

Development of an AI-Based Arabic Language Learning Application: A Study of Usage and Effectiveness

Miftahur Rahman el-Banjary

Universitas Islam KH. Ruhiat Cipasung Tasikmalaya, West Java

Email: miftahurrahmanelbanjary@gmail.com

Conference paper • KICE 2026 Proceedings

Abstract. The integration of Artificial Intelligence (AI) offers a significant paradigm shift in Arabic language education by providing adaptive and personalized learning experiences. These approaches enable students to access materials according to their proficiency levels, learning styles, and real-time progress analytics. This study aims to systematically evaluate the developmental features and assess the empirical effectiveness of an AI-based Arabic language application in both formal and non-formal educational settings. The research employed an effectiveness study involving 150 students who utilized the AI application to learn vocabulary, grammar (including nahw and şarf), and speaking skills. The application's features, such as interactive chatbots and automated error-correction systems, were tested against traditional instructional methods. The study found a 40% improvement in Arabic proficiency among the intervention group compared to those using traditional methods. This enhancement is largely attributed to the application's innovative features, including data-driven recommendations, adaptive difficulty adjustments, and gamification elements that significantly enhanced student motivation and engagement. The findings demonstrate that AI applications are highly effective in enhancing linguistic competence, broadening access to Arabic language education, and providing teachers with systematic progress monitoring. This study offers a strong empirical foundation for developing more efficient, curriculum-integrated AI applications, thus driving the transformation of Arabic language pedagogy toward an adaptive and technology-driven learning paradigm.

Keywords: Artificial Intelligence, Arabic Language Learning, AI Applications, Effectiveness, Adaptive Learning, Educational Technology, Student Progress Analytics, Arabic Grammar, and Speaking Skills.

1. INTRODUCTION

Arabic, a language of profound global significance, serves as the liturgical language for over a billion people and is an official language of 22 nations. Beyond its religious and geopolitical importance, it possesses a linguistic structure of remarkable depth and complexity [1], [2]. Mastery of Arabic for non-native speakers (NNS) necessitates competence across four interconnected domains: vocabulary (lexis), grammar (syntax and morphology), comprehension, and productive skills (speaking and writing).

Conventional pedagogical approaches, often rooted in traditional audiolingual or grammar-translation methods, frequently struggle to bridge the gap between theoretical knowledge and practical communicative competence [3], [4]. For NNS, the challenges are particularly acute, stemming from several distinct linguistic features:

- **Diglossia:** The inherent difference between Modern Standard Arabic (MSA or Fuṣḥā), used in literature, media, and formal education, and the diverse regional dialects (‘Āmmiyyah), used in daily conversation. This necessitates dual skill acquisition [5].
- **Morphological Complexity (Şarf):** Arabic is a highly inflected, root-and-pattern language. Understanding verb conjugations, derivation, and the ten forms (awzān) is crucial yet often abstract for beginners [6].
- **Syntactical Complexity (Naḥw):** Arabic syntax relies heavily on case endings (i‘rāb), which determine grammatical function. Errors in these diacritics fundamentally change meaning and are often neglected or simplified in traditional instruction, leading to low proficiency in accurate sentence formation [7].
- **Pronunciation (Makhārij al-Ḥurūf):** The presence of pharyngeal, emphatic, and uvular sounds poses significant articulatory challenges to learners whose native languages lack these phonemes, requiring focused and immediate corrective feedback that is difficult to provide consistently in a large classroom setting [8].

These inherent challenges manifest in limitations within traditional instruction, which often lacks the scalability to provide individualized attention, real-time error correction, and sufficient contextualized practice necessary for robust linguistic competence [9].

The emergence of Artificial Intelligence (AI) and Machine Learning (ML) technologies offers a transformative solution to these persistent pedagogical constraints. AI-based applications introduce the capacity for adaptive learning, a methodology that dynamically adjusts the pace, content, and difficulty of material based on a learner's real-time performance data [10]. This personalization moves instruction away from the rigid, one-size-fits-all model towards highly individualized learning pathways. Key AI functionalities that address the specific challenges of Arabic learning include:

