International Journal of Arabic Language Teaching
P-ISSN: 2684-690X E-ISSN: 2686-214X
https://e-journal.metrouniv.ac.id/index.php/IJALT

Balancing Language Skills in Arabic Assessment: Challenges of Curriculum Implementation in an Islamic Integrated School

Adibah Ulaya Putri[1]*, Maman Abdurrahman[2], Mad 'Ali[3]
[1,2,3]Universitas Pendidikan Indonesia, Indonesia

Article History:
Received : 10 April 2026
Revised : 11 June 2026
Published : 26 June 2026

Keywords:
Curriculum Implementation; Educational Evaluation; Islamic Integrated Junior High School; Learning Assessment.

***Correspondence Address:**
adibahulayap05@upi.edu

©authors 2026
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

DOI: 10.32332/ijalt.v8i01.13326

Abstract: This study examines Arabic language assessment practices within curriculum implementation at an Islamic Integrated Junior High School (SMPIT), employing a qualitative instrumental case study design with teacher as participant. Data were collected through semi-structured interviews and document analysis, then analyzed using inductive thematic analysis. The findings reveal that, within this specific institutional context, assessment practices are predominantly structured around written tests that emphasize vocabulary (*mufradat*), grammar (*qawa'id*), and translation activities, a pattern driven primarily by the school's curriculum requirements rather than the teacher's individual choices. Assessment of the four language skills is imbalanced: speaking is partially addressed through dialogue and video projects, while listening, reading, and writing receive limited attention in the formal assessment system. Digital platforms such as Google Forms and Quizizz have been integrated into assessment, though their alignment with pedagogical goals requires further development. Key challenges include low student motivation, limited *hijaiyyah* literacy, and demanding curriculum targets. The study concludes that Arabic language assessment practices in this context are shaped more by institutional curriculum structures rather than by individual teacher agency.

INTRODUCTION

Arabic language learning is fundamentally oriented not only toward mastery of linguistic rules but also toward students' ability to use the language as a means of communication[1]. Therefore, the success of Arabic language instruction cannot be adequately measured solely through students' ability to memorize vocabulary or understand grammatical structures but must also be assessed through their integrated proficiency in listening, speaking, reading, and writing in Arabic[2]. Within this context, assessment constitutes an essential component of

[1] Muflihah, Sholehan, and Muhammad Baihaqi, "Integration of Four Language Skills in Arabic Language Learning" 16, no. 1 (2024): 69–87, https://doi.org/10.24042/albayan.v16i1.18381.
[2] Pradi Khusufi Syamsu et al., "Scrutinizing Integrative Learning in Arabic Instruction" 02, no. 01 (2023): 107–23, https://doi.org/10.54298/ijith.v2i1.78 Scrutinizing.

the instructional process, as it functions to determine the extent to which learning objectives have been achieved, provide insights into the effectiveness of instructional strategies implemented by teachers, and offer feedback for improving subsequent instructional planning. In line with this perspective, Brookhart argues that every assessment practice should yield information useful for educational decision-making at the student, classroom, and school levels[3]. Thus, assessment practices are directly associated with the direction, approach, decision-making process, and overall quality of classroom instruction.

In Arabic language learning, assessment possesses relatively greater complexity compared to other subjects because it encompasses four language skills: *maharah al-istima'* (listening skills), *al-kalam* (speaking skills), *al-qira'ah* (reading skills), and *al-kitabah* (writing skills), each of which requires different forms of evaluation[4]. Assessing reading and writing skills is generally easier to administer through written examinations, whereas assessing listening and speaking skills requires direct practice, observation, and performance-based assessment[5]. The challenges associated with assessing listening and speaking skills transcend local contexts and are not unique to Indonesia.

A study conducted in Turkey, for example, found that post-listening activities tended to dominate listening instruction. In contrast, pre-listening activities were insufficient and not fully aligned with the standards outlined in the Common European Framework of Reference for Languages (CEFR) for the teaching of listening skills[6]. Meanwhile, a study conducted in Malaysia revealed that Arabic language learners at UiTM experienced high levels of speaking anxiety, with fear of evaluation during speaking activities also reported at a relatively high level, largely due to limited opportunities for communicative practice[7].This situation strengthens the argument that the discrepancy between formal assessment practices and learners' actual communicative competence represents a systemic issue that requires sustained attention in Arabic language education across different countries.

Recent shifts in the orientation of Arabic language instruction have also influenced evaluation practices implemented in schools. Current curriculum

[3] Susan M Brookhart, "Educational Assessment Knowledge and Skills for Teachers Revisited," no. 14 (2024): 751, https://doi.org/10.3390/ educsci14070751.

[4] Umareni, Unaisah Soehardin, and Eko Ngabdul Shodikin, "Evaluasi Pembelajaran Bahasa Arab Kelas 7 Di Marhalah Salafiyah Wustho Pondok Pesantren Islamic Center Bin Baz Putri Yogyakarta" 02, no. 01 (2024): 27–35, https://doi.org/10.61553/ascent.v2i1.157.

[5] Ihwan Mahmudi et al., "Analysis Of Online Arabic Language Proficiency Test Questions Based On Bloom ' s Cognitive Taxonomy" 8, no. 2 (2025): 4–11, https://doi.org/10.18860 /ijazarabi. V8i2.29595 I.

[6] Zeynep Arkan, "Evaluation of Listening Activities in Arabic Language Teaching Textbooks Used in Turkey in Light of the CEFR Criteria," *Asian-Pacific Journal of Second and Foreign Language Education*, 2026, https://doi.org/10.1186/s40862-025-00366-5.

[7] Hanis Najwa Shaharuddin et al., "The Level of Student Anxiety in Arabic Speaking Skills," *International Journal Of Research And Innovation In Social Science (IJRISS)* VIII, no. 2454 (2024): 2563–71, https://doi.org/10.47772/IJRISS.

implementation tends to emphasize student-centered learning, strengthening communicative competence, and integrating authentic assessment into the learning process[8]. In language learning, authentic assessment may take the form of speaking practice, presentations, dialogues, projects, and other performance-based assignments.

The principle of assessment for learning has been widely discussed in international literature on second and foreign language education. In their systematic review, Vassiliou et al. found that formative assessment in second language learning has received considerably less attention than summative assessment, despite evidence demonstrating its greater effectiveness in supporting continuous student learning and improvement[9]. Furthermore, Weng and Shen argued that Language Assessment Literacy (LAL) is a critical factor influencing the quality of classroom assessment practices. However, many foreign language teachers, including Arabic language teachers, continue to face challenges in designing and implementing comprehensive assessment instruments. Consequently, teachers are expected not only to deliver instructional materials effectively but also to design evaluation systems that align with the characteristics of language learning[10].

Nevertheless, the implementation of Arabic language assessment in educational practice continues to encounter various challenges. Classroom evaluation practices are still frequently dominated by written tests focusing on mastery of grammar, translation, and vocabulary memorization. The assessment of speaking skills and listening skills tends not to be conducted optimally due to limited instructional time, large class sizes, insufficient learning media, and teachers' limited readiness to implement communicative skill-based assessment[11]. A similar tendency has also been reported in Islamic educational settings in other countries. Studies conducted in Islamic schools in Malaysia indicate that language teachers continue to rely heavily on summative assessment and standardized testing. In contrast, the diversification of assessment methods remains limited due

[8] Siti Nur Salsabyila and Muhammad Taufiq, "The Implementation of the Merdeka Curriculum in Arabic Language Education : A Critical Analysis" 8, no. 2 (2025): 175–92, https://doi.org/10.58223/al-irfan.v8i2.535; Anwar Ridwan Nulloh et al., "Arabic Language Learning Model Based On The Merdeka Curriculum" 4, no. 1 (2025): 129–43, https://doi.org/10.15575/ta.v4i1.44922.

[9] Skevi Vassiliou, Salomi Papadima Sophocleous, and Christina Nicole Giannikas, *Formative Assessment in Second Language Learning: A Systematic Review and an Annotated Bibliography* (France: Research-publishing.net, 2022), https://doi.org/10.14705/rpnet.2022.60.9782383720133.

[10] Bin Shen, "Language Assessment Literacy of Teachers," *Frontiers in Psychology* 13, no. May (2022), https://doi.org/10.3389/fpsyg.2022.864582.

[11] Tirto Luqmanul Hakim, Abdul Haris, and Mokhammad Miftakhul Huda, "Assessment of Speaking Skills (Maharah Kalam) in Arabic Language Lesson at Madrasah Ibtidaiyah Riyadlus Sholihien Jember" 5, no. 2 (2025): 599–605, https://doi.org/10.58905/jse.v5i2.485 Assessment; Muhamad Fuad Hasim et al., "Bridging Assessment and Proficiency : Developing Speaking Skill Test Instruments for Arabic Learners" 7, no. 2 (2025): 526–37, https://doi.org/10.21154/zk6nbg28; Lina Marlina and Khairun Nisa Hasugian, "Analysis of the Difficulties Faced by Seventh Grade Students of Mts Al-Haq Margahayu in Mastering Listening (Maharah Istima ') and Speaking Skills (Maharah Kalam) in Arabic Language Learning," 2025, https://doi.org/10.30997/dt.v6i1.16232.

to time constraints and insufficient professional preparedness[12]. These conditions indicate a gap between the curriculum's orientation and the evaluation practices implemented in classrooms.

