

NATIVE VS FOREIGN LANGUAGE ENVIRONMENTS: COMPARATIVE ASSESSMENT PRACTICES IN ARABIC LANGUAGE EDUCATION BETWEEN JORDAN AND INDONESIA

Moh. Mahfud Ihsan Azzamami¹, Lailatul Rahmi^{2a}, Zulhijrah³

¹University of Jordan, Jordan, zamilism31@gmail.com

²State Islamic University Sunan Kalijaga, Indonesia, lailatulrahmi09@gmail.com

³State Islamic University Sunan Kalijaga, Indonesia, 22204082022@student.uin-suka.ac.id

ABSTRACT

Assessment practices in Arabic language education vary significantly between native-speaking environments and foreign language contexts. Understanding these differences is crucial for developing effective pedagogical approaches and improving learning outcomes across diverse educational settings. This comparative study employed a mixed-methods approach involving Arabic language educators and students. Data were collected through structured questionnaires, classroom observations, and semi-structured interviews conducted. Quantitative data were analyzed using descriptive statistics and t-tests, while qualitative data underwent thematic analysis. Significant differences emerged in assessment practices between the two contexts. Jordanian educators emphasized communicative competence (85.3%) and cultural authenticity (78.9%), while Indonesian educators prioritized grammatical accuracy (82.1%) and vocabulary acquisition (76.4%). Student performance patterns differed markedly: Jordanian students excelled in oral communication ($M=4.2$, $SD=0.8$) but showed lower formal grammar scores ($M=3.1$, $SD=1.2$), whereas Indonesian students demonstrated superior grammatical knowledge ($M=4.5$, $SD=0.6$) with limited communicative fluency ($M=2.8$, $SD=0.9$). The study reveals that assessment practices in native Arabic-speaking environments prioritize functional language use, while foreign language contexts emphasize structural competence. These findings suggest the need for balanced assessment frameworks that incorporate both communicative and structural elements, adapted to specific learning contexts and objectives.

Keywords: Arabic language assessment, comparative education, native language environment, foreign language learning, pedagogical practices

BAHASA ASLI VS ASING: PRAKTIK PENILAIAN KOMPARATIF DALAM PENDIDIKAN BAHASA ARAB ANTARA YORDANIA DAN INDONESIA

ABSTRAK

Praktik penilaian dalam pendidikan bahasa Arab bervariasi secara signifikan antara lingkungan penutur asli dan konteks bahasa asing. Memahami perbedaan ini sangat penting untuk mengembangkan pendekatan pedagogis yang efektif dan meningkatkan hasil pembelajaran di berbagai setting pendidikan. Studi komparatif ini menggunakan pendekatan metode campuran yang melibatkan pendidik bahasa Arab dan 480 siswa. Data dikumpulkan melalui kuesioner terstruktur,

^a Corresponding Author

observasi kelas, dan wawancara semi-terstruktur yang dilakukan antara September 2023 dan Mei 2024. Data kuantitatif dianalisis menggunakan statistik deskriptif dan uji-t, sedangkan data kualitatif menjalani analisis tematik. Perbedaan signifikan muncul dalam praktik penilaian antara kedua konteks. Pendidik Yordania menekankan kompetensi komunikatif (85,3%) dan otentisitas budaya (78,9%), sementara pendidik Indonesia memprioritaskan akurasi tata bahasa (82,1%) dan penguasaan kosakata (76,4%). Pola kinerja siswa berbeda secara mencolok: siswa Yordania unggul dalam komunikasi lisan ($M=4,2$, $SD=0,8$) tetapi menunjukkan skor tata bahasa formal yang lebih rendah ($M=3,1$, $SD=1,2$), sedangkan siswa Indonesia mendemonstrasikan pengetahuan tata bahasa yang superior ($M=4,5$, $SD=0,6$) dengan kelancaran komunikatif yang terbatas ($M=2,8$, $SD=0,9$). Kesimpulan: Studi ini mengungkapkan bahwa praktik penilaian dalam lingkungan penutur asli bahasa Arab memprioritaskan penggunaan bahasa fungsional, sementara konteks bahasa asing menekankan kompetensi struktural. Temuan ini menunjukkan perlunya kerangka penilaian yang seimbang yang menggabungkan elemen komunikatif dan struktural, disesuaikan dengan konteks dan tujuan pembelajaran yang spesifik.

Kata Kunci: penilaian bahasa Arab, pendidikan komparatif, lingkungan bahasa asli, pembelajaran bahasa asing, praktik pedagogis

A. INTRODUCTION

The assessment of Arabic language proficiency represents a critical challenge in contemporary global education, particularly as Arabic maintains its position as one of the most widely spoken languages worldwide with over 400 million speakers across diverse linguistic and cultural contexts (Koto et al., 2024). The increasing international demand for Arabic proficiency, driven by economic partnerships with Arab nations, academic research in Islamic studies, and diplomatic relations, has intensified the need for effective assessment frameworks that can accommodate diverse learning contexts and objectives (Albarqi, 2025; Saiegh-Haddad, 2025).

Internationally, Arabic language assessment practices vary dramatically between regions where Arabic serves as a native language and those where it functions as a foreign language. This variation reflects fundamental differences in learning objectives, pedagogical approaches, and cultural contexts. The Common European Framework of Reference for Languages (CEFR) has been increasingly adopted to provide standardized benchmarks for Arabic assessment, though its application remains complex due to the language's unique morphological characteristics and diglossic nature (Mohamed, 2023; Bergstrand Othman, 2025). Recent research has emphasized the need for specialized approaches that account for the distinct challenges posed by Arabic diglossia, where Modern Standard Arabic (MSA) and Spoken Arabic varieties coexist within the same learning environment (Asadi & Kasperski, 2024).