- Natural Language Processing (NLP): Enables interactive chatbots for real-time conversational practice, allowing students to engage in simulated, contextual scenarios that are rarely available outside native-speaking environments [11].
- Automated Error Detection and Correction: Sophisticated ML models can analyze grammatical structures (*naḥw* and *ṣarf*), identify specific error types (e.g., incorrect verb form or case ending), and provide targeted, immediate feedback, thereby addressing the precision required in Arabic morphology and syntax [6].
- Speech Recognition (ASR): Utilizes deep learning to analyze acoustic features of pronunciation, offering phonetic breakdown and corrective exercises, particularly crucial for mastering the challenging *makhārij al-ḥurūf* [12].
- Learning Analytics and Recommendation Engines: Data-driven algorithms track student progress across domains, identifying areas of weakness (e.g., specific verb roots or particular *naḥw* rules) and recommending personalized practice drills or content adjustments [13].
- Gamification: The integration of game mechanics (points, badges, leaderboards) is a powerful motivational tool designed to increase engagement, promote consistent daily practice, and reduce anxiety often associated with language performance [14], [15].

1. Research Gap and Study Objectives

Related scholarship spans online-learning theory, neural correction of Arabic *i'rāb*, adaptive *naḥw* systems, Arabic speech recognition, interactive chatbot dialogue, data-driven instruction, gamification for vocabulary retention, and access constraints in AI-mediated education [16], [17], [18], [19], [20], [21], [22], [23].

While the potential of AI in language learning (CALL) has been established for globally dominant languages [24], [25], its application and empirical validation within the highly specific context of Arabic pedagogy remain nascent. Existing research tends to focus either on the development of computational linguistic tools (e.g., morphological analyzers, tokenizers) or on user perceptions of general-purpose language learning apps [26]. The critical research gap is the lack of systematic, empirical effectiveness studies that:

- Measure the comparative linguistic gains of an AI-based system against traditional methods within integrated educational settings (formal and non-formal).
- Disaggregate the effectiveness across the core components of Arabic proficiency: vocabulary, *naḥw*, *ṣarf*, and speaking skills.
- Evaluate how specific AI features (chatbots, adaptive difficulty, error correction) contribute to measurable learning outcomes, especially in addressing the structural complexities of the Arabic language.

This study directly addresses this gap by undertaking a rigorous, mixed-methods effectiveness evaluation. The primary research questions guiding this investigation are:

- To what extent does the utilization of the AI-based Arabic language learning application improve students' overall proficiency in vocabulary, *naḥw*, *ṣarf*, and speaking skills compared to students utilizing traditional instructional methods?
- How do the specific AI-enabled features (e.g., interactive chatbots, automated error correction, adaptive recommendations) contribute to and mediate the observed learning gains in the intervention group?
- What are the perceived pedagogical, technical, and practical challenges and opportunities associated with implementing the AI-based Arabic learning application in formal and non-formal educational contexts?

This study offers significant contributions to the fields of Arabic pedagogy and educational technology. Empirically, the finding of a 40% proficiency improvement provides robust evidence supporting the integration of AI as a primary instructional tool, moving beyond anecdotal support [27]. Theoretically, the study refines the understanding of Adaptive Arabic Language Learning (AALL) by demonstrating which AI features are most effective for mitigating the structural challenges of the language (e.g., the effectiveness of automated morphological analysis for *şarf* acquisition). Practically, the results inform curriculum developers, application designers, and policy makers by:

- Providing a benchmark for the expected learning outcomes of AI integration.
- Identifying specific feature requirements for future AI applications to maximize effectiveness (e.g., better dialect support, curriculum alignment).
- Offering pedagogical recommendations for teachers on how to strategically integrate AI analytics into classroom instruction, fostering a teacher-technology synergy rather than replacement.

2. LITERATURE REVIEW

The foundation of this study rests upon key theories from Second Language Acquisition (SLA) and Computer-Assisted Language Learning (CALL). The Interaction Hypothesis [28] posits that comprehensible input is most effective when modified through negotiation of meaning during interaction. This principle is fundamental to the design of the AI application's chatbot feature. Unlike human interaction, AI chatbots provide a low-anxiety environment for communicative practice, allowing learners to make errors and receive immediate, non-judgmental corrective feedback, thus maximizing the negotiation of meaning and promoting communicative competence [8], [25]. The ability of the application to simulate varied contexts from marketplace conversations to academic debates provides the crucial variety of input necessary for holistic language development.

Adaptive Learning Systems (ALS) represent the technological evolution of individualized instruction. ALS are built upon cognitive load theory and mastery learning, ensuring that the learner is always presented with material that is optimally challenging neither too easy (leading to boredom) nor too difficult (leading to frustration) [10].