In addition, developments in educational technology have influenced assessment practices in Arabic language learning. Digital platforms such as Quizizz, Google Forms, Kahoot, and other evaluation applications have increasingly been utilized by teachers as part of the assessment process. The use of these platforms is considered capable of enhancing student engagement and creating a more interactive evaluation process[13]. However, the implementation of digital-based assessment is also accompanied by several challenges, including limited facilities, teachers' readiness, and students' ability to utilize educational technology effectively[14].

In the international context, similar challenges have also been reported regarding digital assessment. The effective implementation of technology-based assessment requires not only adequate technological infrastructure but also teachers' competence in integrating technology into meaningful assessment practices, particularly in light of the rapid advancement of emerging technologies such as Generative Artificial Intelligence (GenAI), which has recently become increasingly prevalent in educational settings, including higher education[15]. Consequently, current assessment practices in Arabic language learning exist within a relatively complex context, as they must accommodate curriculum demands, technological developments, and the realities of classroom instruction.

This phenomenon becomes increasingly significant when examined within the context of Islamic Integrated Junior High Schools (*Sekolah Menengah Pertama Islam Terpadu*/SMPIT). As Islamic integrated educational institutions, SMPITs possess instructional characteristics that differ from those of general public schools. Learning in SMPITs emphasizes not only academic achievement but also the integration of Islamic values into the instructional process[16]. In Arabic language instruction, this condition influences learning objectives, the selection of

[12] Anwar Farhan Mohamad Marzaini et al., "A Discovery On Malaysian ESL Teachers' Adherence To Policy Mandates In Classroom-BAsed Assesment Practices," *Malaysian Journal of Learning and Instruction* 1, no. 1 (2024): 249–84, https://doi.org/10.32890/mjli2024.21.1.9.

[13] Balqis Fauzatul Rohman and Choirul Anam, "Analysis of Digital-Based Elementary School Learning Assessment Applications" 02, no. 01 (2023): 1–15, https://doi.org/10.71392/ejip.v2i1.47.

[14] Lina Tri Astuty Beru Sembiring and Shafa Salsabila Sembiring, "Practical Application of Assessment Theories for English Teacher through Digital Technologies" 15, no. 2 (2024): 378–87, https://doi.org/10.26877/eternal.v15i2.830; Anzela Jurane Bremane, "The Digital Transformation Of Assessment: Challenges and Opportunities," *Human, Technologies and Quality of Education*, 2021, 352–63, https://doi.org/10.22364/htqe.2021.25.

[15] Olga Viberg et al., "Exploring Teachers ' (Future) Digital Assessment Practices in Higher Education : Instrument and Model Development," *British Journal of Educational Technology*, no. December 2023 (2024): 2597–2616, https://doi.org/10.1111/bjet.13462.

[16] Eka Yuliana Nurohmah and Siti Ma'rifah, "From Paper-Based to Digital Assessment : Adoption and Challenges of Learning Evaluation Applications in Islamic Education" 3, no. 1 (2025), https://doi.org/10.70376/jerp.Received.

teaching materials, and the evaluation practices teachers implement. In SMPIT, Arabic is positioned not merely as a foreign language subject but also as a language closely connected to Islamic studies, with specific targets related to mastery of vocabulary, grammar, and the four language skills.

The dual status of Arabic as a language of religion and a foreign language generates distinctive assessment challenges. Teachers are expected not only to evaluate students' communicative proficiency but also to ascertain the extent to which their linguistic understanding is integrated with their comprehension of Islamic knowledge and values. Similar challenges have also been identified in Arabic language teaching as a community language in the United Kingdom, where the objectives of Arabic instruction often extend beyond purely linguistic goals.

In many cases, Arabic is taught not only for communicative purposes but also to fulfill broader religious and cultural needs, such as understanding the Qur'an and communicating with Arabic-speaking individuals, including family members[17]. Therefore, the assessment practices implemented by Arabic language teachers in SMPITs warrant further investigation, particularly to examine how teachers conduct learning evaluation amid curriculum demands that increasingly emphasize communicative learning and authentic assessment. Several previous studies have examined evaluation methods employed in Arabic language learning across various schools. To systematically position the present study within the existing body of knowledge, a review of relevant prior studies is presented, organized by their respective areas of focus.

Table 1. Mapping of Previous Studies on Assessment in Arabic Language Learning

No.	Research Focus	Research Gaps
1	Assessment instrument development[18]	Comprehensive assessment instruments in SMPIT contexts
2	Digital media in assessment[19]	Integration of digital media in the assessment of the four Arabic language skills in SMPITs
3	Assessment of speaking skills[20]	Performance-based assessment of speaking in SMPITs
4	Curriculum implementation and assessment[21]	Assessment practices within the implementation of the SMPIT curriculum

[17] Rasha Soliman and Saussan Khalil, "The Teaching of Arabic as a Community Language in the UK," *International Journal of Bilingual Education and Bilingualism ISSN:* 0050 (2024), https://doi.org/10.1080/13670050.2022.2063686.

[18] Khusnul Masrifah and Mohammad Ahsanuddin, "Development of HOTS-Based Final Semester Assessment Questions in the Al- ' Ashri Arabic Textbook for 10th-Grade Students of Islamic Senior High School," *Arabiyatuna: Jurnal Bahasa Arab* 9, no. 2 (2025), https://doi.org/10.29240/jba.v9i2.13889.

[19] Sundirah, Sutaman, and Danial Hilmi, "Development Of Interactive Arabic Language Learning Evaluation Medua Using the Wordwall Platform for Students at MIN 1 Malang" 10, no. 3 (2025): 528–35, https://doi.org/10.18860/abj.v10i3.32979; Rohman and Anam, "Analysis of Digital-Based Elementary School Learning Assessment Applications."

[20] Budi Sanjaya et al., "Evaluasi Keterampilan Berbicara Bahasa Arab Siswa Madrasah Aliyah Di Provinsi Jambi," *Arabi : Journal of Arabic Studies* 6, no. 2 (2021): 220–35, https://doi.org/10.24865/ajas.v6i2.339.

[21] Ujiani Pujiani, "Assessment System for Arabic Language Learning within the Independent Curriculum" 02, no. 03 (2024): 165–73, https://doi.org/10.18860/kitaba.v2i3.25565.

5	Integrated Islamic school context[22]	Comprehensive analysis of Arabic language teachers' assessment practices in SMPITs
6	Cross-national Arabic language assessment[23]	Comparative studies of Arabic language assessment practices in Indonesian Islamic schools and international contexts

Based on the mapping presented above, it is evident that previous studies remain fragmented and tend to focus on specific aspects of Arabic language assessment. Some studies have concentrated on the development of assessment instruments, others on the use of digital media, while several have examined the assessment of individual language skills. Based on studies of previous research, no study has comprehensively investigated Arabic language teachers' assessment practices by simultaneously examining assessment forms, the evaluation of the four language skills, the integration of digital media, and the challenges encountered within a specific institutional context, its SMPIT.

Furthermore, research on Arabic language assessment in Indonesia has largely focused on madrasahs and general schools. In contrast, SMPIT, as integrated Islamic educational institutions characterized by a distinctive curriculum framework, have received limited attention in the existing literature. At the international level, efforts to promote more communicative and authentic Arabic language assessment practices in Islamic schools continue to face a range of pedagogical and structural challenges. This situation underscores the importance of studies that specifically examine Arabic language assessment practices in integrated Islamic schools. This context has thus far received limited attention in the existing literature.

Based on the foregoing discussion, this study seeks to examine assessment practices in Arabic language learning within curriculum implementation at an SMPIT. The research focuses on the assessment methods used by Arabic language teachers, the implementation of evaluation across language skills, the use of digital media in the evaluation process, and the various challenges teachers encounter in conducting learning assessments. This study is important because assessment practices are among the primary indicators of how Arabic language instruction is implemented in the classroom. Furthermore, examining evaluation practices may provide insight into the extent to which communicative learning approaches have been implemented in Arabic language instruction within SMPITs. This study is expected to contribute not only to the improvement of Arabic language education in Indonesia but also to provide insights for

[22] Adella Ira Wanti et al., "Model Dan Praktik: Asesmen Formatif Non Paper-Based Dalam Pembelajaran Bahasa Arab," *Arabi : Journal of Arabic Studies* 7, no. 1 (2022): 76–92, https://doi.org/10.24865/ajas.v7i1.419.

[23] Sara Akkuş and Ömer İshakoğlu, "Arapça Öğretiminde Geleneksel - Alternatif Ölçme ve Değerlendirme Yöntemi Kullanımının Bazı Değişkenlere Göre İncelenmesi *," *International Social Sciences Studie Journal* 10 (2024): 0–2, https://doi.org/10.5281/zenodo.125 92780.

comparative research on Arabic language assessment practices in integrated Islamic schools across different countries.

Accordingly, this study is intended to address several research questions: (1) how are assessment practices implemented by teachers in Arabic language learning at SMPITs; (2) how is the evaluation of the four Arabic language skills conducted; (3) how are digital media utilized in Arabic language learning evaluation; and (4) what challenges are encountered by teachers in implementing Arabic language assessment. In line with these questions, this study aims to analyze assessment practices in Arabic language learning within curriculum implementation at SMPITs, including the evaluation methods employed by teachers, the implementation of language-skill assessment, the use of digital media in evaluation, and the challenges encountered by teachers during the assessment process. Through this study, it is expected that a more contextualized understanding of Arabic language assessment practices within curriculum implementation in SMPIT will be obtained, while also contributing to the development of more communicative and contextually appropriate Arabic language evaluation practices in classroom learning.