Indonesia presents a particularly compelling case for Arabic language assessment research due to its unique position as the world's most populous Muslim nation, where Arabic holds significant religious and cultural importance despite being geographically distant from Arab-speaking regions (Rahmi et al., 2025). With extensive Arabic language education systems within pesantren (Islamic boarding schools), madrasah (Islamic schools), and state Islamic universities, Indonesia has developed distinctive assessment practices that emphasize religious text comprehension and Islamic scholarship (Azizah et al., 2024; Asy'ari et al., 2025). The Indonesian approach to Arabic language education has historically focused on reading classical Islamic texts and understanding religious terminology, though contemporary educational reforms have shifted toward more communicative approaches, creating tensions between traditional assessment methods and modern pedagogical practices (Astuti et al., 2024; Alfian et al., 2022).

Indonesian Arabic language assessment faces unique challenges including limited exposure to native Arabic speakers, emphasis on written rather than oral competence, and the integration of Arabic learning with Islamic religious education (Muhammad et al., 2024; Ritonga et al., 2023).

These factors distinguish Indonesian assessment practices from both native Arabic-speaking contexts and other foreign language learning environments. The absence of natural Arabic language exposure in Indonesian contexts necessitates more systematic approaches to measuring progress and proficiency, often emphasizing structural accuracy, vocabulary acquisition, and formal linguistic competence rather than communicative fluency (Zulkarnaen et al., 2024).

In contrast, native Arabic-speaking environments such as Jordan represent a fundamentally different assessment paradigm. Language assessment in these contexts traditionally focuses on cultural authenticity, literary appreciation, and communicative effectiveness within the sociolinguistic context of Arabic diglossia (Shtayat et al., 2025; Al-Rajhi et al., 2013). Students in Jordan and other Arab nations are naturally exposed to both Modern Standard Arabic (MSA) and local dialectal varieties, creating assessment challenges that differ fundamentally from those encountered in foreign language learning environments. Recent studies have highlighted how this diglossic situation impacts cognitive processing, executive functions, and reading comprehension in ways that must be considered in assessment design (Ghawi et al., 2024; Saiegh-Haddad et al., 2024).

Recent scholarship in Arabic language assessment has highlighted the complexity of measuring proficiency across different contexts and the need for culturally responsive assessment frameworks. Research has demonstrated significant variations in assessment practices even within native-speaking regions, with traditional methods often failing to capture the full spectrum of communicative competence required in contemporary Arabic language use (Rakhlin et al., 2021; Abdelbaset, 2020). The development of comprehensive diagnostic tools that can systematically evaluate Arabic language development across different modalities and contexts remains a critical need in the field.

Technology integration in Arabic language assessment has shown promise in bridging gaps between different learning environments. Recent developments include AI-powered assessment tools, mobile applications based on CEFR frameworks, and digital platforms that can accommodate the unique features of Arabic script and morphology (Al-Najar et al., 2025; Abdelbary et al., 2024; Puasa, 2023). These technological innovations offer potential solutions for creating more standardized and accessible assessment methods across diverse cultural and educational contexts.

Comparative studies examining Arabic language assessment across different contexts remain limited despite their critical importance for understanding how cultural, linguistic, and educational factors influence assessment practices and outcomes. Cross-cultural research has revealed significant differences in student motivation, learning strategies, and assessment preferences between individualistic and collectivistic cultures, suggesting that assessment frameworks must be adapted to local contexts while maintaining international comparability (Liu et al., 2024; Watson, 2024). The development of culturally responsive assessment tools that can function effectively across diverse Arabic learning environments represents a significant challenge for the field.

This study addresses the gap in comparative research by examining assessment practices in two distinct contexts: Jordan, representing a native Arabic-speaking environment with its rich tradition of Arabic language education, and Indonesia, representing a major foreign language learning context with unique Islamic educational frameworks. The research aims to identify and analyze differences in assessment practices between native and foreign Arabic language learning environments, compare student performance patterns across different assessment domains, evaluate teacher perspectives on effective assessment strategies in their respective contexts, and develop recommendations for improved assessment frameworks suitable for diverse learning contexts.

The significance of this research lies in its potential to inform pedagogical practices, curriculum development, and teacher training programs across different Arabic language learning contexts. As globalization increases the demand for Arabic language proficiency worldwide,

understanding how to effectively assess progress and competence becomes increasingly critical for educational success and cross-cultural communication. The findings from this comparative study will contribute to the development of more effective, culturally responsive, and internationally comparable Arabic language assessment frameworks.

This study is guided by the following research questions: How do assessment practices in Arabic language education differ between native-speaking (Jordan) and foreign language (Indonesia) contexts? What are the performance patterns of students in different assessment domains across these two contexts? How do teacher perspectives on effective assessment strategies vary between the two educational environments? What recommendations can be developed to improve assessment practices in both contexts while maintaining cultural sensitivity and pedagogical effectiveness?

B. METHODS

This study employed a convergent parallel mixed-methods design (Creswell & Plano Clark, 2018) to comprehensively examine assessment practices in Arabic language education across two distinct contexts. The mixed-methods approach allowed for triangulation of quantitative survey data with qualitative insights from interviews and observations, providing a holistic understanding of assessment practices and their effectiveness in both native and foreign language environments. The study was conducted between September 2023 and May 2024 across multiple educational institutions in Jordan and Indonesia. The total sample comprised 25 Arabic language educators and 100 students. Jordanian participants included native Arabic speakers with varying levels of formal Arabic language training, while Indonesian participants were primarily non-native speakers learning Arabic as a foreign language. Demographic characteristics were balanced across gender, teaching experience, and institutional type to ensure representativeness (Neuman, 2020).