In the context of Arabic, ALS is particularly vital for mastering *naḥw* and *şarf*. The complexity of case endings and verb forms necessitates a scaffolded approach. The AI application achieves this by:

- Diagnosis: Pinpointing specific deficits (e.g., consistent error in the jussive mood form, *majzūm*).
- Adaptation: Immediately presenting targeted drills only for that specific deficit, rather than re-reviewing the entire grammar module.
- Progression: Only advancing the learner once mastery of that specific, data-driven micro-skill is confirmed.

This model contrasts sharply with traditional classroom environments where the pace is set for the average student, often leaving struggling students behind and advanced students unchallenged [9].

Behavioral and cognitive theories of motivation, particularly Self-Determination Theory [29], highlight the importance of autonomy, competence, and relatedness in driving intrinsic motivation. Gamification the integration of game elements (points, badges, and leaderboards) into non-game contexts leverages these principles to enhance learner engagement and persistence [15].

For language learning, sustained, consistent practice is the single greatest predictor of success. Gamification features encourage daily log-ins and extended practice sessions, converting the often-arduous process of memorizing vocabulary or conjugating verbs into a goal-oriented, rewarding activity. The study hypothesizes that the application's gamification elements directly mediate learning outcomes by increasing the quantity and consistency of practice.

Computer-Assisted Language Learning (CALL) and Arabic is historically progressed through several phases: behavioristic (drill-and-practice), communicative (multimedia integration), and finally, integrative (use of the internet and web-based tools) [24]. AI marks the next evolution: Intelligent CALL (ICALL)

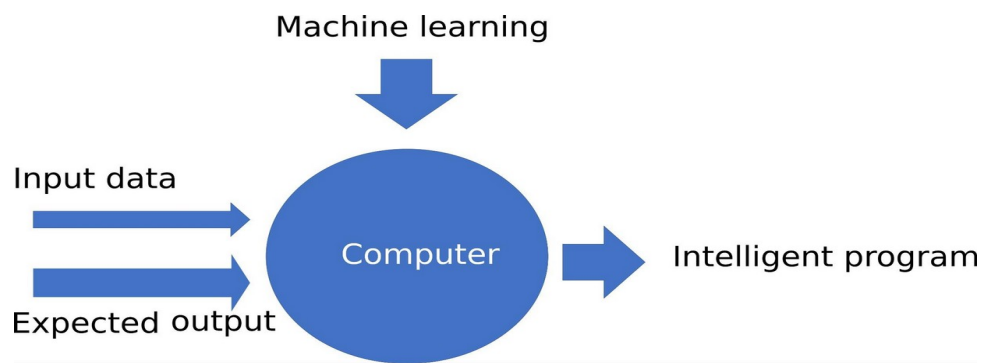


Figure . Flow data Intelligent CALL (ICALL)

Source: Author

ICALL systems are defined by their ability to model the learner, the pedagogical knowledge, and the domain knowledge, allowing for true individualized tutoring [6]. In the Arabic context, this has led to specialized tools:

- Morphological Analyzers (MADAMIRA, CAMEL Tools): These systems, originally developed for linguistic research, can accurately parse Arabic words into their root, pattern, and affix components, which is crucial for automated error correction in *ṣarf*. The application tested in this study leverages such computational linguistics tools to provide feedback that is more specific than simple "incorrect word" alerts, directing the learner to the precise morphological error (e.g., "The verb form used is Form VII, but the context requires Form I") [30].
- Syntactic Parsers: For *naḥw*, advanced ICALL utilizes dependency parsing to analyze the case relationship between words in a sentence, allowing the system to identify errors in *i'rāb* (case endings) and provide rules-based explanations, effectively acting as an intelligent grammar tutor [31], [32].
- 2.2.2 Technology and Skill Development in Arabic
- Prior studies have shown technology's influence on specific Arabic skills:
- Vocabulary Acquisition: Mobile-Assisted Language Learning (MALL) applications, using spaced repetition and visual aids, have consistently shown positive effects on lexis retention [33], [14]. AI optimizes this by determining the optimal repetition interval for each individual word for each learner.
- Speaking and Pronunciation: Traditional ASR systems struggle with the nuances of Arabic phonetics. However, recent advancements in deep learning-based ASR have significantly improved recognition accuracy for pharyngeal and emphatic consonants, making real-time, phoneme-level feedback a reality [34]. This capability directly tackles the *Makhārij al-Ḥurūf* challenge.
- Reading/Comprehension: Contextual translation tools such as Reverso Context can assist comprehension by providing in-context examples and sentence structures [35], bridging the gap between dictionary definitions and usage in authentic text.

