (San Francisco: Jossey-Bass Publishers, 1998).

<sup>22</sup> Sonca Vo, "Exploring the Construct of Interactional Competence in Different Types of Oral Communication Assessment," *Interaction Studies*.

assessment indicators were developed to represent the components of speaking competence in a pragmatic manner.<sup>23</sup>

To ensure the instrument's quality, both validity and reliability were systematically established. Content validity was examined through expert judgment involving five experts in Arabic language education and educational assessment. The level of agreement among experts was quantified using Aiken's V, which indicated high content validity.<sup>24</sup> In addition, the Content Validity Index (CVI) was used to confirm the relevance and clarity of each assessment item.<sup>25</sup>

Reliability was assessed using internal consistency analysis with Cronbach's Alpha, which demonstrated a high level of reliability. Furthermore, inter-rater reliability was evaluated within the performance assessment component to ensure consistency in scoring among assessors. These procedures confirm that the instrument is both valid and reliable for assessing students' speaking competence within a pragmatic, technology-enhanced framework.<sup>26</sup> These indicators include some aspects (see Table 2).<sup>27</sup>

---

*Social Behaviour and Communication in Biological and Artificial Systems* 25, no. 1 (2024): 1–35, <https://doi.org/10.1075/is.00022.vo>.

<sup>23</sup> Ahmed Alduais et al., "Arabic Validation of the Pragmatic Language Skills Inventory to Assess Pragmatic Language Development in Preschoolers with and without Pragmatic Language Impairment," *Children* 9, no. 6 (2022): 809, <https://doi.org/10.3390/children9060809>.

<sup>24</sup> Lewis R. Aiken, "Three Coefficients for Analyzing the Reliability and Validity of Ratings," *Educational and Psychological Measurement* 45, no. 1 (1985): 131–42, <https://doi.org/10.1177/0013164485451012>.

<sup>25</sup> Andrea Spoto et al., "Improving Content Validity Evaluation of Assessment Instruments through Formal Content Validity Analysis," *Psychological Methods* 30, no. 2 (2025): 203–22, <https://doi.org/10.1037/met0000545>.

<sup>26</sup> Mostafa Morady Moghaddam, "Enhancing Language Learners' Pragmatic Skills through Augmented Reality: The Speech Act of Request in the Spotlight," *Innovation in Language Learning and Teaching*, 2025, 1–21, <https://doi.org/10.1080/17501229.2025.2544018>.

<sup>27</sup> R. Umi Baroroh and Nafisatun Nisa, "Non-Test Assessment Innovation Performance for Maharah Kalam Through Youtube in the

Table 2  
*Mahārah al-kalām instrument indicators*

Aspects of Pronunciation ( <i>an-nuthq</i> )	Reflect the student's ability to pronounce Arabic sounds accurately and clearly so that the meaning can be well understood by the listener
Fluency ( <i>as-sur'ah</i> )	Demonstrate the ability of students to produce speech smoothly and spontaneously without excessive pauses, which indicates that the speaker is able to organize ideas directly during the communication process
Vocabulary ( <i>al-mufrodat</i> )	Demonstrate the student's ability to select and use contextually appropriate Arabic vocabulary, including the range and precision of word choice to convey meaning accurately and effectively in communicative situations
Grammar ( <i>al-qowa'id</i> )	Reflects the accuracy of the syntactic structure and morphological forms used to convey ideas correctly and clearly
Comprehension	Demonstrate students' ability to understand the speech heard and provide meaningful responses, which is an important element in interactive communication

Overall, these indicators represent pragmatic competence in authentic digital communication, where the speaker must manage pronunciation clarity, grammatical accuracy, vocabulary selection, and the ability to respond interactively to convey meaning effectively.

#### *Standardization of Assessment Scales*

To ensure consistency in the assessment process, this study uses a five-level scale (1–5) for all aspects of the assessment. The interpretation of the scale is as follows: Scale 1: very poor; Scale 2: deficient; Scale 3: adequate; Scale 4: good;

---

Pandemic Era," *Ijaz Arabi Journal of Arabic Learning* 5, no. 1 (2022): 222–33, <https://doi.org/10.18860/ijazarabi.v5i1.14789>.

Scale 5: very good.<sup>28</sup> The following is an example of a *mahārah al-kalām* assessment rubric table that measures according to four main aspects (See Table 3).

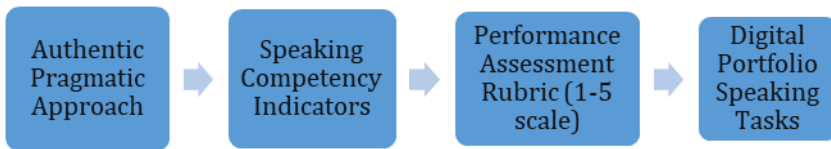
*Table 3*  
*Mahārah al-kalām assessment rubric*