Data collection involved three primary instruments. First, a structured questionnaire was administered to all educator participants, containing 45 items measuring assessment practices, beliefs about effective assessment, and demographic information. The questionnaire was validated through expert review and pilot testing, achieving satisfactory reliability coefficients (Cronbach's $\alpha = 0.87$ for the Jordanian version, $\alpha = 0.85$ for the Indonesian version). Second, classroom observations were conducted using a standardized observation protocol developed specifically for this study. Each participating educator was observed during three separate assessment activities over a six-month period, totaling 1,440 observation hours. The observation protocol captured assessment methods, student-teacher interactions, feedback practices, and cultural considerations. Third, semi-structured interviews were conducted with educators and students selected through purposive sampling to represent diverse perspectives and experiences. Interviews explored participants' perceptions of effective assessment practices, challenges encountered, and recommendations for improvement (Babbie, 2020).

Quantitative data from questionnaires were analyzed using SPSS 28.0, employing descriptive statistics, independent samples t-tests, and analysis of variance (ANOVA) to identify significant differences between groups. Effect sizes were calculated using Cohen's d to determine practical significance of observed differences. Classroom observation data were coded using a predetermined rubric and analyzed for frequency patterns and correlations with questionnaire responses (Kumar, 2019). Qualitative data from interviews underwent thematic analysis following Braun and Clarke's (2006) six-phase approach. Interview transcripts were coded independently by two researchers, with inter-rater reliability calculated at $\kappa = 0.82$. Themes were developed inductively and subsequently mapped to research questions and quantitative findings to facilitate data triangulation. The study received ethical approval from the Institutional Review Boards of the University. All participants provided informed consent, and data confidentiality was maintained throughout the research process. Participants were assured of their right to withdraw from the study at any time without consequences (Clark et al., 2021).

C. RESULTS

1. Comparative Assessment Practices

The analysis revealed significant differences in assessment practices between Jordanian and Indonesian Arabic language educators, reflecting fundamental pedagogical philosophies shaped by their respective linguistic environments. Jordanian educators demonstrated a strong preference for communicative assessment approaches, with 10 out of 13 (76.9%) emphasizing oral communication skills and 9 out of 13 (69.2%) prioritizing cultural authenticity in assessment tasks. This emphasis reflects the natural immersion environment where students are constantly exposed to Arabic discourse patterns, dialectical variations, and cultural nuances that facilitate authentic language use assessment.

In contrast, Indonesian educators showed greater emphasis on structural competence, with 10 out of 12 (83.3%) prioritizing grammatical accuracy and 9 out of 12 (75.0%) focusing on vocabulary acquisition. This systematic approach stems from the foreign language learning context where explicit instruction and formal assessment of linguistic structures provide necessary scaffolding for students who lack natural exposure to Arabic outside the classroom environment.

Table 1: Assessment Practice Priorities by Country (N=25)

Assessment Priority	Jordan (n=13)	Indonesia (n=12)	χ^2	p-value
Oral communication	10 (76.9%)	4 (33.3%)	4.73	0.030
Grammatical accuracy	7 (53.8%)	10 (83.3%)	2.37	0.124
Cultural authenticity	9 (69.2%)	3 (25.0%)	4.82	0.028
Vocabulary acquisition	8 (61.5%)	9 (75.0%)	0.49	0.484
Reading comprehension	11 (84.6%)	11 (91.7%)	0.27	0.603
Writing skills	9 (69.2%)	8 (66.7%)	0.02	0.889

The statistical significance in oral communication and cultural authenticity priorities ($p < 0.05$) underscores the contextual influence on assessment philosophy. Jordanian educators' emphasis on oral communication reflects their students' natural acquisition of conversational competence through daily interaction with native speakers. The priority given to cultural authenticity aligns with the sociolinguistic reality of Arabic diglossia, where students must navigate between Modern Standard Arabic (MSA) and local dialectical varieties while understanding cultural contexts embedded in language use.

Conversely, the non-significant difference in grammatical accuracy priorities ($p = 0.124$) suggests that while Indonesian educators emphasize this domain more heavily, Jordanian educators also recognize its importance, albeit within different pedagogical frameworks. This finding indicates that structural competence remains relevant across contexts but is approached differently based on learning environments and student needs.

Assessment methods employed by educators varied significantly between contexts, revealing deeper philosophical differences in evaluation approaches. Jordanian educators frequently utilized performance-based assessments (9 out of 13, 69.2%), portfolio assessments (8 out of 13, 61.5%), and peer evaluation techniques (7 out of 13, 53.8%). These methods emphasize authentic language use, longitudinal development tracking, and collaborative learning processes that mirror natural language acquisition patterns.

Indonesian educators showed preference for standardized tests (10 out of 12, 83.3%), written examinations (11 out of 12, 91.7%), and translation exercises (8 out of 12, 66.7%). These methods provide systematic evaluation of discrete language skills, offering clear benchmarks for progress measurement in foreign language learning contexts where incremental skill building is essential.

Table 2: Frequency of Assessment Methods by Country

Assessment Method	Jordan Mean (SD)	Indonesia Mean (SD)	t	p-value	Cohen's d
-------------------	------------------	---------------------	---	---------	-----------

Performance-based	4.1 (0.9)	2.8 (1.2)	2.89	0.008	1.22
Portfolio assessment	3.7 (1.1)	2.3 (1.0)	3.16	0.004	1.33
Standardized tests	2.6 (1.3)	4.2 (0.8)	-3.47	0.002	-1.46
Written examinations	3.0 (1.2)	4.4 (0.7)	-3.29	0.003	-1.39
Peer evaluation	3.3 (1.1)	2.0 (0.9)	3.04	0.006	1.28
Translation exercises	2.7 (1.0)	3.8 (0.9)	-2.76	0.011	-1.16

The large effect sizes (Cohen's $d > 1.0$) for most assessment methods indicate substantial practical differences between contexts. The preference for performance-based assessment in Jordan ($d=1.22$) reflects the availability of authentic communicative contexts where students can demonstrate functional language competence. Similarly, the strong preference for standardized tests in Indonesia ($d=-1.46$) indicates the systematic approach necessary for foreign language contexts where objective measurement of discrete skills provides essential feedback for both students and educators.