METHOD

This study employed a qualitative approach within an interpretive paradigm and adopted a case study design. This approach was selected because the study focused on understanding assessment practices in Arabic language learning within curriculum implementation at an SMPIT, based on teachers' experiences and interpretations in the classroom. Epistemologically, the interpretive paradigm was adopted based on the assumption that assessment practices are inherently contextual phenomena that cannot be adequately understood through numerical measurement alone. Rather, they require an in-depth exploration of the meaning's individuals construct within their real-life contexts.

The study employed an instrumental case study which focuses on understanding the case itself, an instrumental case study uses a particular case to gain broader insights into a specific issue or phenomenon. In the present study, the case of an Arabic language teacher at a single SMPIT was selected not primarily because of its uniqueness as an individual case, but because it serves as a paradigmatic representation of a teacher actively engaged in assessment practices within the context of curriculum implementation in an educational institution characterized by distinctive Islamic, curricular, and pedagogical features. The selection of a single participant from one school was guided by the methodological consideration that the depth and richness of data obtained from an intensively documented case would provide greater value in addressing the

exploratory and descriptive research questions than a broader but less in-depth investigation.

The participant in this study was an Arabic language teacher at an SMPIT, selected purposively based on the following criteria: (1) having actively taught Arabic at the Integrated Islamic Junior High School level for at least two years; (2) being directly involved in the planning, implementation, and assessment of Arabic language instruction; and (3) teaching classes that had implemented the current curriculum. A more detailed profile of the participant, including educational background, teaching experience, grade levels taught, and the number of students per class, is presented in the findings section to provide sufficient contextual information for interpreting the study's findings.

The research data were collected through semi-structured interviews and document analysis. The interviews were conducted to obtain information on the assessment methods used by teachers, the evaluation of the four language skills, the use of digital media in evaluation, and the challenges encountered in implementing Arabic language-learning assessment. The interview instrument was developed based on the research focus and was flexibly adjusted throughout the data collection process. The interview guide was developed based on four main aspects that reflected the research questions: (1) assessment practices, including the forms of assessment employed, the frequency of assessment, the coverage of the four language skills, and the use of authentic assessment; (2) curriculum implementation within the assessment system, including the alignment of assessment practices with curriculum requirements and the influence of the foundation's distinctive curriculum; (3) the use of digital media in assessment, including the types of media utilized and their perceived effectiveness; and (4) teachers' challenges and reflections, including obstacles encountered in conducting assessments and teachers' perspectives on ideal assessment practices. These four aspects subsequently served as the basis for the initial coding process during data analysis. The interview was conducted face-to-face, audio-recorded with the participant's consent, and subsequently transcribed verbatim for analytical purposes.

In addition to interviews, this study also employed document analysis to strengthen the research findings. The analyzed documents included the syllabus, semester program, teaching modules, assessment instruments, and sample test items used in Arabic language instruction. Document analysis was conducted to examine the relationship between instructional planning and the assessment practices teachers implemented in the classroom. The use of these two data sources (interviews and documents) constituted data source triangulation aimed at enhancing the credibility of the findings. This was achieved by comparing the

teacher's verbal accounts with documentary evidence contained in the instructional planning documents that she has produced.

The data collection process was conducted in several stages: a review of instructional documents, followed by semi-structured interviews, transcription, coding, and conclusion. Data were analyzed using an inductive thematic analysis approach based on the procedures proposed by Miles, Huberman, and Saldaña (2014), which comprise three main stages: (1) data condensation, involving the reduction and selection of raw data through initial coding; (2) data display, involving the organization of codes into categories and themes; and (3) drawing and verifying conclusions[24]. The themes were developed in a data-driven (inductive) manner, meaning that they emerged directly from the empirical data rather than being imposed by a predetermined theoretical framework.

In order to ensure the trustworthiness of the study, the researcher adopted the framework proposed by Lincoln and Guba, which encompasses four criteria: credibility, transferability, dependability, and confirmability[25]. Credibility was established through: (a) data source triangulation between interview data and instructional documents; and (b) member checking, whereby the researcher returned the preliminary interpretations to the participant for confirmation of their accuracy. Dependability was strengthened through detailed and consistent documentation of the research procedures. Confirmability was maintained by ensuring that all interpretations of the findings could be traced back to the original data rather than relying solely on the researcher's subjective judgments. Transferability was enhanced by rich, thick descriptions of the research context, enabling readers to assess the relevance of the findings to comparable settings.

The study was conducted in accordance with established research ethics principles. Prior to data collection, permission was obtained from the school, and written informed consent was secured from the participant. The participant was informed of the study's purpose, the right to withdraw from participation at any time without penalty, and the measures taken to ensure confidentiality. To protect privacy and maintain confidentiality, both the participant's identity and the school's name were anonymized in the research report.

[24] Matthew B Miles, A Michael Huberman, and Johny Saldana, *Qualitative Data Analysis*, ed. Helen Salmon Editorial, 3rd ed. (United States of America: SAGE Publications, 2014), https://www.metodos.work/wp-content/uploads/2024/01/Qualitative-Data-Analysis.pdf.

[25] Yvonna S. Lincoln and Egon G. Guba, *Naturalistic Inquiry* (Beverly Hills, CA: SAGE Publications, 1985).

RESULT AND DISCUSSION

Before presenting the findings, a brief profile of the research participant should be provided. The participant was a female Arabic language teacher with more than five years of teaching experience at an SMPIT. She holds a degree in Arabic Language Education from an Islamic higher education institution and currently teaches Arabic to students in Grades VII through IX. The number of students in each class ranges from approximately 25 to 30. This profile is important to consider when interpreting the findings, as teaching experience, educational background, and classroom conditions may indirectly influence the teacher's choice of assessment strategies and instruments[26].

The findings and discussion are presented according to the four research questions that informed the study. Within each section, themes emerging from the interview data are discussed in relation to: (1) assessment practices; (2) evaluation of the four Arabic language skills conducted; (3) the use of digital media in assessment; and (4) teachers' challenges and reflections on assessment practices. These themes should not be viewed as isolated phenomena; rather, they reflect underlying structural tensions among the distinctive curricular demands of the educational foundation, teachers' pedagogical autonomy, and the principles of communicative assessment emphasized in literature.

Based on the results of the interviews and the analysis of instructional documents, assessment practices in Arabic language learning at a SMPIT demonstrate an interrelationship among the foundation's curriculum targets, the forms of assessment employed by teachers, and the actual conditions of classroom instruction. The research data was obtained through interviews with an Arabic language teacher and analysis of the syllabus, teaching modules, assessment instruments, and sample test items used in Arabic language learning.

The results of the document analysis indicate that Arabic language instruction at the school had incorporated structured assessment planning within the syllabus, as illustrated in the following table.

Table 2. Semester 1 Syllabus

Session	Main Topic	Learning Process
1	*Muqoddimah* (Introduction)	Teacher introduces the course overview, learning objectives, and Arabic particle system (huruf); students share prior knowledge through Q&A; teacher explains the relevance of particles in Quranic and daily Arabic texts
2	*Harf Jar* + Translation	Teacher explains the definition and types of harf jar (e.g., مِنْ، إِلَى، عَنْ، عَلَى، فِي، رُبَّ، بِ، لِ، كَ) with examples from the Quran; students identify harf jar in sample sentences; guided translation practice in pairs

[26] Hailay Tesfay Gebremariam and Abate Demissie Gedamu, "The Influence of K-12 Language Teachers ' Demographic Profiles on Authentic Assessment Practice," *Indonesiann Journal Of Applied Linguistics* 13, no. 2 (2023): 443–54, https://doi.org/10.17509/ijal.v13i2.63085.

3	*Harf Jar* + Translation	Students review previous material through oral drill; teacher provides contextual sentences for translation; students practice constructing sentences using harf jar independently; class discussion and correction
4	Daily Assessment 1	Written test covering *Harf Jar* material: particle identification, sentence completion, and Arabic-Indonesian translation
5	*Harf Istifham* + Translation	Teacher introduces interrogative particles (e.g., هَلْ، مَنْ، مَا، أَيْنَ، كَيْفَ، مَتَى); students practice forming questions from declarative sentences; translation exercise from Arabic to Indonesian
6	*Harf Istifham* + Translation	Communicative drill: students ask and answer questions using harf istifham; group activity translating short Arabic dialogues containing interrogative particles; teacher provides feedback
7	*Muroja'ah* (Review)	Structured review of *Harf Jar* and *Harf Istifham*; students complete a gap-fill and matching exercise; peer-correction followed by teacher clarification of common errors
8	Mid-Semester Assessment	Written mid-semester exam covering Sessions 1–7: particle identification, translation, and short sentence construction
9	*Harf Syarat* + Translation	Teacher explains conditional particles (e.g., إِنْ، إِذَا، مَنْ، مَا، مَهْمَا) and their grammatical effect; students analyze conditional sentences from the Quran; guided translation practice
10	*Harf Syarat* + Translation	Students construct their own conditional sentences; translation exercise (Arabic to Indonesian and vice versa); class presentation of sentences for peer review
11	*Harf Athaf* + Translation	Teacher introduces conjunctive particles (e.g., وَ، فَ، ثُمَّ، أَوْ، بَلْ); students identify them in Quranic verses; guided practice connecting words and clauses; translation exercise
12	Daily Assessment 2	Written test covering *Harf Syarat* and *Harf Athaf*: identification, sentence construction, and translation
13	*Harf Athaf* + Translation	Extended practice: students translate compound sentences using harf athaf; group activity producing a short paragraph with conjunctions; teacher feedback and correction
14	*Harf Jawazim* + Translation	Teacher explains *jazm*-inducing particles (e.g., لَمْ، لَمَّا، لاَ النَّاهِيَة، لاَمْ الأَمْر) and their effect on the *mudhari'* verb; students practice verb conjugation in context; translation exercise
15	*Muroja'ah* (Review)	Comprehensive review of Sessions 9–14; students complete a mixed-particle exercise; teacher-led discussion addressing common misconceptions; preparation tips for end-of-semester exam
16	End-of-Semester Assessment	Written end-of-semester exam covering all Semester 1 material: particle recognition, grammatical analysis, and full-sentence translation