Aspect	Score 1	Score 2	Score 3	Score 4	Score 5
Fluency (Speech Flow)	Very frequent pauses; speech breaks down; message unclear	Frequent hesitation; fragmented speech; meaning difficult to follow	Noticeable pauses, but meaning generally understandable	Minor pauses; speech generally smooth	Natural flow with minimal hesitation; clear delivery
Accuracy (Grammar & Vocabulary)	Frequent errors impede understanding	Many errors; partial comprehension possible	Some errors, but the message is still clear	Minor errors that do not hinder meaning	Accurate grammar and vocabulary; native-like control
Pragmatic Competence (speech acts, sociolinguistic appropriateness, contextual adaptation)	Unable to perform basic speech acts appropriately; responses are inappropriate to the context	Limited ability to perform speech acts; frequent pragmatic inappropriateness	Can perform simple speech acts but inconsistently; occasional pragmatic errors	Appropriate use of common speech acts; mostly contextually appropriate	Performs a range of speech acts effectively; sociolinguistically appropriate and contextually adaptive
Interactional Ability (turn-taking, responsiveness, negotiation of meaning)	Unable to respond or maintain interaction	Limited response; struggles with turn-taking	Basic interaction; delayed or minimal responses	Appropriate responses and turn-taking	Highly interactive; extends conversation and negotiates meaning

<sup>28</sup> Erta Mahyudin et al., *Strategi Mengembangkan Keterampilan Berbahasa Arab Aktif Produktif dan Reseptif* (Jakarta: Publica Indonesia Utama, 2025).

This rubric also serves as an objective guideline for teachers and students, making the assessment process more transparent, consistent, and reflective. Conceptually, the relationship between assessment approaches, indicators, and rubrics can be described in Figure 1.<sup>29</sup>

Figure 1  
Conceptual framework of the technology-integrated portfolio assessment model for mahārah al-kalām



### Portfolio Assessment Instrument Design

#### Self-Assessment

Self-assessment is a form of authentic assessment that positions the self as both the subject and the object of assessment. This assessment uses a 5-point Likert scale to ensure consistency and ease of interpretation of the results, with the following criteria: 1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree. In this case, each student assesses themselves. The format for self-assessment of speaking skills can be seen in the form in Table 4.<sup>30</sup>

Table 4  
Format for self-assessment of speaking skills

<b>Self-Assessment of Speaking Skills</b>	
Name:	Date:
This week I used Arabic to speak with:	

<sup>29</sup> Novanda Alim Setya Nugraha et al., "Development of a Speaking Skills Assessment Rubric Through the English Reporting News Project for High School Students in Kampung Inggris Purbalingga," *Journal on English Language, Teaching and Learning, Linguistic and Literature* 13, no. 1 (2025): 1842–50, <https://doi.org/10.24256/ideas.v13i1.7039>.

<sup>30</sup> Imam Asrori et al., *Evaluasi Dalam Pembelajaran Bahasa Arab*, 5th ed. (Malang: Misykat Indonesia, 2019).

<b>Statement</b>	<b>SD</b>	<b>D</b>	<b>N</b>	<b>A</b>	<b>SA</b>
I think I am successful.					
The person I am talking to understands me.					
I feel comfortable speaking Arabic with other people.					
I understand everything that other people say to me.					
I can do this without any problems.					

### Peer Assessment

Peer assessment is a technique that involves students evaluating each other's performance in relation to the process and level of competency mastery, in both quality and quantity.<sup>31</sup> This assessment technique not only helps teachers collect assessment data but also provides opportunities for students to learn from one another.<sup>32</sup>

This assessment uses a 5-point Likert scale to ensure consistency and ease of interpretation of the results, with the following criteria: 1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree.

The format for peer evaluations of speaking skills is provided in the instrument in Table 5.<sup>33</sup>

*Table 5*  
*Format for peer evaluations of speaking skills*

<b>Peer Assessment: Explaining the Process</b>	
Speaker Name:	Date:
Your Name:	

<sup>31</sup> Nurlev Avana, *Asesmen Pembelajaran: Strategi Evaluasi, Teknik Penyusunan Soal, dan Pelaporan Hasil Belajar* (Purwokerto: PT Revormasi Jangkar Philosophia, 2025).

<sup>32</sup> Desy Seplyana et al., *Metodologi Penelitian dalam Evaluasi Pendidikan* (Bandung: Widina Media Utama, 2025).

<sup>33</sup> Asrori et al., *Evaluasi Dalam Pembelajaran Bahasa Arab*.

Statement	SD	D	N	A	SA
I understand what the speaker is saying.					
The speaker describes how something works.					
The speaker explains at a pace I can follow.					
I feel I can do this myself.					
The speaker uses appropriate Arabic vocabulary and expressions.					

### Performance Assessment

Performance assessment is a form of authentic assessment conducted through direct observation of students' activities as they perform specific tasks or demonstrate specific skills.<sup>34</sup> This assessment uses a rating scale that allows the assessor to rate the level of competency mastery on a continuous, more objective scale.<sup>35</sup>

In this study, the rating scale was adjusted to a 5-point Likert scale to ensure consistency across instruments, with the following criteria: 1 = Very Incompetent; 2 = Incompetent; 3 = Somewhat Competent; 4 = Competent; 5 = Very Competent.

The performance assessment instrument operationalizes the theoretical indicators outlined in Table 2 into four practical scoring dimensions: Fluency, Accuracy of Language, Content and Ideas, and Activity, which collectively represent the linguistic and pragmatic competencies required in authentic Arabic speaking tasks.

<sup>34</sup> Anjin Hu et al., "Digital Technologies in Authentic Assessment in Higher Education: A Systematic Literature Review and Narrative Synthesis," *SAGE Open* 15, no. 3 (2025), <https://doi.org/10.1177/21582440251357198>.

<sup>35</sup> Santi Yudhawati Darmo, "Asesmen Pembelajaran Inklusif," in *Asesmen Pembelajaran: Teori dan Praktik*, ed. Andri Cahyo Purnomo (Serang: Sada Kurnia Pustaka, 2025).