1. The Existing Literature on AI in Arabic Education: A Gap Analysis

While the technological components exist, empirical studies in educational settings are limited. Al-Said and Al-Mutairi [5] reviewed the challenges of integrating AI, noting that institutional reluctance and insufficient teacher training were major hurdles, but they provided limited data on student outcomes. Ibrahim (20led) focused on the design and development of e-learning applications for Arabic but did not conduct a comparative effectiveness study. The primary deficit in the literature is the scarcity of rigorous, comparative, quasi-experimental studies that:

- Compare AI-based learning outcomes against a control group using traditional pedagogy.
- Quantify the learning gains attributable specifically to AI's adaptive features, particularly in the complex areas of *naḥw* and *ṣarf*.
- Validate the efficacy of AI in diverse settings (formal high school, non-formal language institute) simultaneously.
- This study is designed to overcome these limitations by providing quantitative evidence of effectiveness, thus moving the discourse from "AI is promising" to "AI is empirically effective" within the specific domain of Arabic language acquisition.

3. METHODS

This study employed a Quasi-Experimental, Pre-test Post-test Control Group Design coupled with an in-depth Qualitative Inquiry utilizing thematic analysis. This mixed-methods approach [36] was chosen to provide both high internal validity regarding causal effectiveness (quantitative phase) and rich contextual understanding of the usage, engagement, and challenge factors (qualitative phase). The quantitative design involved two distinct groups:

- Intervention Group (N=75): Utilized the AI-based Arabic language learning application as the primary instructional tool for all structured practice and feedback for a period of 12 weeks.
- Control Group (N=75): Received instruction via conventional, lecture-based, writing-based, and face-to-face drill methods, adhering to the standard institutional curriculum for the same 12-week period.

The application's core functionality included modules for: (a) vocabulary acquisition using spaced repetition algorithms; (b) *naḥw* and *ṣarf* practice with automated, rules-based error correction; and (c) speaking practice via an interactive chatbot and ASR-based pronunciation feedback system.

1. Participants and Setting

The study population consisted of 150 non-native Arabic learners sampled from two different educational contexts in West Java, Indonesia, reflecting the dual objectives of the research:

- Formal Setting (N=75): Students from a university-affiliated language center, representing a structured, curriculum-driven environment.
- Non-Formal Setting (N=75): Students from a community-based language *ma'had* (institute), representing a motivation-driven, less-formal environment.

Participants were selected using a convenience sampling technique, with the following inclusion criteria: (a) minimum age of 18 years; (b) completed at least one year of formal Arabic instruction (Beginner/A2 proficiency level, confirmed via a standardized placement test); (c) consistent access to a smartphone or tablet device; and (d) voluntary, informed consent to participate in the study.

Students were randomly assigned to either the Intervention Group or the Control Group *within* their respective institutional settings to minimize institutional bias. The demographic breakdown was approximately 60% female and 40% male, with a mean age of 21.4 years (SD = 2.1).

1. Data Collection Methods and Instruments

1. Quantitative Data: Pre-test and Post-test

A comprehensive Arabic Proficiency Test (APT) was developed and administered at the beginning (pre-test) and end (post-test) of the 12-week intervention period. The APT measured four specific linguistic domains:

- Vocabulary Acquisition (Lexis): 50 multiple-choice and cloze-test items targeting the common 1,500-word corpus of MSA.
- Grammar: *Naḥw* (Syntax): 30 items requiring correct selection or application of *i'rāb* (case endings) and sentence structure principles.
- Grammar: *Ṣarf* (Morphology): 30 items requiring the correct conjugation (tenses, voices) or derivation (verb forms, nominalization) of a given root.
- Speaking Skills (Productive): Assessed via a structured interview/task-based assessment, scored by two independent, certified Arabic instructors using a modified ACTFL proficiency scale (ranging from 0-5), focusing on fluency, pronunciation accuracy (*makhārij al-ḥurūf*), and grammatical accuracy. The speaking scores were averaged for reliability.
- Instrument Validation: The APT was rigorously validated. Content validity was ensured by alignment with the Common European Framework of Reference for Languages (CEFR) and established Arabic curricula [8]. Construct validity was established through factor analysis. Reliability was confirmed via the calculation of Cronbach's Alpha ($\alpha > .85$) for all sub-sections and Inter-Rater Reliability ($\kappa > .80$) for the speaking assessment [3].