Table 3. Semester 2 Syllabus

Session	Main Topic	Learning Process
1	*Muroja'ah* (Review)	Recap of Semester 1 material through oral questioning and a short written warm-up quiz; teacher identifies learning gaps and previews Semester 2 topics
2	*Harf Nawashib* + Translation	Teacher explains nasb-inducing particles for mudhari' verbs (e.g., أَنْ، لَنْ، كَيْ، حَتَّى، لِ) with Quranic examples; students identify affected verbs; guided translation practice
3	*Harf Nawashib* + Translation	Students drill verb conjugation after *harf nawashib*; pair work translating sentences; independent written exercise with teacher feedback

4	Daily Assessment 3	Written test covering *Harf Jawazim* (Sem. 1, Session 14) and *Harf Nawashib:* verb form identification, particle selection, and translation
5	*Harf Nawasikh* + Translation	Teacher introduces *nawasikh* (كَانَ وَأَخَوَاتُهَا and إِنَّ وَأَخَوَاتُهَا) and their effect on *mubtada'-khabar*; students analyze example sentences; guided translation from Arabic to Indonesian
6	*Harf Nawasikh* + Translation	Students practice constructing sentences with *nawasikh*; group translation activity using Quranic verses; class discussion on grammatical changes and meaning shifts
7	*Muroja'ah* (Review)	Structured review of *Harf Nawashib* and *Harf Nawasikh*; mixed exercise combining identification, transformation, and translation; peer-checking with teacher guidance
8	Mid-Semester Assessment	Written mid-semester exam covering Sessions 1–7 of Semester 2: particle function, grammatical analysis, and translation
9	*Harf Laa* + Translation	Teacher explains the various functions of لَا (*laa nafi, laa nahiyah, laa nafi lil jins*) with contextual examples; students distinguish between types; translation practice
10	*Harf Laa* + Translation	Students complete sentence transformation exercises using different types of laa; group activity translating Quranic verses; teacher correction and reinforcement
11	*Harf Nida* and *Istiqbal* + Translation	Teacher introduces vocative particles (يَا، أَيَا، هَيَا) and future/expectation particles (سَ، سَوْفَ); students practice usage in context; translation exercise
12	Daily Assessment 4	Written test covering *Harf Laa* and *Harf Nida* & *Istiqbal*: identification, function differentiation, and translation
13	*Huruf Nida* and *Istiqbal* + Translation	Extended communicative practice: students construct vocative and future-tense sentences; group translation of short Arabic texts; peer presentation and feedback
14	*Huruf Istitsna* and *Taukid* + Translation	Teacher explains exceptive particles (إِلَّا، غَيْرُ، سِوَى) and emphasis particles (إِنَّ، لَامُ التَّوْكِيد، نُونُ التَّوْكِيد); students analyze and translate example sentences; independent practice
15	*Muroja'ah* (Review)	Comprehensive review of all Semester 2 material; students complete a full mixed-particle worksheet; teacher-led Q&A; study strategies shared for the final exam
16	End-of-Year Assessment	Written end-of-year exam covering all Semester 1 and Semester 2 material: particle recognition, grammatical analysis, sentence construction, and translation

The document indicates that learning evaluation was conducted periodically through Daily Assessments (*Penilaian Harian*/PH), Mid-Semester Assessments (*Penilaian Tengah Semester*/PTS), and Final Semester Assessments (*Penilaian Akhir Semester*/PAS). This finding is consistent with the teacher's explanation during the interview, in which it was stated that evaluation was conducted in almost every chapter of the instructional material.

> *"There are four PH assessments. In the first semester, the sequence is PH 1, PTS, PH 2, and PAS. The second semester follows the same pattern: PH 3, PTS, PH 4, and PAT. So, in one academic year, there are four PH assessments, and they are always evaluated and formally reported."*

These findings indicate that Arabic language learning evaluation had been implemented systematically and in a structured manner in accordance with the established instructional planning. This is consistent with the view of Jovita et al.

(2025), who argue that learning evaluation is a continuous process conducted to monitor students' progress and learning achievement, while also enabling teachers to provide the feedback necessary to support students' learning development[27]. In this context, evaluation functions not only as a tool for measuring learning outcomes but also as a form of academic accountability within the instructional process.

Arabic Language Assessment Practices in Classroom Implementation

Although evaluation was conducted periodically, the interview results revealed that the dominant form of assessment remained written tests. The teacher explained that written assessments were considered more objective and more feasible to implement in classrooms with relatively large numbers of students.

> *"In PH assessments, we usually review the material together because even the original scores often do not reach the minimum competency standard. Therefore, the scores are usually supplemented with daily performance, attendance, and classroom participation. In my opinion, objective assessment is best represented through written tests because, in oral assessment, some students perform while others do not. Although written tests are not entirely relevant for assessing maharah al-kalam, they are still used to maintain objectivity in assessment."*

These findings demonstrate that evaluation practices continue to emphasize cognitive and administrative aspects of assessment. The teacher also adjusted students' scores based on classroom participation and daily performance, as written evaluations frequently failed to meet the minimum competency criteria. Written tests were preferred because they were considered capable of minimizing subjectivity in the assessment process, particularly in classrooms with a large number of students. On the other hand, the teacher also acknowledged that using written tests to assess *maharah al-kalam* did not fully reflect students' actual speaking abilities.

This perspective warrants critical examination. The belief that written tests are inherently more "objective" is rooted in a positivist epistemology that prioritizes standardization and quantitative measurement. However, within the language assessment literature, such formal notions of objectivity often come at the expense of construct validity, that is, the extent to which an assessment instrument accurately measures the construct it is intended to assess[28]. In the context of *maharah al-kalam*, the use of written tests measures knowledge of speaking rather than the ability to speak. This situation reflects a discrepancy between what the teacher conceptually understands (namely, that written tests are not relevant for *maharah al-kalam*) and what is actually implemented in practice. In

[27] Jovita Ponomariovien, Daiva Jakavonyte-Staškuviene, and Frederic Torterat, "Implementing Competency-Based Education Through the Personalized Monitoring of Primary Students ' Progress and Assessment," 2025, https://doi.org/10.3390/educsci15020252.

[28] Vahid Aryadoust, "The Vexing Problem of Validity and the Future of Second Language Assessment," *Language Testing* 40(1) 8 –1 (2023): 8–14, https://doi.org/10.1177/02655322221125204.

the language assessment literature, such a discrepancy is often described as a gap between knowledge and practice in teachers' language assessment literacy[29]. Assessment should be repositioned as an integral component of the learning process (*assessment for learning*) through the provision of constructive feedback. However, in practice, evaluation is still frequently understood as a means of measuring final learning outcomes rather than as a means of gradually developing students' language abilities.

Analysis of the evaluation items used in instruction further revealed that the assessment forms remained predominantly focused on *mufradat* and translation-based questions. This was evident in the evaluation tasks administered via Google Forms, which primarily emphasized students' ability to memorize vocabulary and translate sentences, as illustrated in the table below.

Table 4. Practice Activity in Quizizz

No	Question	Option A	Option B	Option C	Option D
1	The Arabic word for "inside / within" is…	مِنْ	أَيْ	فِي	مَنْ
2	The meaning of the letter إِلَى is…	To / Towards	From	With	If
3	The meaning of the word يُقِيمُونَ in the verse الَّذِينَ يُقِيمُونَ الصَّلَاةَ is…	They Establish	They Spend / Donate	They Believe	They Fear Allah (Taqwa)
4	The meaning of the word مُتَّقِينَ in the verse ذَٰلِكَ الْكِتَابُ لَا رَيْبَ ۛ فِيهِ ۛ هُدًى لِّلْمُتَّقِينَ is…	People who fear Allah (righteous)	Good people	People who believe	People who act / practice
5	The meaning of the word أُولَٰئِكَ in the verse أُولَٰئِكَ عَلَىٰ هُدًى مِّن رَّبِّهِمْ is…	He / She	They (those)	You	We

The dominance of vocabulary- and translation-based assessment cannot be separated from the foundation's distinctive curriculum, which emphasizes mastery of vocabulary and the ability to translate the Qur'an. The teacher explained that the objectives of Arabic language learning at the SMPIT were primarily focused on translation mastery rather than on communicative Arabic use.