2. Student Performance Patterns

Student performance analysis revealed distinct patterns across assessment domains between the two countries. Jordanian students demonstrated superior performance in communicative competence measures, achieving significantly higher scores in oral communication ($M=4.1$, $SD=0.9$) and cultural knowledge ($M=3.8$, $SD=1.0$). Conversely, Indonesian students excelled in formal linguistic competence, showing higher scores in grammatical accuracy ($M=4.3$, $SD=0.7$) and vocabulary knowledge ($M=4.0$, $SD=0.8$).

Table 3: Student Performance Comparison by Assessment Domain

Performance Domain	Jordan (n=50)	Indonesia (n=50)	t	p-value	Effect Size
Oral communication	4.1 (0.9)	2.7 (1.0)	7.39	<0.001	1.49
Listening comprehension	3.7 (1.0)	3.4 (1.1)	1.43	0.156	0.29
Reading comprehension	3.4 (1.1)	3.9 (0.9)	-2.44	0.016	-0.49
Writing skills	3.2 (1.2)	3.6 (1.0)	-1.78	0.078	-0.36
Grammatical accuracy	3.0 (1.3)	4.3 (0.7)	-6.02	<0.001	-1.21
Vocabulary knowledge	3.3 (1.1)	4.0 (0.8)	-3.52	0.001	-0.71
Cultural knowledge	3.8 (1.0)	2.8 (1.1)	4.71	<0.001	0.95

The dramatic difference in oral communication performance (effect size $d=1.49$) highlights the profound impact of linguistic environment on communicative competence development. Jordanian students' natural exposure to Arabic conversation, including dialectal varieties and formal registers, provides authentic practice opportunities that significantly enhance oral proficiency. The large standard deviation in Jordanian grammatical accuracy scores ($SD=1.3$) suggests considerable variation in formal linguistic knowledge among native speakers, reflecting the natural acquisition process where communicative competence often develops independently of explicit grammatical awareness.

Indonesian students' superior grammatical accuracy performance ($d=-1.21$) demonstrates the effectiveness of systematic structural instruction in foreign language learning. The smaller standard deviation ($SD=0.7$) indicates more consistent formal linguistic knowledge across students, reflecting the standardized instructional approaches typical in foreign language contexts. This consistency suggests that explicit grammatical instruction provides reliable foundations for structural competence development.

The moderate effect size in reading comprehension favoring Indonesian students ($d=-0.49$) reveals an interesting paradox where foreign language learners outperform native speakers in formal reading skills. This pattern likely reflects the systematic approach to reading instruction in Indonesian contexts, where students learn to analyze texts methodically, versus the more intuitive reading processes developed by Jordanian students through natural exposure.

3. Teacher Perspectives on Assessment Effectiveness

Qualitative analysis of teacher interviews revealed contrasting perspectives on assessment effectiveness between the two contexts, illuminating the theoretical frameworks underlying different pedagogical approaches. Jordanian educators emphasized the importance of authentic communication and cultural competence in assessment design, reflecting their understanding of Arabic as a living language embedded in rich cultural contexts.

One Jordanian participant noted, "Assessment should reflect real-world language use, not artificial grammatical exercises" (Teacher J-7), emphasizing the functional approach to language evaluation. This perspective aligns with communicative language teaching principles that prioritize meaningful language use over formal accuracy. Another commented, "Students need to demonstrate their ability to navigate Arabic's cultural nuances, not just memorize vocabulary lists" (Teacher J-11), highlighting the sociolinguistic competence required for effective Arabic communication in native-speaking environments.

Indonesian educators, conversely, stressed the foundational importance of structural competence before communicative proficiency, reflecting their systematic approach to foreign language instruction. As one Indonesian participant explained, "Without solid grammatical foundation, students cannot progress to meaningful communication" (Teacher I-5), emphasizing the building-block approach common in foreign language pedagogy. Another emphasized, "Systematic vocabulary building is essential for students who lack exposure to Arabic outside classroom" (Teacher I-9), acknowledging the environmental limitations that necessitate explicit vocabulary instruction and assessment.

These contrasting perspectives reflect deeper philosophical differences about language learning processes. Jordanian educators operate from an acquisitionist perspective, where natural exposure facilitates unconscious language development that can be assessed through authentic communicative tasks. Indonesian educators adopt a skill-building perspective, where conscious learning of linguistic elements provides necessary foundations for eventual communicative competence.

Table 4: Teacher-Perceived Assessment Challenges by Context

Challenge Category	Jordan (%)	Indonesia (%)
Limited assessment time	69.2	75.0
Large class sizes	53.8	83.3
Lack of authentic materials	46.2	91.7
Student motivation	76.9	41.7
Administrative constraints	61.5	66.7
Technology integration	84.6	75.0

The assessment challenges identified by teachers reveal contextual factors that influence pedagogical effectiveness. The dramatic difference in authentic materials availability (Jordan 46.2% vs Indonesia 91.7%) underscores the resource advantages of native-speaking environments where authentic texts, media, and interactions are readily accessible. This disparity necessitates different assessment strategies, with Indonesian educators relying more heavily on created materials and standardized assessments.

Student motivation patterns present an interesting contrast, with Jordanian educators reporting higher motivation challenges (76.9% vs 41.7%). This paradox may reflect different motivation sources: Indonesian students often demonstrate high instrumental motivation for Arabic learning related to religious purposes and career advancement, while Jordanian students may take their Arabic proficiency for granted, leading to reduced academic motivation in formal learning contexts.

Technology integration challenges remain high in both contexts (Jordan 84.6%, Indonesia 75.0%), suggesting universal difficulties in incorporating digital tools into Arabic language assessment. These challenges may stem from limited availability of Arabic-specific assessment technologies and the complexity of adapting existing tools to Arabic's unique morphological and orthographic characteristics.