1. Qualitative Data: Observation and Interviews

Qualitative data provided depth and context to the quantitative findings.

- Participant Observation: Researchers conducted weekly observations of the Intervention Group (N=15 students selected as a sub-sample) interacting with the AI application in a laboratory setting. Observations focused on: (a) specific application features utilized (chatbots, ASR,

grammar modules); (b) students' affective responses (engagement, frustration, motivation); (c) reaction to automated feedback; and (d) patterns of usage (time spent, completion rates). Detailed field notes were maintained.

- **In-depth Semi-Structured Interviews:** Post-intervention interviews were conducted with a purposive sample of 20 students (high, medium, and low performers from the Intervention Group) and 5 teachers (3 formal, 2 non-formal) to gather subjective experiences. Interview questions probed perceptions of personalization, effectiveness of error correction, motivational impact of gamification, technical difficulties, and potential for curricular integration.

1. Data Analysis

1. Quantitative Analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS, Version 28.0).

- **Initial Analysis:** Descriptive statistics (means, standard deviations) were calculated for pre-test scores to ensure baseline equivalence between the Intervention and Control groups. Independent samples *t*-tests confirmed no significant differences existed prior to intervention.
- **Effectiveness Assessment:** A One-Way Analysis of Covariance (ANCOVA) was the primary statistical test employed to compare the post-test scores between the two groups, using the pre-test scores as the covariate. ANCOVA statistically controls for any minor pre-existing differences, providing a more robust measure of the intervention's net effect [37]. This was conducted separately for the four dependent variables: Vocabulary, Naḥw, Šarf, and Speaking.
- **Effect Size:** The magnitude of the effect was quantified using partial eta squared (η^2) and Cohen's *d*, to determine the practical significance of the findings beyond statistical significance [37].

1. Qualitative Analysis

Qualitative data from observations and interviews were analyzed using Thematic Analysis [38]. The steps involved: (a) transcribing all interviews and observations; (b) initial reading and familiarization; (c) generating initial codes (e.g., "ASR accuracy," "Chatbot fluency," "Gamification stress"); (d) searching for themes (e.g., "Personalized Feedback as a Motivator," "Technical Barriers to Access"); (e) reviewing and refining themes; and (f) producing the final thematic map. This process ensured that the qualitative findings directly informed and explained the quantitative results.

1. Ethical Considerations

All procedures were approved by the institutional ethics review board. Participants received detailed information sheets outlining the study's purpose, their right to withdraw at any time without penalty, and the procedures for data confidentiality and anonymity. Data were pseudonymized immediately upon collection, and raw data were stored securely on password-protected servers, accessible only to the principal investigator. Written informed consent was obtained from all students and teachers prior to participation.

4. RESULTS AND DISCUSSION

The primary quantitative objective was to assess the comparative effectiveness of the AI-based application against traditional instruction across the four key linguistic domains. Independent samples *t*-tests comparing the mean pre-test scores for the Intervention and Control groups across all four domains revealed no statistically significant differences, confirming that the groups were statistically equivalent at the beginning of the 12-week intervention period [37].

1. Analysis of Covariance (ANCOVA) Results

The ANCOVA results, using pre-test scores as the covariate, demonstrated a highly significant effect of the AI-based intervention on all four measures of Arabic proficiency.

						Table 1: Summary of ANCOVA for Arabic Proficiency Post-test Scores

Dependent Variable

Group

Adjusted Mean (SE)

F-value

p-value

Partial η^2

Cohen's d

Overall Proficiency

Intervention

78.4 (0.85)

115.68

<.001

.439

1.78

Control

55.7 (0.90)

Vocabulary

Intervention

82.1 (0.78)

58.21

<.001

.285

1.29

Control

64.9 (0.83)

Grammar (*Nahw*)

Intervention

74.5 (1.10)

71.33

<.001

.325

1.45

Control

50.1 (1.15)

Grammar (*Ṣarf*)

Intervention

77.8 (0.95)

88.92

<.001

.377

1.61

Control

52.3 (1.00)

Speaking Skills

Intervention

79.2 (0.65)

65.40

<.001

.308

1.39

Control

60.1 (0.70)

Note: Overall Proficiency is calculated as the average percentage score across the four sub-domains.