> *"The foundation expects students, by the time they graduate, to have memorized the translation of one juz of the Qur'an, whereas Arabic language competitions outside the school tend to focus more on everyday communicative language. Therefore, the contexts are different."*

This statement reflects a difference in orientation between the foundation's curriculum targets and the communicative competency demands of Arabic

[29] Indah Puspawati, Maharani Khansa, and Utami Widiati, "Developing EFL Teachers ' Language Assessment Literacy : A Systematic Literature Review on Teacher Training Programs," *The Electronic Journal for English as a Second Language Developing* 28, no. 2 (2024): 1–21, https://doi.org/10.55593/ej.28110a8.

language learning. This tension reflects what the curriculum literature describes as a conflict between the institutional curriculum and the communicative curriculum. Institutional demands encourage teachers to prioritize memorization and translation targets, whereas contemporary language education emphasizes the development of authentic communicative competence. From a curricular perspective, Arabic language instruction in the SMPIT remains strongly influenced by a vocabulary- and translation-oriented approach, which, in turn, shapes the teacher's assessment methods. A similar pattern has also been reported in the context of Indonesian madrasahs, where institutional curricular orientations are a primary factor shaping teachers' assessment practices, often exerting a stronger influence than either professional training or the national curriculum[30].

Balancing the Assessment of the Four Language Skills

The interview findings revealed that the evaluation of Arabic language learning had not yet comprehensively and proportionally covered the four language *maharah*. The teacher explained that the only skill currently receiving more concrete evaluation attention was *maharah al-kalam*, particularly in Grade VII.

> *"Here, we have only begun emphasizing maharah al-kalam, and even that applies only to seventh-grade students. So, besides evaluating qawa'id, there is also assessment in maharah al-kalam through text translation and hiwar practice with classmates."*

In addition, the teacher assigned a project-based task to create a simple how to video on the topic of residence, which students subsequently uploaded to social media as part of the learning assessment. The use of video-based projects indicates the teacher's effort to implement authentic assessment in Arabic language learning.

However, this initiative toward authentic assessment warrants critical scrutiny, as its scope remains highly limited. It is confined to *maharah al-kalam* at a single grade level and has not been meaningfully extended to *maharah al-istima'*, *maharah al-qira'ah*, or *maharah al-kitabah*. This imbalance should not be viewed merely as a technical issue, rather, it reflects deeper structural challenges. First, the assessment of *maharah al-istima'* was almost absent from the instructional documents analyzed, suggesting that listening skills have not yet been systematically incorporated into the formal assessment framework. This is particularly noteworthy given that studies across various contexts have consistently shown that *maharah al-istima'* is among the most neglected skills in Arabic language assessment, largely due to limitations in instructional media and

[30] Wahidmurni, Samsul Susilawati, and Achmad Anwar Abidin, "Resistance to Curriculum Changes among Teachers of Madrasah Ibtidaiyah in Indonesia," *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam* 9, no. 1 (2024): 16–27, https://doi.org/10.31538/ndh.v9i1.4300.

teachers' expertise in designing valid listening assessment instruments[31]. Second, the assessment of *maharah al-kitabah* primarily consisted of copying and translation rather than the production of creative, communicative written texts. As a result, the claim that "all four language skills have been assessed" is not fully supported by the data and should therefore be acknowledged as a limitation of the findings.

These results suggest that the school's assessment framework has begun to expand beyond traditional written examinations by incorporating authentic assessment practices, primarily through video-based speaking tasks and dialogue performance. However, this transition is only partially realized. Authentic assessment is almost exclusively limited to *maharah al-kalām* and is only implemented at a single grade level, rather than representing a comprehensive transformation of assessment practices. As a result, the assessment system continues to demonstrate a significant disparity in the four Arabic language skills, indicating that the curriculum implementation has not yet achieved the integrated competence that communicative language teaching and competency-based assessment envision. This imbalance should be interpreted as evidence of broader structural constraints that are entrenched within the enactment of the curriculum, rather than as an isolated pedagogical shortcoming.

The instructional documents and assessment instruments that were examined were largely devoid of *maharah al-istimāʿ* (listening), which is the most significant lacuna. Its absence implies that the school's formal assessment framework has not yet systematically integrated listening. This discovery is indicative of a persistent challenge that has been identified in the field of Arabic language education. Listening is frequently marginalized due to the fact that instructors frequently lack the necessary audio resources, technological support, and expertise to develop reliable and valid listening assessment tasks. Consequently, learners are afforded inadequate opportunities to exhibit receptive oral comprehension, despite the fact that listening is a fundamental aspect of communicative competence and serves as the foundation for subsequent speaking development.

The evaluation of *maharah al-kitābah* (writing) demonstrates a comparable imbalance. Despite the formal inclusion of writing activities, they were primarily limited to translation assignments and copying exercises. The primary objective of these activities is to assess lexical recall and orthographic accuracy, rather than the capacity to compose written discourse that is contextually appropriate, meaningful, and coherent. As a result, the evaluation prioritizes lower-order linguistic knowledge while disregarding higher-order writing competencies, including organization, audience awareness, creativity, and communicative

[31] Arkan, "Evaluation of Listening Activities in Arabic Language Teaching Textbooks Used in Turkey in Light of the CEFR Criteria."

effectiveness. This discovery implies that writing assessments continue to be product-oriented rather than process-oriented and has not yet incorporated the most recent principles of communicative and performance-based language assessment.

In aggregate, these results argue against the notion that all four Arabic language skills have been evaluated in a balanced and integrated manner. The depth, frequency, and authenticity of the assessments of each skill change significantly, despite the fact that they are all present in the curriculum to varying degrees. The continued reliance on traditional techniques for listening, reading, and writing, as well as the dominance of speaking among performance-based assessments, suggest that assessment practices are more influenced by institutional curriculum priorities and implementation constraints than by a comprehensive pedagogical commitment to balanced language competence. Consequently, it is imperative to recognize this imbalance as a significant direction for curriculum development and a limitation of the current assessment system. In the future, the expansion of authentic assessment across all four language skills should be supported by teacher professional development, appropriate assessment instruments, and digital technologies that are capable of facilitating valid, reliable, and communicative evaluation.

Integration of Digital Media in Arabic Language Assessment

In addition to project-based assignments and Google Forms, the teacher also used Quizizz as an assessment tool. *"I have used Quizizz during PH implementation because, in my opinion, it is very helpful, especially since students today are highly interested in using smartphones."* The use of Quizizz represents a concrete implementation of the teaching module, as illustrated in this table.

Table 5. Teaching Module

Practice (Quizizz Media)				
• Teacher's Role: Distribute a Quizizz link or code to students. • Student's Role: Complete the quiz individually.				
Sample Quizizz Question				
No	**Question**	**Option A**	**Option B**	**Option C**
1	أَكْلُ الْفَاكِهَةَ	I eat fruit	I eat bread	I eat vegetables
2	What is the Arabic word for "meat"?	دَجَاجٌ	سَمَكٌ	لَحْمٌ
3	كُلَّ يَوْمٍ	Every moment / time	Every day	Every week
Reflection				
• Teacher & Student: Review and discuss the Quizizz results together. • Summarize the lesson material.				

During the interview, the teacher explained that the use of digital media increased students' engagement in the evaluation process, given that

contemporary students are a generation that has been familiar with digital technology from an early age in their daily lives. Therefore, the use of digital media in learning evaluation was found to be more effective at attracting students' attention and enhancing their participation than conventional forms of assessment alone. This finding is consistent with previous studies demonstrating that gamified platforms such as Quizizz can enhance student motivation, engagement, and formative feedback in language learning contexts[32]. Quizizz, in particular, supports formative assessment by enabling teachers to monitor students' responses in real time and provide immediate feedback, both of which are central principles of assessment for learning[33].

Interestingly, in several sections of the assessment items, the teacher incorporated memes and humorous images as interludes to create a more relaxed, engaging evaluation atmosphere. According to the teacher, this approach was intended to reduce students' tension while completing the assessment and to enhance their engagement throughout the evaluation process.

The use of memes within assessment instruments is particularly noteworthy and may be examined from two competing perspectives. On the one hand, research on the use of humor and memes in educational settings suggests that culturally relevant humorous content can reduce students' anxiety and encourage active participation, thereby fostering a more supportive and engaging learning environment[34]. This is consistent with the affective dimension of language assessment, which recognizes that students' emotional states can influence their performance during evaluation. On the other hand, it is important to distinguish between memes incorporated merely as entertaining interludes and those designed pedagogically to assess specific linguistic competencies. If memes function solely as decorative elements or humorous breaks without a substantive connection to the construct being measured, their inclusion may undermine the content validity of the assessment instrument[35]. In other words, the critical question is not whether memes can be used in assessment, but whether their use is pedagogically purposeful and contributes meaningfully to the measurement of the intended competencies.

---

[32] Shen, "Language Assessment Literacy of Teachers."

[33] Helen Moreira and Martha Lucia Lara Freire, "Promoting Formative Assesment with Quizizz: A Classroom Action Research Study," *Ciencia Latina Revista Científica Multidisciplinar* Vol. 8, No (2024): 590, https://doi.org/10.37811/cl_rcm.v8i2.10511.