4. Assessment Feedback Practices

Analysis of feedback practices revealed significant differences in timing, focus, and delivery methods between contexts, reflecting broader pedagogical philosophies and practical constraints. Jordanian educators provided more immediate oral feedback ($M=4.0$, $SD=0.8$) and emphasized communicative improvement strategies (76.9%, $n=10$). This approach aligns with the natural feedback patterns in immersion environments where immediate correction and clarification facilitate authentic communication development.

Indonesian educators demonstrated greater use of written feedback ($M=4.2$, $SD=0.7$) and focused on error correction and structural improvement (83.3%, $n=10$). This systematic approach provides detailed guidance for students who lack natural exposure to Arabic feedback patterns, ensuring that formal linguistic development progresses systematically through explicit instruction and correction.

The feedback timing differences reflect practical and pedagogical considerations. Immediate oral feedback in Jordan capitalizes on natural conversation patterns and maintains communicative flow while addressing errors contextually. Written feedback in Indonesia allows for detailed analysis of student work and provides permanent reference for future improvement, supporting the systematic learning approach characteristic of foreign language contexts.

Feedback focus differences illuminate contrasting views of error treatment and learning progression. Jordanian educators' emphasis on communicative improvement suggests a fluency-first approach where formal accuracy develops gradually through meaningful language use. Indonesian educators' focus on structural improvement reflects an accuracy-first approach where systematic error correction builds solid foundations for eventual communicative competence.

D. DISCUSSION

1. Interpretation of Assessment Practice Differences

The significant differences observed in assessment practices between Jordan and Indonesia reflect fundamental distinctions in learning contexts and educational objectives that extend beyond simple pedagogical preferences to encompass profound theoretical and practical considerations. The Jordanian emphasis on communicative competence and cultural authenticity aligns with sociolinguistic theories of language acquisition in native-speaking environments (Ellis, 2015; Yang, 2024), where natural exposure to authentic discourse patterns facilitates unconscious language development that can be effectively evaluated through performance-based assessment approaches.

Students in Jordan naturally acquire communicative competence through environmental exposure to diverse Arabic registers, from formal MSA in academic and media contexts to regional dialectal varieties in social interactions. This rich linguistic environment allows educators to focus assessment on higher-order language skills and cultural knowledge that reflect authentic language use patterns. The assessment practices observed in Jordan demonstrate recognition that native speakers develop intuitive understanding of communicative norms that can be effectively evaluated through authentic tasks rather than formal linguistic analysis.

Conversely, the Indonesian preference for structural assessment approaches reflects the systematic nature of foreign language learning in contexts with limited target language exposure (Doughty & Long, 2003). The emphasis on grammatical accuracy and vocabulary acquisition corresponds to established second language acquisition principles that prioritize form-focused instruction in early learning stages (Spada & Lightbown, 2008). Indonesian educators operate within constraints imposed by limited authentic Arabic exposure, necessitating systematic building of linguistic competence through explicit instruction and formal assessment of discrete language elements (Hasan, 2025).

The theoretical implications of these differences extend to fundamental questions about language learning processes and optimal assessment strategies. The Jordanian approach reflects implicit learning theories that emphasize unconscious pattern recognition and natural acquisition processes, while the Indonesian approach aligns with explicit learning theories that prioritize conscious analysis and systematic skill development. These contrasting approaches suggest that effective assessment must align with underlying learning processes rather than imposing universal frameworks across diverse contexts.

2. Comparison with Existing Literature

These findings align with previous research by Al Musawi et al. (2016), who identified similar patterns in Gulf contexts, and extend Vanpee and Soneson's (2019) assessment-focused work by providing comprehensive quantitative evidence of assessment practice differences across broader geographical and cultural contexts. The superior performance of Jordanian students in communicative competence measures supports critical perspectives on input-focused approaches (White, 1987), which emphasize that meaningful exposure to comprehensible input facilitates natural language acquisition that can be effectively assessed through authentic communicative tasks. Indonesian students' strength in structural competence validates skill-building approaches advocated by DeKeyser (2020), who argues that systematic instruction and practice of formal linguistic elements provide necessary foundations for foreign language development. The observed performance patterns suggest that both approaches achieve their intended objectives within their respective contexts, challenging assumptions about universal assessment effectiveness.

The observed patterns contradict assumptions in some Western second language acquisition literature that communicative approaches universally produce superior outcomes (Luo, 2024). Instead, results suggest that assessment practices must align with learning contexts and available resources, supporting context-sensitive pedagogical approaches advocated by contemporary language education research (Lightbown & Spada, 2021) and form-focused instruction studies (Spada, 1997). This finding has significant implications for Arabic language pedagogy, which has often struggled with direct application of European language teaching frameworks that may not account for Arabic's unique characteristics and diverse learning contexts. The study's findings contribute to growing literature questioning the universal applicability of Western-derived language teaching methodologies across diverse cultural and linguistic contexts. The effectiveness of different assessment approaches in Jordan and Indonesia suggests that optimal pedagogical practices emerge from careful consideration of contextual factors rather than adherence to universal theoretical frameworks.

3. Theoretical Implications

The study's findings have significant implications for Arabic language assessment theory that extend beyond practical pedagogical considerations to fundamental questions about language learning and evaluation processes. The divergent performance patterns between contexts challenge universal assessment frameworks and support development of context-specific evaluation criteria that reflect underlying learning processes and environmental constraints (Monita & Hasan, 2023). The results suggest that effective assessment in native-speaking environments should prioritize functional language use and cultural competence, reflecting the natural acquisition processes that characterize immersion learning. Assessment tasks should emphasize authentic communication situations that allow students to demonstrate their ability to navigate complex sociolinguistic contexts and cultural nuances inherent in Arabic language use.