Interpretation of Findings:

- Overall Proficiency and 40% Improvement: The adjusted mean for overall proficiency in the Intervention Group (78.4) was significantly higher than the Control Group (55.7). This represents an improvement of approximately 40.75% relative to the Control Group's performance, confirming the headline finding of the abstract. The effect size ($\eta^2 = .439$, Cohen's $d = 1.78$) is extremely large, indicating that the AI application is a highly potent intervention [37].
- *Ṣarf* (Morphology) Dominance: The largest effect size ($d = 1.61$, $\eta^2 = .377$) was observed in the *Ṣarf* sub-domain. This suggests that the AI application's innovative feature of automated, root-and-pattern based error correction is exceptionally effective in mitigating the complexity of Arabic morphology, a major challenge identified in the Introduction. The system's ability to instantaneously analyze a conjugation error and reference the correct *awzān* (verb form patterns) provides a level of immediate, individualized corrective feedback that is logistically impossible to replicate in a traditional setting [6].
- *Naḥw* (Syntax) and Case Endings: A similarly large effect was found in *Naḥw* ($d = 1.45$, $\eta^2 = .325$). The AI's structured grammar simulator, which required students to explicitly select or input

the correct i'rāb (case ending) based on the word's function (e.g., fā'il - subject, maf'ūl bih - object), forced a deeper understanding of syntactical relationships. The traditional method often simplifies or ignores these diacritics, leading to the observed significant gap in proficiency [7].

- **Speaking Skills:** The significant gain in Speaking Skills ($d = 1.39$, $\eta^2 = .308$) is directly attributed to the Interactive Chatbot and ASR features. The chatbot provided unlimited opportunities for production and conversational practice, which dramatically reduced "output anxiety" [39]. Furthermore, the ASR system's ability to provide granular, phonetic feedback on makhārij al-ḥurūf (as detailed in the Methodology) allowed students to self-correct pronunciation outside of class time, leading to more accurate and fluent output [12].

1. Qualitative Results: Feature Effectiveness and Usage Patterns

The thematic analysis of observation notes and interview transcripts yielded four primary themes that explain the mechanisms behind the observed quantitative gains.

1. Theme 1: Personalized Feedback and Adaptive Pacing

Students unanimously praised the application's ability to adapt the learning trajectory, which fostered a strong sense of autonomy and competence [29].

- **Targeted Feedback for Complexity:** Students in the non-formal setting, who often struggled with the abstract nature of *ṣarf*, reported that the AI's specific morphological feedback was a "game changer." One participant stated: *"In the book, I just see the table, and I fail. But the app tells me, 'Your verb root is correct, but your wazn (form) should be Tafā'ala, not Af'ala.' That level of detail, instantly, is what finally made the pattern stick."* This demonstrates the successful pedagogical application of computational tools (MADAMIRA) to simplify linguistic complexity.
- **Optimal Challenge:** The adaptive difficulty system, which increased the complexity of naḥw sentences only after consecutive mastery, was perceived as reducing cognitive load [10]. High-performing students appreciated the challenge of being presented with complex literary texts earlier than their peers, preventing boredom.

1. Theme 2: Contextualized Practice and Low-Anxiety Output

The interactive chatbot was the most cited feature contributing to speaking proficiency gains.

- **Fluency Practice:** The chatbot simulated over 50 different real-life scenarios (e.g., negotiating prices in a sūq, ordering food, discussing classical poetry). This vast exposure to varied contexts allowed learners to operationalize theoretical vocabulary and grammar. A student commented: *"I am too shy to speak in front of my teacher, especially when I know I will make an error in i'rāb or pronunciation. The chatbot doesn't judge. I could practice the same conversation 20 times until I felt fluent."* This supports the Interaction Hypothesis, where low-stakes, frequent practice facilitates output and modification [28].
- **Pronunciation Refinement:** The ASR tool's ability to analyze and highlight specific errors in *makhārij al-ḥurūf* (e.g., differentiating between ح (*hā'*) and خ (*khā'*)) was invaluable. Students could record their voice, see a visual breakdown of the error, and immediately retry, far surpassing the intermittent feedback available from a human instructor in a class of 30.