Table 6 shows the performance assessment instrument for *mahārah al-kalām*.<sup>36</sup>

*Table 6*  
*Performance assessment instrument for mahārah al-kalām*

No	Name	Aspects assessed				Total Score
		Fluency	Accuracy Language	Content Ideas	Activity	
1						
2						
3						
4						

Assessment criteria: 1 = Very Incompetent; 2 = Incompetent; 3 = Somewhat Competent; 4 = Competent; 5 = Very Competent. The assessment criteria can be applied as follows: (a) If a student scores 22–25, they are categorized as Highly Competent; (b) If a student scores 18–21, they are categorized as Competent; (c) If a student scores 13–17, they are categorized as Fairly Competent; (d) If a student scores 8–12, they are categorized as Incompetent; (e) If a student scores 0–7, they are categorized as Very Incompetent.

With this adjustment, the assessment categories align directly with the five levels of the Likert scale, making the interpretation of results more accurate and proportional.

#### *Instrument Validity and Reliability*

To ensure the academic quality and psychometric soundness of the developed assessment instrument, a multi-step validation and reliability procedure was implemented. Content validity was evaluated through expert judgment involving five specialists in Arabic language education and language assessment. Each expert rated the relevance, clarity, and representativeness of each indicator in relation to the construct of *mahārah al-kalām*. The ratings were

---

<sup>36</sup> Abdul Munip, *Penilaian Pembelajaran Bahasa Arab* (Yogyakarta: Fakultas Ilmu Tarbiyah dan Keguruan, UIN Sunan Kalijaga Yogyakarta, 2017).

quantitatively analyzed using Aiken's V index. The computed Aiken's V value of 0.87 indicates a high level of agreement among experts and confirms that the indicators adequately represent the intended construct.<sup>37</sup>

Reliability was examined using an internal consistency analysis with Cronbach's alpha. The Cronbach's alpha of 0.91 demonstrates excellent internal consistency, indicating that the items within the instrument consistently measure the same underlying construct of pragmatic speaking competence. In addition, qualitative feedback from experts was systematically reviewed and used to refine the wording of the rubric's indicators and descriptors. Through these procedures, the instrument demonstrates strong evidence of content validity and reliability, supporting its suitability for measuring students' speaking ability within an authentic pragmatic framework.<sup>38</sup>

## ***Develop***

### *Technology-Integrated Portfolio Assessment*

At the development stage, the *mahārah al-kalām* assessment instrument was refined by integrating digital technology into portfolio-based speaking tasks derived from the *Malzamah* textbook used in the Akbarnas Program. This stage aimed to develop a feasible, pedagogically grounded assessment model that integrates authentic speaking tasks, standardized assessment rubrics, and digital documentation of student performance.<sup>39</sup> The development process involved several systematic steps, including expert validation and the

---

<sup>37</sup> Aiken, "Three Coefficients for Analyzing the Reliability and Validity of Ratings."

<sup>38</sup> Slamet Widodo et al., *Buku Ajar Metodologi Penelitian* (Pangkal Pinang: CV Science Techno Direct, 2023).

<sup>39</sup> Fumiyo Nakatsuhara and Vivien Berry, "Use of Innovative Technology in Oral Language Assessment," *Assessment in Education: Principles, Policy & Practice* 28, no. 4 (2021): 343-49, <https://doi.org/10.1080/0969594X.2021.2004530>.

preparation of the final prototype of the technology-integrated portfolio assessment model.<sup>40</sup>

First, the initial prototype of the assessment instrument and scoring rubric was evaluated through expert validation involving specialists in Arabic language education and educational assessment. The validation focused on three main aspects: (1) the relevance of the assessment indicators to the objectives of *mahārah al-kalām* learning, (2) the clarity and consistency of rubric descriptors, and (3) the alignment between the designed speaking tasks and authentic pragmatic communication practices. Feedback from the experts indicated that the instrument's overall structure was appropriate; however, several revisions were recommended to improve the clarity of the rubric descriptors and to ensure greater consistency across the assessment scale. These suggestions were subsequently incorporated into the revised version of the instrument.

Second, the revised instrument underwent a limited trial implementation within the Akbarnas Program learning environment. In this stage, students completed a series of speaking tasks organized as digital portfolio projects. The tasks were designed to reflect authentic communicative situations and included activities such as self-introductions, role-play conversations, picture descriptions, explanations of daily activities, and expressions of opinions on social topics.<sup>41</sup> Students documented their speaking performances through short video recordings with a duration of approximately three

---

<sup>40</sup> Khadeegha Alzouebi, "Electronic Portfolio Development and Narrative Reflections in Higher Education: Part and Parcel of the Culture?," *Education and Information Technologies* 25, no. 2 (2020): 997–1011, <https://doi.org/10.1007/s10639-019-09992-2>.

<sup>41</sup> Grant Wiggins, "The Case for Authentic Assessment," *Practical Assessment, Research, and Evaluation* 2, no. 2 (1990): 1–3, <https://doi.org/10.7275/ffb1-mm19>.

to five minutes. These recordings were then submitted through digital platforms as part of their individual speaking portfolios.

The integration of digital platforms, such as YouTube, Instagram, and WhatsApp, was selected based on pedagogical and practical considerations. These platforms are widely known to students, easily accessible on mobile devices, and allow for flexible documentation and sharing of speaking performances. From a pedagogical perspective, the use of digital media supports authentic language practice, encourages reflective learning, and enables continuous documentation of students' speaking development over time.