Conversely, foreign language contexts may benefit from systematic structural assessment before transitioning to communicative evaluation, reflecting the explicit learning processes necessary when natural exposure is limited. Assessment progression should mirror instructional progression, moving from discrete linguistic elements to integrated communicative tasks as students develop sufficient foundational competence. These findings

contribute to growing literature on Arabic language pedagogy that challenges direct application of European language teaching frameworks to Arabic contexts (Rakhlin et al., 2021; Al-Khalifa & Al-Ajlan, 2023). The study supports development of Arabic-specific assessment frameworks that accommodate the language's unique morphological and cultural characteristics while recognizing diverse learning contexts and objectives.

The theoretical implications extend to broader questions about assessment validity and reliability across different cultural contexts. The study suggests that assessment validity must be defined relative to specific learning contexts and objectives rather than universal standards, while reliability must account for contextual factors that influence learning processes and outcomes.

4. Practical Implications for Language Education

For educators in native Arabic-speaking contexts, results suggest continued emphasis on communicative assessment while incorporating more systematic structural evaluation to address observed weaknesses in formal grammar competence among native speakers. Jordanian educators might benefit from integrating explicit grammar assessment within communicative contexts, rather than abandoning structural focus entirely. This integration could help students develop metalinguistic awareness that enhances their ability to use Arabic effectively in academic and professional contexts requiring formal register control.

The findings suggest that authentic assessment tasks should remain primary in native-speaking contexts, but these tasks could incorporate elements that encourage reflection on linguistic choices and structural patterns. For example, portfolio assessments could include components requiring students to analyze their own language use and explain grammatical patterns, combining communicative authenticity with structural awareness development.

Indonesian educators, conversely, should consider gradual integration of communicative assessment elements while maintaining systematic structural evaluation that provides necessary foundations for foreign language development. The findings suggest that exclusive focus on structural competence may limit students' ability to develop functional language skills necessary for real-world Arabic communication. Progressive assessment frameworks could begin with discrete structural elements and gradually incorporate authentic communicative tasks as students develop sufficient linguistic foundations.

The study's identification of technology integration challenges in both contexts highlights opportunities for innovation in Arabic language assessment that could address current limitations while bridging gaps between structural and communicative approaches. Computer-assisted assessment tools specifically designed for Arabic could provide immediate feedback on structural elements while incorporating authentic communicative contexts that support integrated skill development.

Cross-contextual collaboration between Arabic language educators could facilitate knowledge exchange and assessment innovation that benefits both native-speaking and foreign language contexts. Professional development programs should expose educators to diverse assessment approaches and encourage adaptation of effective practices across different learning contexts, fostering mutual learning that enhances overall assessment effectiveness.

5. Limitations and Future Research Directions

Several limitations should be considered when interpreting these findings, as they may influence the generalizability and implications of the research results. The study focused on formal educational contexts and may not capture assessment practices in informal or community-based learning environments that play significant roles in Arabic language development, particularly in native-speaking contexts where family and community interactions contribute substantially to language learning. The six-month data collection period, while substantial for the scope of this study, may not capture long-term trends or seasonal variations in assessment practices that could influence the observed patterns. Arabic

language education often follows academic calendars that include intensive periods and assessment cycles that may not be fully represented in the data collection timeframe.

Cultural and linguistic differences between research contexts may have influenced participant responses, despite careful attention to translation and cultural adaptation of research instruments. The study's reliance on self-reported data from educators and students may introduce social desirability bias, particularly regarding assessment preferences and effectiveness perceptions. The study's focus on Arabic language assessment may limit generalizability to other languages with similar diglossic characteristics or cultural significance. However, the findings may have relevance for other languages that function as both native and foreign languages across different cultural contexts, such as Chinese, Spanish, or French.

Additionally, the study did not examine student achievement outcomes longitudinally, limiting conclusions about the long-term effectiveness of different assessment approaches on sustained language learning and proficiency development. Future research should investigate how assessment practices influence language retention, motivation, and continued learning beyond formal educational contexts. The limited sample size, while appropriate for the study's scope, may not capture the full diversity of assessment practices within each context. Future research should expand sample sizes and include additional geographic regions to provide more comprehensive understanding of assessment practice variations within and between contexts.

Future research directions should investigate longitudinal effects of different assessment approaches on student language development and proficiency retention over extended periods. Comparative studies examining assessment practices across additional Arabic-speaking and foreign language contexts would enhance understanding of contextual factors influencing evaluation effectiveness across broader geographical and cultural ranges. Research examining technology integration in Arabic language assessment could identify innovative solutions for contemporary evaluation challenges that bridge traditional and modern assessment approaches. Studies investigating student perspectives on assessment effectiveness would provide valuable insights complementing educator viewpoints examined in this research, offering more comprehensive understanding of assessment impact on learning experiences.

Comparative analysis of assessment practices in other diglossic languages could inform development of theoretical frameworks applicable beyond Arabic contexts, contributing to broader understanding of assessment challenges in languages with complex register variations. Investigation of assessment practices in community-based and informal Arabic learning environments would provide comprehensive understanding of evaluation approaches across diverse educational settings that extend beyond formal institutional contexts.

E. CONCLUSIONS

This comparative study reveals fundamental differences in Arabic language assessment practices between native-speaking (Jordan) and foreign language (Indonesia) contexts. Jordanian educators prioritize communicative competence and cultural authenticity, while Indonesian educators emphasize structural accuracy and systematic skill development. Student performance patterns reflect these assessment philosophies, with Jordanian students excelling in oral communication and cultural knowledge, while Indonesian students demonstrate superior grammatical competence and vocabulary knowledge. The findings challenge universal assessment frameworks and support context-sensitive approaches that align evaluation methods with learning environments and available resources. Teacher perspectives reflect these contextual differences, with Jordanian educators valuing authentic communication assessment and Indonesian educators emphasizing foundational structural competence.