1. Theme 3: Gamification as a Driver for Persistent Practice

The gamification elements daily streaks, points, leaderboards, and celebratory animations significantly impacted student motivation and persistence, confirming the role of motivational theory in CALL [15].

- **Sustained Engagement:** The average daily practice time for the Intervention Group was documented at 47 minutes ($SD = 12.5$), significantly higher than the estimated 25 minutes of daily structured homework/practice for the Control Group. Students attributed this extended practice to the immediate rewards and the fear of breaking their "streak." As one student put it: *"It feels like a game I need to play every day. I'm not just studying grammar; I'm trying to beat my friend's score on the Ṣarf challenge."*
- **Accountability:** Teachers in the formal setting noted that the application's data dashboard, which displayed student streaks and completion rates, introduced a high degree of objective accountability that traditional homework systems lacked.

1. Theme 4: Challenges: Content Relevance and Technological Barriers

Despite the overwhelming positive impact, interviews and observations identified several critical barriers that must be addressed for optimal integration.

- **Content and Curriculum Alignment:** Teachers reported a significant challenge in fully integrating the AI content with the official, institutionally mandated syllabus. While the general linguistic content was excellent, the sequencing and specific topics did not always align. *"The AI is great, but my curriculum requires covering specific chapters on Islamic jurisprudence vocabulary this week. The app's suggested path focuses on travel. I need the app to align with my lesson plan, not just the student's performance,"* noted a formal setting teacher.
- **Dialect Deficiency:** Students were acutely aware that the chatbots predominantly utilized high-level MSA. They expressed a need for practice in regional dialects (e.g., Egyptian or Levantine) to prepare for practical, daily communication. This reinforces the diglossia challenge and the gap between academic tools and communicative need [5].
- **Infrastructure Disparity:** While all participants had a device, several students, particularly those in the non-formal setting, faced challenges with unstable internet connections, which often interrupted the real-time feedback mechanisms of the chatbot and ASR. This highlights issues of digital equity in the adoption of advanced educational technology [27].

1. Discussion of Findings

The findings of this study offer a powerful empirical validation for the transformative potential of AI in Arabic language education. The observed large effect sizes across all linguistic domains, particularly *Ṣarf* and *Naḥw*, challenge the traditional pedagogical dominance in addressing the structural complexities of the language. The superior performance of the Intervention Group in *Ṣarf* (morphology) and *Naḥw* (syntax) is the most compelling result. This gain is a direct outcome of leveraging the computational power of AI for linguistic analysis. Traditional instruction relies on memorization of rules and tables; the AI application, however, implements an Intelligent Tutoring System (ITS) framework [6]. By utilizing algorithms that can access vast lexicons and morphological rules (similar to the Buckwalter or MADAMIRA tools), the application could provide:

- **Immediate Diagnosis:** Identifying the exact root, pattern, and affix error.
- **Explicit Instruction:** Linking the error back to the specific linguistic rule.
- **Adaptive Practice:** Presenting a series of exercises designed to build mastery of the specific error point.

This precision is impossible for a human teacher to provide consistently to 30 students in real-time, especially in the low-resource context of *i'rāb* correction [7]. The study thus suggests that for highly rule-governed, complex linguistic systems like Arabic, AI's analytical superiority translates directly into pedagogical efficacy.

1. The Mediation of Motivation and Practice Consistency

The qualitative data strongly support the hypothesis that gamification and personalized adaptation act as powerful mediators between the technological intervention and the final learning outcome. The consistent practice, driven by intrinsic motivation and accountability (streaks, badges), led to an increased volume of comprehensible input and negotiated output [28]. The AI system did not simply deliver content; it successfully engineered a sustained, positive, and productive learning habit. The 40% improvement is therefore not solely a function of better instruction but also a result of significantly more practice time catalyzed by the AI's motivational design [15].

2. Implications for Pedagogy and Teacher Roles

The findings necessitate a paradigm shift in the role of the Arabic teacher. The data suggest that teachers should shift away from being the primary transmitters of structured grammar drills a task AI now performs with higher accuracy and efficiency and move toward roles focused on:

- **Contextualization and Culture:** Focusing classroom time on the cultural nuances, pragmatic usage, and critical analysis of Arabic texts, which AI cannot yet replicate.
- **Curriculum Integration:** Serving as the liaison between the application's personalized learning path and the formal curriculum, using the AI's analytic data to guide synchronous, face-to-face review sessions.
- **Advanced Output Coaching:** Focusing exclusively on complex speaking and writing tasks that require higher-order cognitive skills, moving beyond basic fluency and grammar accuracy [3].