Ethical considerations were also addressed in the implementation of the technology-based portfolio system. Students were instructed to upload their videos using private or unlisted settings to protect personal data and ensure that the recordings were shared only within the learning group. In addition, alternative submission methods were provided for students with limited internet access in order to mitigate potential issues related to digital inequality and technological accessibility.<sup>42</sup>

The results of the limited trial indicate that the technology-integrated portfolio assessment instrument was well-received by both teachers and students within the Akbarnas Program learning environment. Students demonstrated the ability to complete the assigned speaking tasks independently using the digital platforms provided, and teachers reported that the rubric descriptors were sufficiently clear to guide consistent scoring. The self-assessment and peer assessment components were observed to encourage student reflection and active engagement during the speaking activities. These observations suggest that the instrument is practically implementable and

---

<sup>42</sup> Shahid Abrar-ul-Hassan et al., "Revisiting Second Language Portfolio Assessment in a New Age," *System* 103 (2021): 102652, <https://doi.org/10.1016/j.system.2021.102652>.

contextually appropriate for authentic Arabic speaking assessment in course-based learning settings.

To enhance readability and avoid an overly detailed presentation of individual task questions, the speaking activities were synthesized into broader categories of pragmatic communication tasks. These categories include self-introduction, opinion expression, storytelling, role-play interaction, procedural explanation, and descriptive narration. Each task is associated with specific pragmatic competencies, appropriate digital platforms for submission, and recommended performance duration.<sup>43</sup> The synthesis of these tasks is presented in a table that provides an overview of the relationships among learning topics, pragmatic speaking tasks, digital platforms, and task duration.

*Table 4*  
*Synthesis Table*

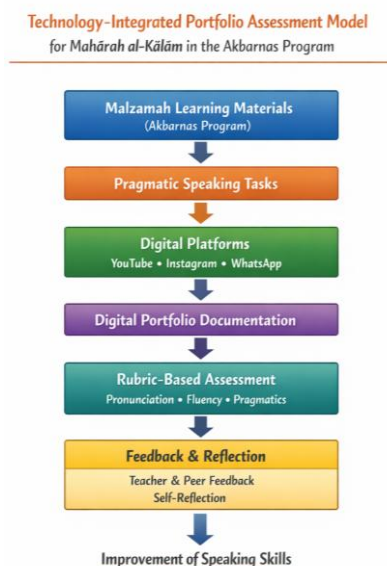
<b>Task Category</b>	<b>Pragmatic Competence</b>	<b>Digital Platform</b>	<b>Duration</b>
Self-Introduction	Self-presentation and identity expression	WhatsApp/ Instagram	3-5 minutes
Role-Play Dialogue	Turn-taking and situational role enactment	YouTube/ WhatsApp	3-5 minutes
Picture Description	Descriptive narration and contextual interpretation	Instagram/ YouTube	3-5 minutes
Expressing Opinions	Opinion expression and argumentation	WhatsApp/ Instagram	3-5 minutes

<sup>43</sup> Khaled Al Masaeed, "Researching and Measuring Second Language Pragmatic Development in Study Abroad: Insight from Arabic," in *Designing Second Language Study Abroad Research: Critical Reflections on Methods and Data*, ed. Janice McGregor and John L. Plews (Cham: Springer International Publishing, 2022), 155-70, [https://doi.org/10.1007/978-3-031-05053-4\\_8](https://doi.org/10.1007/978-3-031-05053-4_8).

Storytelling/ Recounting Experiences	Narrative competence and experiential recounting	YouTube	3–5 minutes
Explaining Procedures	Instructional speech and sequential explanation	YouTube	3–5 minutes
Describing Daily Activities	Informational speech and habitual action description	Instagram	3–5 minutes
Interview/ Question- Answer	Information exchange and interactive questioning	WhatsApp/ Instagram	3–5 minutes

The overall structure of the technology-integrated portfolio assessment model developed in this study is illustrated in the figure below.

*Figure 1*  
*Technology-Integrated Portfolio Assessment Model*



Thus, the final product of the development stage is a technology-integrated portfolio assessment model for *mahārah al-kalām* that combines authentic pragmatic speaking tasks, standardized assessment rubrics, and digital

documentation of students' speaking performances within a digital portfolio framework.<sup>44</sup>

## Discussion

The findings of this study demonstrate that developing a technology-integrated portfolio assessment instrument provides a more comprehensive framework for assessing *mahārah al-kalām* in Arabic language learning. Unlike conventional assessment practices that often rely on spontaneous oral testing or memorization-based performance tasks—such as those identified in the Akbarnas Program, where assessments lacked standardized rubrics, digital documentation, and continuity<sup>45</sup>—the proposed model emphasizes a process-oriented, authentic, and systematically documented assessment of speaking competence. By integrating digital portfolios, students' speaking performances can be recorded, reviewed, and evaluated continuously, allowing teachers to monitor learning progress over time rather than relying solely on single-session oral examinations.<sup>46</sup>

From a theoretical perspective, the authentic pragmatic approach applied in this study aligns with the view that speaking competence should be evaluated not only in terms of linguistic accuracy but also in learners' ability to use language appropriately in specific communicative contexts.<sup>47</sup> Scholars in language assessment emphasize that speaking ability involves integrating linguistic knowledge, discourse competence, and pragmatic

---

<sup>44</sup> Ahmed Mohammed Alkaabi and Asma Khaleel Abdallah, "Portfolio Practices in the Principal Evaluation Process: A Qualitative Case Study," *Heliyon* 10, no. 21 (2024): e39467, <https://doi.org/10.1016/j.heliyon.2024.e39467>.