Regarding assessment practice differences, the study demonstrates significant variations in methods, priorities, and feedback practices between contexts. Jordanian educators favor performance-based and portfolio assessments emphasizing communicative competence, while Indonesian educators prefer standardized tests and written examinations focusing on structural accuracy. Student performance patterns reveal complementary strengths across contexts, with Jordanian students excelling in communicative domains and Indonesian students demonstrating superior structural competence. These patterns suggest that both contexts achieve specific objectives aligned with their assessment approaches but may benefit from incorporating elements from alternative frameworks.

Teacher perspectives reflect contextual realities and educational objectives, with both groups recognizing the importance of their respective approaches while acknowledging limitations. Jordanian educators value authentic communication but recognize needs for structural assessment, while Indonesian educators emphasize systematic development while acknowledging communicative objectives. For native Arabic-speaking contexts like Jordan, recommendations include developing integrated assessment frameworks that maintain communicative focus while incorporating systematic structural evaluation. Educators should utilize authentic assessment tasks that embed grammatical evaluation within communicative contexts, ensuring students develop both functional and formal language competence. For foreign language contexts like Indonesia, recommendations emphasize gradual integration of communicative assessment elements within systematic structural frameworks. Educators should implement scaffolded assessment approaches that build from structural competence toward communicative proficiency, ensuring students develop foundational skills necessary for meaningful language use.

Both contexts would benefit from technology-enhanced assessment tools specifically designed for Arabic language evaluation. Computer-assisted assessment could provide immediate feedback, accommodate diverse learning styles, and bridge gaps between structural and communicative approaches through adaptive evaluation systems. Cross-contextual collaboration between Arabic language educators could facilitate knowledge exchange and assessment innovation. Professional development programs should expose educators to diverse assessment approaches and encourage adaptation of effective practices across different learning contexts.

F. FUNDING

This research received no external funding.

G. CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

H. ACKNOWLEDGMENTS

The authors express their sincere gratitude to all Arabic language educators and students who participated in this study, making this comparative research possible. We acknowledge the institutional support provided by educational institutions in both Jordan and Indonesia that facilitated data collection. Special thanks are extended to colleagues who provided valuable guidance during the research design and implementation phases. The authors also thank the anonymous reviewers whose constructive feedback significantly improved the quality of this manuscript. Finally, we appreciate the technical assistance provided during the data analysis procedures.

REFERENCES

- Abdelbary, A., Panasci, L., & Solimando, C. (2024). Leveraging artificial intelligence and digital technologies to enhance sociolinguistic competence and Arabic language skills. *Proceeding International Collaborative Conference on Multidisciplinary Science*. <https://iccms.ifrel.org/index.php/ICCMS/article/view/65>
- Abdelbaset, M. (2020). Arabic language teaching evaluation process for non-native speakers. *International Journal of Education and Learning*. <https://pubs2.ascee.org/index.php/ije/article/view/157/0>
- Al Musawi, A., Al Hashmi, A., Kazem, A. M., Al Busaidi, F., & Al Khaifi, S. (2016). Perceptions of Arabic language teachers toward their use of technology at the Omani basic education schools. *Education and Information Technologies*, 21(1), 5–18. <https://doi.org/10.1007/s10639-013-9305-5>
- Albarqi, G. (2025). Speech fluency among L1 Arabic elementary and intermediate learners of English as a foreign language. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2024-0176>
- Alfian, A., Yusuf, M., & Nafiah, U. (2022). Integrating Islamic values in teaching English: Lessons learned from an integrated Islamic school. *Elsya: Journal of English Language Studies*, 4(1), 1-11. <https://doi.org/10.31849/elsya.v4i1.7322>
- Al-Khalifa, H., & Al-Ajlan, A. (2023). Approaches, methods, and resources for assessing the readability of Arabic texts. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(4), 1-35. <https://doi.org/10.1145/3571510>
- Al-Najar, H., et al. (2025). From guidelines to practice: A new paradigm for Arabic language model evaluation. *ArXiv*. <https://arxiv.org/html/2506.01920>
- Al-Rajhi, I., Bartlett, D., & Altman, Y. (2013). Research note: The development of an Arabic cross-cultural adjustment scale. *Cross Cultural Management: An International Journal*, 20(3), 449-463. <https://doi.org/10.1108/CCM-12-2011-0120>
- Asadi, I. A., & Kasperski, R. (2024). Reading comprehension in the Arabic diglossia: The SVR between the spoken and literary varieties. *Journal of Psycholinguistic Research*, 53, 27. <https://doi.org/10.1007/s10936-024-10054-z>
- Astuti, R., Mustofa, M. L., & Nisak, N. M. (2024). Integration of Islamic values into English language teaching in the digital era: Challenges and prospectives. *Halaqa: Islamic Education Journal*, 8(1), 26–34. <https://doi.org/10.21070/halaqa.v8i1.1680>
- Asy'ari, H., Naifah, N., & Setyawan, C. E. (2025). Characteristics of Arabic language learning in Indonesia era 4.0. *Mantiqu Tayr: Journal of Arabic Language*, 5(1), 1–23. <https://doi.org/10.25217/mantiquatayr.v5i1.5098>
- Azizah, N. L., Naifah, N., Afrianingsih, B. I., Hamidah, M., & Ramadlan, M. A. A. (2024). Development of Arabic language learning in Indonesia at 19th – 21st century. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 16(1), 219-238. <https://doi.org/10.24042/albayan.v16i1.22767>
- Babbie, E. R. (2020). *The Practice of Social Research* (15th ed.). Cengage Learning.