[34] Raja Altukruni, "A Systematic Literature Review on the Integration of Internet Memes in EFL/ESL Classrooms," *Arab World English Journal (AWEJ)* 13, no. 4 (2022): 237–50, https://doi.org/10.24093/awej/vol13no4.15; Alevtyna Kalyuzhna, Iryna Lavrinenko, and Olena Radchenko, "Presenting Grammar Through Internet Memes In Teaching English For Specific Purposes," *Cognition, Communication, Discourse,* 2023, 68–82, https://doi.org/10.26565/2218-2926-2023-26-04.

[35] Adiyo Roebianto et al., "Content Validity: Definition And Procedure Of Content Validation In Psycological Research," *Testing, Psychometrics, Methodology in Applied Psychology (TPM)* 30, no. 1 (2023): 5–18, https://doi.org/10.4473/TPM30.1.1.

However, based on the interview findings, the implementation of language-skill assessment at the school had not yet comprehensively encompassed all languages skills. The assessment practices implemented primarily focused on *maharah al-kalam*. At the same time, other skills such as *al-istima'*, *al-qira'ah*, and *al-kitabah* had not been optimally integrated into the learning assessment process, as previously discussed. This condition indicates that the implementation of Arabic language evaluation remained partial and did not fully accommodate balanced integration of the four language skills.

During the evaluation of speaking skills, the teacher acknowledged that the results had not yet been optimal because students' actual abilities did not meet the high expectations set by the foundation.

> *"The main issue here is that the emphasis is placed more on quantity than quality. The foundation expects students, upon graduation, to have memorized the translation of one juz of the Qur'an. However, in practice, when students can translate only ten verses from that juz, they are already considered to have mastered the translation of the entire juz, even though this does not necessarily represent their actual translation ability comprehensively. To address this issue, we as Arabic language teachers require students at each level to memorize and present a certain number of Arabic vocabulary items related to the Qur'an as an effort to fulfill these curriculum targets."*

This statement demonstrates that excessively demanding curriculum targets affect assessment practices, directing them more toward achieving material and vocabulary targets than toward developing deeper language proficiency. In contrast, scholars in communicative language assessment argue that language assessment should ideally measure students' ability to use language authentically and communicatively, rather than merely reproduce linguistic forms. There is broad consensus that authenticity in assessment requires classroom language tasks to replicate real-world language use contexts, as this more accurately reflects learners' actual communicative abilities[36].

Furthermore, the teacher's statement that "translating ten verses is considered representative of mastery of one *juz* reveals an issue of sampling validity in assessment, namely the assumption that a limited sample of language performance can adequately represent a much broader domain of competence. In assessment theory, this condition is referred to as construct underrepresentation, whereby the assessment instrument fails to capture the full breadth of the construct it is intended to measure[37]. Consequently, the scores obtained by students in such a system may not accurately reflect their actual Arabic proficiency.

[36] Linyu Liao, Shelly Xueting Ye, and Jinsong Yang, "A Mini Review of Communicative Language Testing," *Frontiers in Psychology*, 2023, https://doi.org/10.3389/fpsyg.2023.1058411.

[37] S. Messick, *Educational Measurement* (New York: American Council on Education / Macmillan Publishing Company, 1989), https://archive.org/details/educationalmeasu0000unse_3ed.

Despite the demands of the foundation's curriculum and students' varying abilities, the teacher implemented one strategy: dividing Arabic instruction into several levels. At each level, students were required to memorize and submit a specified number of *mufradat* as a prerequisite for advancing to the next level. This strategy reflects adaptive teacher agency in response to institutional pressures, in which the teacher accepts the constraints imposed by the system while simultaneously designing a more manageable, pedagogically controllable mechanism for incremental assessment[38]. The level-based vocabulary submission strategy, although more measurable than claims of mastering an entire *juz*, essentially operates within the same underlying logic: assessing memorization as a proxy for language proficiency. Consequently, the assessment practices implemented served not only as evaluation tools but also as adaptive strategies used by teachers to bridge curriculum demands with students' actual conditions, although opportunities for genuinely communicative assessment remain highly limited within the existing system.

Challenges in Implementing Arabic Language Assessment

In practice, however, teachers continued to encounter various challenges in conducting Arabic language-learning assessments. During the interview, the teacher explained that the primary obstacles were related to students' learning motivation and their basic ability to read *hijaiyyah* letters.

> *"So far, the main challenges are students' low motivation and limited interest in Arabic language subjects. Another issue can be seen in seventh grade, where some students still experience difficulties in reading and pronouncing hijaiyyah letters, because Arabic is inevitably connected to Qur'anic studies. For example, some students still cannot distinguish between certain hijaiyyah letters, confuse vowel markings, and so forth, which automatically affects their performance in Arabic language learning."*

These findings indicate a strong relationship between Arabic language learning and students' Qur'anic reading abilities. This condition is particularly distinctive within the SMPIT context and differs from that of general schools. Deficiencies in foundational Arabic literacy, particularly in recognizing and reading the Arabic script (*hijaiyyah*), have layered consequences: they not only hinder the development of Arabic language proficiency but also affect students' ability to read the Qur'an, which constitutes a central educational objective of the institution. As a result, teachers face a dual challenge in designing assessment practices that accommodate substantial variations in students' foundational literacy skills within the same classroom. Therefore, assessment practices in

[38] Yuhong Jiang, Jia Li, and Qiang Wang, "An Ecological Approach to Understanding University English Teachers ' Professional Agency in Implementing Formative Assessment," *Frontiers in Psychology*, no. September (2022): 1–19, https://doi.org/10.3389/fpsyg.2022.916980.

Arabic language learning within the SMPIT context cannot be separated from students' Qur'anic literacy skills.

In addition, students' low learning motivation also emerged as a factor influencing the effectiveness of learning assessment. In the language assessment literature, students' psychological conditions, including motivation, anxiety, and self-perceptions toward the subject, are widely recognized as affective factors that can significantly influence the validity and reliability of assessment outcomes[39]. Therefore, low student motivation should not be viewed merely as a pedagogical issue separate from assessment. Rather, it constitutes an important contextual factor that must be taken into account when interpreting the results of Arabic language assessment in this SMPIT setting.

Nevertheless, the teacher emphasized that the assessment process could still be conducted optimally through individual assessment of each student. When the researcher asked whether all students received adequate assessment, the teacher responded, "*Yes, because every student is assessed individually.*" This statement reflects the teacher's effort to maintain comprehensive student assessment despite students' limitations and the complexity of Arabic language evaluation.

This tendency was further reinforced by the teacher's perspective on the ideal form of evaluation in Arabic language learning. The teacher stated, "*In my opinion, a written-test evaluation is more ideal because it is more objective.*" This perspective reflects an understanding of educational assessment rooted in a positivist measurement paradigm, which stands in contrast to the communicative assessment paradigm that has long been dominant in the language teaching literature. Within the communicative paradigm, no assessment instrument can be considered entirely "objective," as every assessment design decision embodies particular assumptions about what constitutes language competence[40].

In fact, written tests centered on vocabulary and translation may exhibit construct underrepresentation, as they fail to capture the more complex dimensions of communicative competence. This does not imply that the teacher lacks professional competence. Rather, such assessment choices can be understood as rational responses to contextual constraints, including large class sizes, limited instructional time, and institutional requirements for reporting student achievement to the educational foundation. Studies conducted across various Southeast Asian contexts have similarly found that foreign language

[39] Shaharuddin et al., "The Level of Student Anxiety in Arabic Speaking Skills."

[40] Khomeshwaree Mootoosamy and Vahid Aryadoust, "Transitioning from Communicative Competence to Multimodal and Intercultural Competencies : A Systematic Review," *Societies (MDPI)* 14, no. Cc (2024): 3–5, https://doi.org/10.3390/soc14070115.

teachers tend to prioritize the practicality of assessment in large and demanding classroom settings, even when they are fully aware of its pedagogical limitations[41].

Overall, this study's findings reveal a three-layered structural tension within Arabic language assessment practices in SMPITs. First, there is a tension between the foundation's distinctive curriculum, which emphasizes memorization and translation, and the national curriculum's expectation of communicative language competence. Second, there is a tension between teachers' assessment knowledge, which recognizes the limitations of written testing, and their practical preference for written assessments, driven by considerations of efficiency and feasibility. Third, there is a tension between adopting digital technologies, such as Quizizz and Google Forms, which reflect an awareness of the need for innovation, and the reality that these innovations have not yet been accompanied by the development of more comprehensive assessment constructs that adequately address all language skills. These tensions suggest that the gap between communicative language assessment theory and classroom practice in SMPIT is not primarily a knowledge gap on the part of the teacher, but a structural gap produced by competing institutional demands that simultaneously pull assessment practices in opposing directions.

These tensions are not unique to the SMPIT context. The dominance of written, grammar-translation-based assessment aligns with broader findings in the Arabic language education literature, which consistently document the persistence of traditional assessment forms despite curricular shifts toward communicative approaches[42]. However, this study extends those findings by locating the root cause not in teacher habit or lack of training, but in the institutional curriculum architecture of the educational foundation, a factor that previous studies in the Indonesian context have rarely foregrounded as the primary explanatory variable.

Furthermore, the three-layered tension identified here resonates with findings from integrated Islamic school contexts in Malaysia, where teachers similarly navigate between top-down institutional mandates and communicative competence expectations, producing hybrid assessment practices that reflect negotiated compromise rather than full adherence to either paradigm[43]. This

[41] Ima Fitriyah, Fitriatul Masitoh, and Utami Widiati, "Classroom-Based Language Assessment Literacy and Professional Development Need between Novice and Experienced EFL Teachers" 12, no. 1 (2022): 124–34, https://doi.org/10.17509/ijal.v12i1.46539; Puspawati, Khansa, and Widiati, "Developing EFL Teachers ' Language Assessment Literacy : A Systematic Literature Review on Teacher Training Programs"; Marzaini et al., "A Discovery On Malaysian ESL Teachers' Adherence To Policy Mandates In Classroom-BAsed Assesment Practices."