<sup>45</sup> Salwa Mohamed, "Text Classification and Gradation in Arabic Textbooks," *The Language Learning Journal* 52, no. 6 (2024): 629–49, <https://doi.org/10.1080/09571736.2023.2213695>.

<sup>46</sup> Al-Sabbah Saher et al., "Traditional versus Authentic Assessments in Higher Education," *Pegem Journal of Education and Instruction* 12, no. 1 (2022): 283–91, <https://doi.org/10.47750/pegegog.12.01.29>.

<sup>47</sup> Wiggins, "The Case for Authentic Assessment"; Aco Nasir, *Mengenal Keterampilan Berbicara Dasar* (Yogyakarta: KBM Indonesia, 2021).

appropriateness in real communication situations.<sup>48</sup> Similarly, communicative language testing frameworks highlight that speaking assessment should reflect meaningful language use rather than isolated linguistic performance.<sup>49</sup> This theoretical orientation is reflected in the five assessment indicators developed in this study, pronunciation, fluency, vocabulary, grammar, and comprehension, which together represent the linguistic and pragmatic dimensions of authentic speaking competence in Arabic.<sup>50</sup>

These findings are also consistent with earlier studies on Arabic-speaking assessment, such as research conducted by Yuristira et al., which reported that authentic and communicative assessment approaches can enhance student participation and speaking confidence in Arabic language learning.<sup>51</sup> In language learning contexts, portfolios have been shown to promote learner autonomy and provide richer evidence of skill development compared with traditional testing methods.<sup>52</sup> The positive reception of the instrument during the limited trial, where students demonstrated independent task completion and teachers reported clarity in rubric use, further supports this body of evidence and

---

<sup>48</sup> Glenn Fulcher, "Assessment Literacy for the Language Classroom," *Language Assessment Quarterly* 9, no. 2 (2012): 113–32, <https://doi.org/10.1080/15434303.2011.642041>.

<sup>49</sup> Alser Kamath, *Language Testing and Assessment* (New Rochelle: Magnum, 2024).

<sup>50</sup> Roslizawati Saad et al., "The Use of Digital Teaching Tools to Support Arabic Speaking Skills in Secondary School: A Systematic Literature Review," *Educational Process International Journal* 17, no. 1 (2025): e2025404, <https://doi.org/10.22521/edupij.2025.17.404>.

<sup>51</sup> Yuristira et al., "Inovasi Penilaian Otentik Maharah Kalam dengan Pendekatan Komunikatif pada Buku Bahasa Arab MI Kelas III."

<sup>52</sup> Imam Makruf and Anisatul Barokah, "Improving the Quality of ICT-Based Arabic Learning Assessment With Online Applications," *Journal of Higher Education Theory and Practice* 23, no. 11 (2023): 32–40, <https://doi.org/10.33423/jhetp.v23i11.6216>.

indicates that the model is contextually appropriate for course-based Arabic language learning settings.<sup>53</sup>

However, compared with earlier studies that focused primarily on communicative assessment or performance-based speaking tests, the present study extends the existing literature by integrating digital technology into the portfolio assessment framework. The incorporation of digital platforms such as YouTube, Instagram, and WhatsApp enables systematic documentation of students' speaking performances while facilitating flexible submission and feedback processes.<sup>54</sup> This technology integration supports continuous monitoring of learners' speaking development and creates opportunities for teacher feedback, peer evaluation, and learner self-reflection.<sup>55</sup> The selection of these platforms was particularly appropriate given their accessibility on mobile devices and students' familiarity with them, which reduced barriers to participation and supported equitable engagement across the learning group.

Therefore, the novelty of this study lies in the development of a technology-integrated authentic pragmatic portfolio assessment model that combines three essential elements: authentic pragmatic speaking tasks, structured analytic assessment rubrics, and digital portfolio documentation. This integrated approach provides a more holistic framework for assessing speaking competence in Arabic

---

<sup>53</sup> Mohammed Saleem Ismail Al-Mallahma, "Integrating Technology in Arabic Language Instruction: Enhancing Learner Engagement and Proficiency," in *Business Resilience and Business Innovation for Sustainability*, 2025, 1341–53, [https://doi.org/10.1007/978-3-031-87584-7\\_97](https://doi.org/10.1007/978-3-031-87584-7_97).

<sup>54</sup> Mior Syazril Mohamed Sapawi and Nik Mohd Rahimi Nik Yusoff, "Integrating Technology into the Arabic Language Curriculum: A Systematic Review of Trends, Strategies and Cultural Dimensions," *Social Sciences & Humanities Open* 12 (2025): 101974, <https://doi.org/10.1016/j.ssaho.2025.101974>.

<sup>55</sup> Areej ElSary, "The Impact of a Professional Upskilling Training Programme on Developing Teachers' Digital Competence," *Journal of Computer Assisted Learning* 39, no. 4 (2023): 1154–66, <https://doi.org/10.1111/jcal.12788>.

language learning. Despite these contributions, several limitations should be acknowledged. First, the development and implementation of the assessment model were conducted within the specific context of the Akbarnas Program, which may limit the generalizability of the findings to other Arabic language learning institutions with different curricular structures or technological infrastructures. Second, this study primarily focused on the development and feasibility testing of the assessment model rather than evaluating its effectiveness through large-scale experimental research.