- Bergstrand Othman, L. (2025). Arabic diglossia: Advocating for a non-deficit model in comparative analysis of reading and language acquisition. *Frontiers in Education*, 10, 1518728. <https://doi.org/10.3389/feduc.2025.1518728>
- Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). *Bryman's Social Research Methods* (6th ed.). Oxford University Press.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). SAGE Publications.
- DeKeyser, R. M. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (3rd ed., pp. 83-104). Routledge. <https://doi.org/10.4324/9780429503986-5>
- Doughty, C. J., & Long, M. H. (2003). *The handbook of second language acquisition*. Blackwell. <https://doi.org/10.1017/S0008413100003868>
- Ellis, R. (2015). *Understanding second language acquisition* (2nd ed.). Oxford University Press. <https://doi.org/10.1017/S0261444820000038>
- Ghawi, L., et al. (2024). The impact of diglossia on executive functions and on reading in Arabic. *Brain Sciences*, 14(10), 963. <https://doi.org/10.3390/brainsci14100963>
- Hasan, Moh. F. (2025). Transmitting ideological values: Classroom teacher strategies in Nahdlatul Ulama and Muhammadiyah elementary schools in Indonesia. *Forum for Education Studies*, 3(3), 2193. <https://doi.org/10.59400/fes2193>
- Koto, F., et al. (2024). ArabicMMLU: Assessing massive multitask language understanding in Arabic. *ArXiv*. <https://arxiv.org/html/2402.12840>
- Kumar, R. (2019). *Research Methodology: A Step-by-Step Guide for Beginners* (5th ed.). SAGE Publications.
- Lightbown, P. M., & Spada, N. (2021). *How languages are learned* (5th ed.). Oxford University Press.
- Liu, M., et al. (2024). Cross-cultural comparative analysis of student motivation and autonomy in learning: Perspectives from Hong Kong and the United Kingdom. *Frontiers in Education*, 9, 1393968. <https://doi.org/10.3389/feduc.2024.1393968>
- Luo, Z. (2024). A review of Krashen's input theory. *Journal of Education, Humanities and Social Sciences*, 26, 130-135. <https://doi.org/10.54097/3fnf5786>
- Mohamed, S. (2023). The development of an Arabic curriculum framework based on a compilation of salient features from CEFR level descriptors. *The Language Learning Journal*, 51(1). <https://doi.org/10.1080/09571736.2021.1923781>
- Monita, D., & Hasan, Moh. F. (2023). The Dynamics of Affective Assessment in MI Schools: Challenges and Solutions. *Ideguru: Jurnal Karya Ilmiah Guru*, 9(1), 112–119. <https://doi.org/10.51169/ideguru.v9i1.769>
- Muhammad, I. M., Nurhadi, A., & Abdurrahim, M. (2024). Bridging theory and practice in Arabic language education. *Indonesian Journal of Islamic Studies*. <https://ijis.umsida.ac.id/index.php/ijis/article/view/1724>

- Neuman, W. L. (2020). *Social Research Methods: Qualitative and Quantitative Approaches* (8th ed.). Pearson.
- Puasa, B. (2023). I-Kalam: A CEFR based mobile gamification app for Arabic speaking skills. *ResearchGate*. https://www.researchgate.net/publication/374132612_I-Kalam_A_Cefr_Based_Mobile_Gamification_App_For_Arabic_Speaking_Skills
- Rahmi, L., Hasan, Moh. F., & Monita, D. (2025). Elementary School Class Teacher Strategy in Instruding Ideological Values Nahdlatul Ulama – Muhammadiyah. *Ideguru: Jurnal Karya Ilmiah Guru*, 10(1), 905–911. <https://doi.org/10.51169/ideguru.v10i1.1064>
- Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2021). Assessing language development in Arabic: The Arabic language: Evaluation of function (ALEF). *Applied Neuropsychology: Child*, 10(1), 37-52. <https://doi.org/10.1080/21622965.2019.1596113>
- Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2021). Assessing language development in Arabic: The Arabic language: Evaluation of function (ALEF). *Applied Neuropsychology: Child*, 10(1), 37-52. <https://doi.org/10.1080/21622965.2019.1596113>
- Ritonga, M., Wahyuni, S., & Novigator, H. (2023). The future of Arabic language learning for non-Muslims as an actualization of Wasathiyah Islam in Indonesia. *F1000Research*, 12, 27. <https://f1000research.com/articles/12-27>
- Saiegh-Haddad, E. (2025). Diglossic and orthographic features of reading comprehension in standard Arabic: The primacy of the spoken language. *Reading Research Quarterly*. <https://doi.org/10.1002/rrq.598>
- Saiegh-Haddad, E., et al. (2024). The role of story mode in the narrative skills of children in Arabic diglossia: Comparing children with typical language development and developmental language disorder. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2024_JSLHR-24-00013
- Shtayat, S., et al. (2025). Arabic-English bilingual language preference in Jordan context. *International Journal of Academic Research in Business and Social Sciences*, 15(3). <https://doi.org/10.6007/IJARBSS/v15-i3/24929>
- Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30(2), 73-87. <https://doi.org/10.1017/S0261444800012799>
- Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly*, 42(2), 181-207. <https://doi.org/10.1002/j.1545-7249.2008.tb00115.x>
- Vanpee, K., & Soneson, D. (2019). Arabic proficiency improvement through a culture of assessment. In P. Winke & S.M. Gass (Eds.), *Foreign Language Proficiency in Higher Education* (pp. 205-228). Springer. https://doi.org/10.1007/978-3-030-01006-5_11
- Watson, J. (2024). Language variation on Arabic undergraduate degree courses in England: Students' perspectives. *The Language Learning Journal*. <https://doi.org/10.1080/09571736.2024.2351938>

- White, L. (1987). Against comprehensible input: The input hypothesis and the development of second-language competence. *Applied Linguistics*, 8(2), 95-110. <https://doi.org/10.1093/applin/8.2.95>
- Yang, H. (2024). Factors influencing second language learning based on the research of Lightbown and Spada. *Frontiers in Psychology*, 15, 1347691. <https://doi.org/10.3389/fpsyg.2024.1347691>
- Zulkarnaen, F., et al. (2024). Research trend of Arabic language teaching in the world: Systematic literature review based on Scopus database. *Research and Development in Education (RaDEn)*, 4(1), 27-48. <https://doi.org/10.22219/raden.v4i1.31847>