[42] Pujiani, "Assessment System for Arabic Language Learning within the Independent Curriculum"; Abdillah Mahbubi, "Conventional and Contemporary Arabic Language Teaching Methods : A Comparative Analysis" 6, no. 2 (2024): 205–28, https://doi.org/10.21580/alsina.6.2.23307.

[43] Marzaini et al., "A Discovery On Malaysian ESL Teachers' Adherence To Policy Mandates In Classroom-BAsed Assesment Practices"; Mohammad Taufiq Abdul Ghani et al., "Enhancing Arabic Communication Proficiency in Malaysian among Malaysian Students : Barriers , Pedagogical Strategies ,

cross-contextual resonance suggests that the structural tensions documented in this study are not an anomaly specific to one SMPIT, but may represent a systemic pattern across Islamic integrated educational institutions in Muslim-majority countries.

These findings carry implications at three levels. Theoretically, they contribute to the understanding that the quality of Arabic language assessment in integrated Islamic educational institutions is shaped not only by teachers' individual capacities, but more fundamentally by the institutional curriculum architecture that frames and constrains teachers' opportunities to design and implement assessment. This positions institutional curriculum governance rather than teacher competence as the primary unit of analysis in future research on assessment quality in Islamic education contexts.

At the practical level, the findings suggest that efforts to reform Arabic language assessment in SMPIT should begin with constructive curricular dialogue among foundation administrators, teachers, and curriculum developers, rather than relying solely on individual teacher training initiatives. Specifically, foundations need to revisit the alignment between their distinctive curriculum targets and the demands of communicative language assessment, creating space for teachers to implement more balanced language skills evaluation without compromising institutional goals. At the policy level, the Ministry of Religious Affairs and relevant educational authorities may need to develop clearer guidelines for assessment in SMPIT contexts that acknowledge the dual mandate of language and religious learning, providing practical frameworks for integrating communicative assessment within foundation-specific curriculum structures.

This study is not without limitations. The instrumental case study design, while methodologically justified, involved a single teacher at one institution, which necessarily limits the transferability of findings to other SMPIT contexts with different foundation governance, resource levels, or teacher profiles. Additionally, data collection relied exclusively on interviews and document analysis, without classroom observation, which means that spontaneous formative assessment practices enacted in real instructional moments may not have been fully captured. The absence of student voice also represents a gap: student perceptions of assessment fairness, relevance, and impact were not examined, despite their direct relevance to understanding assessment quality. These limitations provide a clear agenda for future research, as elaborated in the conclusion.

and Environmental Influences," *Edelweiss Applied Science and Technology* 9, no. 7 (2025): 2078–89, https://doi.org/10.55214/2576-8484.v9i7.9093.

CONCLUSION

This study examined Arabic language assessment practices within curriculum implementation at an SMPIT through an instrumental case study involving a single Arabic language teacher, using semi-structured interviews and document analysis. The findings reveal that assessment was conducted in a structured manner through Daily, Mid-Semester, and End-of-Semester Assessments, yet remained predominantly oriented toward written tests emphasizing vocabulary (*mufradat*), grammar (*qawa'id*), and Qur'anic translation. The assessment of the four language skills was not implemented in a balanced manner: while *maharah al-kalam* had begun to be assessed through *hiwar* practice and video-based projects, the assessment of *al-istima'*, productive *al-qira'ah*, and creative *al-kitabah* remained largely absent from the formal assessment system. The teacher demonstrated initiative in incorporating Quizizz and Google Forms into evaluation, yet faced structural challenges including low student motivation, limited *hijaiyyah* literacy, and high institutional curriculum targets.

The central argument of this study is that the quality and orientation of Arabic language assessment at SMPIT are shaped more substantially by institutional curricular architecture than by individual teacher capacity alone. The teacher was not simply choosing to avoid communicative assessment, she was operating within a system that structurally prioritizes Qur'anic vocabulary memorization and translation as the primary markers of Arabic language achievement, producing structural tensions between institutional demands, teacher assessment knowledge, and communicative assessment principles. This pattern resonates with findings from integrated Islamic school contexts in Malaysia, where institutional mandates similarly constrain teachers' capacity to diversify assessment practices, strengthening the argument that reforming Arabic language assessment in Islamic integrated schools requires systemic curricular dialogue among educational foundations, teachers, and curriculum developers, rather than relying solely on isolated teacher-training initiatives.

This study is limited by its single-participant design and the absence of classroom observation and student perspectives, which restricts the transferability of findings. Future research should employ multi-site case study designs across SMPIT institutions with varying foundation governance structures, incorporate observational data to capture spontaneous formative assessment practices, and include student perceptions of assessment relevance and fairness. Comparative studies across Muslim-majority countries, examining how Islamic educational institutions negotiate the tension between religious-linguistic curriculum goals and communicative language assessment would further advance the international literature on Arabic language education.

References

Akkuş, Sara, and Ömer İshakoğlu. "Arapça Öğretiminde Geleneksel - Alternatif Ölçme ve Değerlendirme Yöntemi Kullanımının Bazı Değişkenlere Göre İncelenmesi *." *International Social Sciences Studie Journal* 10 (2024): 0–2. https://doi.org/10.5281/zenodo.125 92780.

Altukruni, Raja. "A Systematic Literature Review on the Integration of Internet Memes in EFL/ESL Classrooms." *Arab World English Journal (AWEJ)* 13, no. 4 (2022): 237–50. https://doi.org/10.24093/awej/vol13no4.15.

Arkan, Zeynep. "Evaluation of Listening Activities in Arabic Language Teaching Textbooks Used in Turkey in Light of the CEFR Criteria." *Asian-Pacific Journal of Second and Foreign Language Education*, 2026. https://doi.org/10.1186/s40862-025-00366-5.

Aryadoust, Vahid. "The Vexing Problem of Validity and the Future of Second Language Assessment." *Language Testing* 40(1) 8 –1 (2023): 8–14. https://doi.org/10.1177/02655322221125204.

Bremane, Anzela Jurane. "The Digital Transformation Of Assessment: Challenges and Opportunities." *Human, Technologies and Quality of Education*, 2021, 352–63. https://doi.org/10.22364/htqe.2021.25.

Brookhart, Susan M. "Educational Assessment Knowledge and Skills for Teachers Revisited," no. 14 (2024): 751. https://doi.org/10.3390/ educsci14070751.

Fitriyah, Ima, Fitriatul Masitoh, and Utami Widiati. "Classroom-Based Language Assessment Literacy and Professional Development Need between Novice and Experienced EFL Teachers" 12, no. 1 (2022): 124–34. https://doi.org/10.17509/ijal.v12i1.46539.

Gebremariam, Hailay Tesfay, and Abate Demissie Gedamu. "The Influence of K-12 Language Teachers ' Demographic Profiles on Authentic Assessment Practice." *Indonesiann Journal Of Applied Linguistics* 13, no. 2 (2023): 443–54. https://doi.org/10.17509/ijal.v13i2.63085.

Ghani, Mohammad Taufiq Abdul, Nor Hadisah Mohamed Shalikin, Saipolbari Ramli, ZM Zakaria, WAAW Daud, AZ Amiruddin, and MA Ahmad. "Enhancing Arabic Communication Proficiency in Malaysian among Malaysian Students : Barriers , Pedagogical Strategies , and Environmental Influences." *Edelweiss Applied Science and Technology* 9, no. 7 (2025): 2078–89. https://doi.org/10.55214/2576-8484.v9i7.9093.

Hakim, Tirto Luqmanul, Abdul Haris, and Mokhammad Miftakhul Huda. "Assessment of Speaking Skills (Maharah Kalam) in Arabic Language Lesson at Madrasah Ibtidaiyah Riyadlus Sholihien Jember" 5, no. 2 (2025): 599–605. https://doi.org/10.58905/jse.v5i2.485 Assessment.

Hasim, Muhamad Fuad, M Baihaqi, Nelly Sakinah, Ulil Irzaqotul Humaidah, and Adi Surya Utama. "Bridging Assessment and Proficiency : Developing Speaking Skill Test Instruments for Arabic Learners" 7, no. 2 (2025): 526–37. https://doi.org/10.21154/zk6nbg28.

Jiang, Yuhong, Jia Li, and Qiang Wang. "An Ecological Approach to Understanding University English Teachers ' Professional Agency in Implementing Formative Assessment." *Frontiers in Psychology*, no. September (2022): 1–19. https://doi.org/10.3389/fpsyg.2022.916980.

Kalyuzhna, Alevtyna, Iryna Lavrinenko, and Olena Radchenko. "Presenting Grammar Through Internet Memes In Teaching English For Specific Purposes." *Cognition, Communication, Discourse,* 2023, 68–82. https://doi.org/10.26565/2218-2926-2023-26-04.

Liao, Linyu, Shelly Xueting Ye, and Jinsong Yang. "A Mini Review of Communicative Language Testing." *Frontiers in Psychology*, 2023. https://doi.org/10.3389/fpsyg.2023.1058411.