Future studies are therefore recommended to examine the effectiveness of the technology-integrated portfolio assessment model using experimental or quasi-experimental designs with larger, more diverse student populations. In addition, further research may explore integrating advanced digital learning tools or learning analytics systems to support automated documentation and feedback processes in speaking assessment. From a practical perspective, the results of this study suggest that technology-based portfolio assessment has significant potential to improve Arabic-speaking assessment practices by making them more systematic, reflective, and better aligned with authentic communication contexts. This model may serve as a practical reference for Arabic language teachers in designing authentic speaking assessments that integrate digital documentation and reflective learning processes.

## **Conclusion**

This study developed a technology-integrated portfolio assessment model to evaluate *mahārah al-kalām*, grounded in authentic pragmatic speaking tasks, structured analytic rubrics, and digital portfolio documentation within the Malzamah textbook context of the Akbarnas Program. The integration of self-assessment, peer assessment, and performance assessment provides a more comprehensive, systematic, and reflective evaluation of students' speaking competence. The findings indicate that the instrument

demonstrates strong content validity (Aiken's  $V = 0.87$ ) and high internal consistency (Cronbach's Alpha = 0.91), confirming its validity and reliability for use in real learning contexts. A limited trial involving 32 students further confirmed the instrument's practicality and feasibility within authentic Arabic language learning settings. Theoretically, this study contributes to the development of a holistic language assessment framework that integrates pragmatic competence, authentic assessment, and digital technology. From a practical and policy perspective, the model offers a structured reference for educators and institutions to implement more transparent, continuous, and technology-supported speaking assessments aligned with real communicative needs.

However, this study has several limitations, including its implementation within a single institutional context, a relatively small sample size, and a focus on feasibility rather than on effectiveness testing. These limitations may affect the generalizability of the findings. Therefore, future research is recommended to conduct experimental or quasi-experimental studies with larger, more diverse samples and to explore cross-institutional validation and integration with digital learning systems, such as Learning Management Systems (LMS). Further studies may also investigate the long-term impact of this model on students' pragmatic speaking competence and learner autonomy.

## References

- Abrar-ul-Hassan, Shahid et al. "Revisiting Second Language Portfolio Assessment in a New Age." *System* 103 (2021): 102652.  
<https://doi.org/10.1016/j.system.2021.102652>.
- Aburub, Imad. "An Evaluation of the Arabic Language Textbook for First Grade From the Teachers' Point of View." *Journal of Higher Education Theory and Practice* 23, no. 9 (2023): 131–42.  
<https://doi.org/10.33423/jhetp.v23i9.6133>.

- Ahmed AbdelAl Ibrahim, Khaled, and Yasha Rakhshani. "Portfolio Assessment among Upper-Intermediate Learners: Effects on Learners' Willingness to Communicate, Grit Tendencies, and Motivation." *Asian-Pacific Journal of Second and Foreign Language Education* 9, no. 1 (2024): 36. <https://doi.org/10.1186/s40862-024-00257-1>.
- Aiken, Lewis R. "Three Coefficients for Analyzing the Reliability and Validity of Ratings." *Educational and Psychological Measurement* 45, no. 1 (1985): 131-42. <https://doi.org/10.1177/0013164485451012>.
- Al-Mallahma, Mohammed Saleem Ismail. "Integrating Technology in Arabic Language Instruction: Enhancing Learner Engagement and Proficiency." In *Business Resilience and Business Innovation for Sustainability*, 1341-53, 2025. [https://doi.org/10.1007/978-3-031-87584-7\\_97](https://doi.org/10.1007/978-3-031-87584-7_97).
- Alduais, Ahmed et al. "Arabic Validation of the Pragmatic Language Skills Inventory to Assess Pragmatic Language Development in Preschoolers with and without Pragmatic Language Impairment." *Children* 9, no. 6 (2022): 809. <https://doi.org/10.3390/children9060809>.
- Alkaabi, Ahmed Mohammed, and Asma Khaleel Abdallah. "Portfolio Practices in the Principal Evaluation Process: A Qualitative Case Study." *Heliyon* 10, no. 21 (2024): e39467. <https://doi.org/10.1016/j.heliyon.2024.e39467>.
- Almelhes, Sultan. "Enhancing Arabic Language Acquisition: Effective Strategies for Addressing Non-Native Learners' Challenges." *Education Sciences* 14, no. 10 (2024): 1116. <https://doi.org/10.3390/educsci14101116>.
- Alzouebi, Khadeegha. "Electronic Portfolio Development and Narrative Reflections in Higher Education: Part and Parcel of the Culture?" *Education and Information Technologies* 25, no. 2 (2020): 997-1011. <https://doi.org/10.1007/s10639-019-09992-2>.
- Asrori, Imam et al. *Evaluasi Dalam Pembelajaran Bahasa Arab*. 5th ed. Malang: Misykat Indonesia, 2019.

- Avana, Nurlev. *Asesmen Pembelajaran: Strategi Evaluasi, Teknik Penyusunan Soal, dan Pelaporan Hasil Belajar*. Purwokerto: PT Revormasi Jangkar Philosophia, 2025.
- Baroroh, R. Umi, and Nafisatun Nisa. "Non-Test Assessment Innovation Performance for Maharah Kalam Through Youtube in the Pandemic Era." *Ijaz Arabi Journal of Arabic Learning* 5, no. 1 (2022): 222–33. <https://doi.org/10.18860/ijazarabi.v5i1.14789>.
- Darmo, Santi Yudhawati. "Asesmen Pembelajaran Inklusif." In *Asesmen Pembelajaran: Teori dan Praktik*, edited by Andri Cahyo Purnomo. Serang: Sada Kurnia Pustaka, 2025.
- ElSayary, Areej. "The Impact of a Professional Upskilling Training Programme on Developing Teachers' Digital Competence." *Journal of Computer Assisted Learning* 39, no. 4 (2023): 1154–66. <https://doi.org/10.1111/jcal.12788>.
- Fischer, Johann. "The Underlying Action-Oriented and Task-Based Approach of the CEFR and Its Implementation in Language Testing and Assessment at University." *Language Learning in Higher Education* 10, no. 2 (2020): 301–16. <https://doi.org/10.1515/cercles-2020-2021>.
- Fulcher, Glenn. "Assessment Literacy for the Language Classroom." *Language Assessment Quarterly* 9, no. 2 (2012): 113–32. <https://doi.org/10.1080/15434303.2011.642041>.
- Galante, Angelica, and Le Chen. "Advancing Plurilingual and Pluricultural Competence in the Language Classroom." In *Activating Linguistic and Cultural Diversity in the Language Classroom*, edited by Enrica Piccardo et al., 259–81. Cham: Springer International Publishing, 2022. [https://doi.org/10.1007/978-3-030-87124-6\\_10](https://doi.org/10.1007/978-3-030-87124-6_10).
- Grant P. Wiggins. *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass Publishers, 1998.
- Hu, Anjin et al. "Digital Technologies in Authentic Assessment in Higher Education: A Systematic Literature Review and Narrative Synthesis." *SAGE Open* 15, no. 3 (2025). <https://doi.org/10.1177/21582440251357198>.

- Kamath, Alser. *Language Testing and Assessment*. New Rochelle: Magnum, 2024.
- Latif, Muhammad M. M. Abdel, and Majed M. Alhamad. "Emergency Remote Teaching of Foreign Languages at Saudi Universities: Teachers' Reported Challenges, Coping Strategies and Training Needs." *Education and Information Technologies* 28, no. 7 (2023): 8919–44. <https://doi.org/10.1007/s10639-022-11512-8>.
- Mahyudin, Erta et al. *Strategi Mengembangkan Keterampilan Berbahasa Arab Aktif Produktif dan Reseptif*. Jakarta: Publica Indonesia Utama, 2025.
- Majadly, Haifaa, and Athar Haj Yahya. "Beyond the Language: Arabic Language Textbooks in Arab–Palestinian Society as Tools for Developing Social–Emotional Skills." *Education Sciences* 14, no. 10 (2024): 1088. <https://doi.org/10.3390/educsci14101088>.
- Makruf, Imam, and Anisatul Barokah. "Improving the Quality of ICT-Based Arabic Learning Assessment With Online Applications." *Journal of Higher Education Theory and Practice* 23, no. 11 (2023): 32–40. <https://doi.org/10.33423/jhetp.v23i11.6216>.
- Marinho, Paulo et al. "The Digital Portfolio as an Assessment Strategy for Learning in Higher Education." *Distance Education* 42, no. 2 (2021): 253–67. <https://doi.org/10.1080/01587919.2021.1911628>.
- Masaeed, Khaled Al. "Researching and Measuring Second Language Pragmatic Development in Study Abroad: Insight from Arabic." In *Designing Second Language Study Abroad Research: Critical Reflections on Methods and Data*, edited by Janice McGregor and John L. Plews, 155–70. Cham: Springer International Publishing, 2022. [https://doi.org/10.1007/978-3-031-05053-4\\_8](https://doi.org/10.1007/978-3-031-05053-4_8).
- Maskuri, Maskuri et al. "Analisis Kesalahan dalam Menggunakan Lahjah Arabiyah pada Maharah Kalam." *Lahjah Arabiyah: Jurnal Bahasa Arab dan Pendidikan Bahasa Arab* 4, no. 2 (2023): 159–70. <https://doi.org/10.35316/lahjah.v4i2.159-170>.
- McCombie, Seth, and Khaled Al Masaeed. "L2 Arabic Pragmatics in the Classroom: Insights from Language

- Instructors' Beliefs and Practices." *System* 128 (2025): 103546.  
<https://doi.org/10.1016/j.system.2024.103546>.
- Mohamed, Salwa. "Text Classification and Gradation in Arabic Textbooks." *The Language Learning Journal* 52, no. 6 (2024): 629–49.  
<https://doi.org/10.1080/09571736.2023.2213695>.
- Morady Moghaddam, Mostafa. "Enhancing Language Learners' Pragmatic Skills through Augmented Reality: The Speech Act of Request in the Spotlight." *Innovation in Language Learning and Teaching*, 2025, 1–21.  
<https://doi.org/10.1080/17501229.2025.2544018>.
- Mukaromah, Muhimmatul et al. "Adaptasi Maharah Kalam dalam Kehidupan Sehari-hari Peserta Didik Kelas VII MTs Sulamul Huda Ponorogo." *Jurnal Ihtimam* 6, no. 1 (2023): 107–20. <https://doi.org/10.36668/jih.v6i1.463>.
- Munip, Abdul. *Penilaian Pembelajaran Bahasa Arab*. Yogyakarta: Fakultas Ilmu Tarbiyah dan Keguruan, UIN Sunan Kalijaga Yogyakarta, 2017.
- Nakatsuhara, Fumiyo, and Vivien Berry. "Use of Innovative Technology in Oral Language Assessment." *Assessment in Education: Principles, Policy & Practice* 28, no. 4 (2021): 343–49.  
<https://doi.org/10.1080/0969594X.2021.2004530>.
- Nasir, Aco. *Mengenal Keterampilan Berbicara Dasar*. Yogyakarta: KBM Indonesia, 2021.
- Nassiri, Naoual et al. "Approaches, Methods, and Resources for Assessing the Readability of Arabic Texts." *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, no. 4 (2023): 1–30.  
<https://doi.org/10.1145/3571510>.
- Nugraha, Novanda Alim Setya et al. "Development of a Speaking Skills Assessment Rubric Through the English Reporting News Project for High School Students in Kampung Inggris Purbalingga." *Journal on English Language, Teaching and Learning, Linguistic and Literature* 13, no. 1 (2025): 1842–50.  
<https://doi.org/10.24256/ideas.v13i1.7039>.
- Pasau, Arini Daeng, and Imam Fauji. "Assessment Innovation