Lincoln, Yvonna S., and Egon G. Guba. *Naturalistic Inquiry*. Beverly Hills, CA: SAGE Publications, 1985.

Mahbubi, Abdillah. "Conventional and Contemporary Arabic Language Teaching Methods : A Comparative Analysis" 6, no. 2 (2024): 205–28. https://doi.org/10.21580/alsina.6.2.23307.

Mahmudi, Ihwan, Hanifa Bujanah, Moh Ismail, and Ahmad Saifulloh. "Analysis Of Online Arabic Language Proficiency Test Questions Based On Bloom ' s Cognitive Taxonomy" 8, no. 2 (2025): 4–11. https://doi.org/10.18860 /ijazarabi. V8i2.29595 I.

Marlina, Lina, and Khairun Nisa Hasugian. "Analysis of the Difficulties Faced by Seventh Grade Students of Mts Al-Haq Margahayu in Mastering Listening (Maharah Istima ') and Speaking Skills (Maharah Kalam) in Arabic Language Learning," 2025. https://doi.org/10.30997/dt.v6i1.16232.

Marzaini, Anwar Farhan Mohamad, Shahazwan Mat Yusoff, Noorhayati Zakaria, Mohamad Helme Basal, Wan Nurul Elia, Haslee Sharil, and Kaarthiyany Supramaniam. "A Discovery On Malaysian ESL Teachers' Adherence To Policy Mandates In Classroom-BAsed Assesment Practices." *Malaysian Journal of Learning and Instruction* 1, no. 1 (2024): 249–84. https://doi.org/10.32890/mjli2024.21.1.9.

Masrifah, Khusnul, and Mohammad Ahsanuddin. "Development of HOTS-Based Final Semester Assessment Questions in the Al- ' Ashri Arabic Textbook for 10th-Grade Students of Islamic Senior High School." *Arabiyatuna: Jurnal Bahasa Arab* 9, no. 2 (2025). https://doi.org/10.29240/jba.v9i2.13889.

Messick, S. *Educational Measurement*. New York: American Council on Education / Macmillan Publishing Company, 1989. https://archive.org/details/educationalmeasu0000unse_3ed.

Miles, Matthew B, A Michael Huberman, and Johny Saldana. *Qualitative Data Analysis*. Edited by Helen Salmon Editorial. 3rd ed. United States of America: SAGE Publications, 2014. https://www.metodos.work/wp-content/uploads/2024/01/Qualitative-Data-Analysis.pdf.

Mootoosamy, Khomeshwaree, and Vahid Aryadoust. "Transitioning from Communicative Competence to Multimodal and Intercultural Competencies : A Systematic Review." *Societies (MDPI)* 14, no. Cc (2024): 3–5. https://doi.org/10.3390/soc14070115.

Moreira, Helen, and Martha Lucia Lara Freire. "Promoting Formative Assesment with Quizizz: A Classroom Action Research Study." *Ciencia Latina Revista Científica Multidisciplinar* Vol. 8, No (2024): 590. https://doi.org/10.37811/cl_rcm.v8i2.10511.

Muflihah, Sholehan, and Muhammad Baihaqi. "Integration of Four Language Skills in Arabic Language Learning" 16, no. 1 (2024): 69–87. https://doi.org/10.24042/albayan.v16i1.18381.

Nulloh, Anwar Ridwan, Fiksi Bahruddin Ar Romli, Moch Cecep Abdul Azis, and Muhammad Andhika Silmi. "Arabic Language Learning Model Based On The Merdeka Curriculum" 4, no. 1 (2025): 129–43. https://doi.org/10.15575/ta.v4i1.44922.

Nurohmah, Eka Yuliana, and Siti Ma'rifah. "From Paper-Based to Digital Assessment : Adoption and Challenges of Learning Evaluation Applications in Islamic Education" 3, no. 1 (2025). https://doi.org/10.70376/jerp.Received.

Ponomariovien, Jovita, Daiva Jakavonyte-Staškuviene, and Frederic Torterat. "Implementing Competency-Based Education Through the Personalized Monitoring of Primary Students ' Progress and Assessment," 2025. https://doi.org/10.3390/educsci15020252.

Pujiani, Ujiani. "Assessment System for Arabic Language Learning within the Independent Curriculum" 02, no. 03 (2024): 165–73. https://doi.org/10.18860/kitaba.v2i3.25565.

Puspawati, Indah, Maharani Khansa, and Utami Widiati. "Developing EFL Teachers ' Language Assessment Literacy : A Systematic Literature Review on Teacher Training Programs." *The Electronic Journal for English as a Second Language Developing* 28, no. 2 (2024): 1–21. https://doi.org/10.55593/ej.28110a8.

Roebianto, Adiyo, Setiawati Intan Savitri, Irfan Aulia, Arie Suciyana, and Mubarokah Lailatul. "Content Validity: Definition And Procedure Of Content Validation In Psycological Research." *Testing, Psychometrics, Methodology in Applied Psychology (TPM)* 30, no. 1 (2023): 5–18. https://doi.org/10.4473/TPM30.1.1.

Rohman, Balqis Fauzatul, and Choirul Anam. "Analysis of Digital-Based Elementary School Learning Assessment Applications" 02, no. 01 (2023): 1–15. https://doi.org/10.71392/ejip.v2i1.47.

Salsabyila, Siti Nur, and Muhammad Taufiq. "The Implementation of the Merdeka Curriculum in Arabic Language Education : A Critical Analysis" 8, no. 2 (2025): 175–92. https://doi.org/10.58223/al-irfan.v8i2.535.

Sanjaya, Budi, Wahyu Hidayat, Universitas Islam, Negeri Sulthan, and Thaha Saifuddin. "Evaluasi Keterampilan Berbicara Bahasa Arab Siswa Madrasah Aliyah Di Provinsi Jambi." *Arabi : Journal of Arabic Studies* 6, no. 2 (2021): 220–35. https://doi.org/10.24865/ajas.v6i2.339.

Sembiring, Lina Tri Astuty Beru, and Shafa Salsabila Sembiring. "Practical Application of Assessment Theories for English Teacher through Digital Technologies" 15, no. 2 (2024): 378–87. https://doi.org/10.26877/eternal.v15i2.830.

Shaharuddin, Hanis Najwa, Hazrati Yahaya, Ibrahim Youssef Abdelhamid, Muhammad Hatta Shafri, and Nurul Najwa. "The Level of Student Anxiety in Arabic Speaking Skills." *International Journal Of Research And Innovation In Social Science (IJRISS)* VIII, no. 2454 (2024): 2563–71. https://doi.org/10.47772/IJRISS.

Shen, Bin. "Language Assessment Literacy of Teachers." *Frontiers in Psychology* 13, no. May (2022). https://doi.org/10.3389/fpsyg.2022.864582.

Soliman, Rasha, and Saussan Khalil. "The Teaching of Arabic as a Community Language in the UK." *International Journal of Bilingual Education and Bilingualism ISSN:* 0050 (2024). https://doi.org/10.1080/13670050.2022.2063686.

Sundirah, Sutaman, and Danial Hilmi. "Development Of Interactive Arabic Language

Learning Evaluation Medua Using the Wordwall Platform for Students at MIN 1 Malang" 10, no. 3 (2025): 528–35. https://doi.org/10.18860/abj.v10i3.32979.

Syamsu, Pradi Khusufi, Achmad Satori Ismail, Syukron Kamil, and Muhbib Abdul Wahab. "Scrutinizing Integrative Learning in Arabic Instruction" 02, no. 01 (2023): 107–23. https://doi.org/10.54298/ijith.v2i1.78 Scrutinizing.

Umareni, Unaisah Soehardin, and Eko Ngabdul Shodikin. "Evaluasi Pembelajaran Bahasa Arab Kelas 7 Di Marhalah Salafiyah Wustho Pondok Pesantren Islamic Center Bin Baz Putri Yogyakarta" 02, no. 01 (2024): 27–35. https://doi.org/10.61553/ascent.v2i1.157.

Vassiliou, Skevi, Salomi Papadima Sophocleous, and Christina Nicole Giannikas. *Formative Assessment in Second Language Learning: A Systematic Review and an Annotated Bibliography*. France: Research-publishing.net, 2022. https://doi.org/10.14705/rpnet.2022.60.9782383720133.

Viberg, Olga, Chantal Mutimukwe, Stefan Hrastinski, Teresa Cerratto- Pargman, and Joakim Lilliesköld. "Exploring Teachers ' (Future) Digital Assessment Practices in Higher Education : Instrument and Model Development." *British Journal of Educational Technology*, no. December 2023 (2024): 2597–2616. https://doi.org/10.1111/bjet.13462.

Wahidmurni, Samsul Susilawati, and Achmad Anwar Abidin. "Resistance to Curriculum Changes among Teachers of Madrasah Ibtidaiyah in Indonesia." *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam* 9, no. 1 (2024): 16–27. https://doi.org/10.31538/ndh.v9i1.4300.

Wanti, Adella Ira, A Qurrotul, Dewi Chamidah, Universitas Islam, Negeri Maulana, and Malik Ibrahim. "Model Dan Praktik: Asesmen Formatif Non Paper-Based Dalam Pembelajaran Bahasa Arab." *Arabi : Journal of Arabic Studies* 7, no. 1 (2022): 76–92. https://doi.org/10.24865/ajas.v7i1.419.