- of Mahārah Kalam in Arabic Language Textbooks for First Grade of Islamic Elementary School.” *Research and Development in Education (RaDEn)* 5, no. 1 (2025): 644–51. <https://doi.org/10.22219/raden.v5i1.39459>.
- Purnomo, Hamdan Hadi et al. “Inovasi Penerapan Penilaian Portofolio dalam Mahārah Qirā’ah Mata Pelajaran Bahasa Arab.” *Jurnal Alfazuna: Jurnal Pembelajaran Bahasa Arab dan Kebahasaaraban* 9, no. 2 (2025): 243–62. <https://doi.org/10.15642/alfazuna.v9i2.4565>.
- Riswar, Hanan, and Umi Baroroh. “Inovasi Tes Maharah Kalam pada Buku Durus Al-Lughah Al-‘Arabiyah Jilid 3 Berbasis Penilaian Kurikulum Merdeka.” *JiIP - Jurnal Ilmiah Ilmu Pendidikan* 8, no. 2 (2025): 1531–39. <https://doi.org/10.54371/jiip.v8i2.6850>.
- Saad, Roslizawati et al. “The Use of Digital Teaching Tools to Support Arabic Speaking Skills in Secondary School: A Systematic Literature Review.” *Educational Process International Journal* 17, no. 1 (2025): e2025404. <https://doi.org/10.22521/edupij.2025.17.404>.
- Saher, Al-Sabbah et al. “Traditional versus Authentic Assessments in Higher Education.” *Pegem Journal of Education and Instruction* 12, no. 1 (2022): 283–91. <https://doi.org/10.47750/pegegog.12.01.29>.
- Sapawi, Mior Syazril Mohamed, and Nik Mohd Rahimi Nik Yusoff. “Integrating Technology into the Arabic Language Curriculum: A Systematic Review of Trends, Strategies and Cultural Dimensions.” *Social Sciences & Humanities Open* 12 (2025): 101974. <https://doi.org/10.1016/j.ssaho.2025.101974>.
- Septyana, Desy et al. *Metodologi Penelitian dalam Evaluasi Pendidikan*. Bandung: Widina Media Utama, 2025.
- Setyawan, Cahya Edi et al. “Implementasi Penilaian Non-Tes Mahārah Kalām: Studi Kasus Multisitus di Perguruan Tinggi Yogyakarta.” *Alsina: Journal of Arabic Studies* 7, no. 1 (2025): 1–28. <https://doi.org/10.21580/alsina.7.1.26108>.
- Shannaq, Boumedyen. “Digital Formative Assessment as a Transformative Educational Technology.” In *Advances in Information and Communication*, edited by Kohei Arai,

- 471–81. Cham: Springer, 2024.  
[https://doi.org/10.1007/978-3-031-54053-0\\_32](https://doi.org/10.1007/978-3-031-54053-0_32).
- Spoto, Andrea et al. “Improving Content Validity Evaluation of Assessment Instruments through Formal Content Validity Analysis.” *Psychological Methods* 30, no. 2 (2025): 203–22. <https://doi.org/10.1037/met0000545>.
- Ubaidillah, Moch et al. “Digital Portfolio as an Authentic Assessment Instrument in Arabic Language Learning: A Literature Review.” *Jurnal Penelitian Ilmu Pendidikan Indonesia* 4, no. 2 (2025): 838–46. <https://doi.org/10.31004/jpionv4i2.486>.
- Vo, Sonca. “Exploring the Construct of Interactional Competence in Different Types of Oral Communication Assessment.” *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems* 25, no. 1 (2024): 1–35. <https://doi.org/10.1075/is.00022.vo>.
- Wicaksana, Muhlis Fajar. *Belajar Mengembangkan Model Penilaian Autentik dalam Pembelajaran Bahasa Indonesia*. Yogyakarta: CV Budi Utama, 2020.
- Widodo, Slamet et al. *Buku Ajar Metodologi Penelitian*. Pangkal Pinang: CV Science Techno Direct, 2023.
- Wiggins, Grant. “The Case for Authentic Assessment.” *Practical Assessment, Research, and Evaluation* 2, no. 2 (1990): 1–3. <https://doi.org/10.7275/ffb1-mm19>.
- Yuristira, Afif Naufal Andrean et al. “Inovasi Penilaian Otentik Maharah Kalam dengan Pendekatan Komunikatif pada Buku Bahasa Arab MI Kelas III.” *Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab dan Kebahasaaraban* 8, no. 2 (2025): 558–68. <https://doi.org/10.35931/am.v8i2.5226>.