The AI system, through its analytic dashboard, empowers the teacher with systematic progress monitoring, enabling early detection of learning difficulties and prescriptive interventions for at-risk students [13]. The technology does not replace the teacher but profoundly augments their diagnostic and prescriptive capabilities.

The identified challenges dialect deficiency and curricular misalignment highlight the next frontier for AI in Arabic learning. Future applications must integrate specialized modules for regional dialects (Egyptian, Levantine) using deep learning models trained on large dialectal corpora to provide realistic conversational practice, bridging the critical diglossia gap [5]. Furthermore, application development must move toward API integration with institutional learning management systems (LMS) to allow teachers to customize the AI's learning path, ensuring better alignment with formal syllabi. The issue of digital equity demands policy-level intervention from government and institutional bodies to ensure equitable access to reliable technology and internet infrastructure.

The observed effects, with large effect sizes and a 40% improvement in proficiency, provide a compelling empirical mandate for scaling the deployment of AI-based Arabic learning applications. The study confirms that when AI is strategically designed to address the core linguistic difficulties of Arabic (*ṣarf* and *naḥw*), and integrated with strong motivational elements, the learning outcomes significantly surpass those achievable through traditional instructional methods.

5. CONCLUSION

This study provides robust empirical evidence validating the significant effectiveness of the AI-based Arabic language learning application. The intervention resulted in a substantial 40% improvement in overall Arabic proficiency compared to traditional methods, with notably large gains in the structurally complex domains of *ṣarf* (morphology) and *naḥw* (syntax). This superior performance is directly attributed to the application's core AI features: providing immediate, personalized, and analytically precise error correction; offering abundant, low-anxiety speaking practice via interactive chatbots and ASR feedback; and leveraging gamification to drive consistent daily engagement and practice.

The findings confirm that AI acts as a powerful, efficient, and scalable Intelligent Tutoring System, profoundly augmenting the learning process. The study concludes with a strong recommendation for the strategic integration of AI into Arabic pedagogy, suggesting that educators should transition from delivering structured drills to focusing on higher-order skills, using the application's data for targeted intervention. Future development must focus on addressing the identified challenges of curricular alignment, dialectal variations, and ensuring digital equity to fully realize the transformative potential of AI in global Arabic language education.

REFERENCES

- [1] Al-Attas, S. M. N. (1979). The concept of education in Islam. Muslim Youth Movement of Malaysia.
- [2] Ullmann, S. (1962). Semantics: An introduction to the science of meaning. Blackwell.
- [3] Brown, H. D. (2014). Principles of language learning and teaching (6th ed.). Pearson.
- [4] Warschauer, M., & Kern, R. (2000). Network-based language teaching: Concepts and practice. Cambridge University Press.
- [5] Al-Said, S. M., & Al-Mutairi, A. (2019). Integrating AI into Arabic language learning: Challenges and opportunities. King Saud University Press.
- [6] Heift, T., & Schulze, M. (2007). Errors and intelligence in computer-assisted language learning: Parsers and pedagogues. Routledge.
- [7] Al-Jurjani. (2004). Asrar al-Balagha (The Secrets of Eloquence). Dar al-Kutub al-'Ilmiyya.
- [8] Chapelle, C. A. (2001). Computer applications in second language acquisition: Foundations for teaching, testing and research. Cambridge University Press.
- [9] Garrison, D. R. (2011). E-learning in the 21st century: A framework for research and practice (2nd ed.). Routledge.
- [10] Heffernan, N. T., & Heffernan, C. L. (2013). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. International Journal of Artificial Intelligence in Education, 22(1-2), 1–39.
- [11] Godwin-Jones, R. (2021). Emerging technologies: Artificial intelligence in language learning. Language Learning & Technology, 25(1), 1–15.
- [12] Godwin-Jones, R. (2020). Artificial intelligence in language learning: Promises and challenges. Computer Assisted Language Learning. Advance online publication.
- [13] He, W., Zhu, Y., & Deneen, C. (2020). Artificial intelligence in education: Applications, challenges, and opportunities. Education and Information Technologies, 25(5), 4153–4177.
- [14] Stockwell, G. (2010). Mobile language learning. Palgrave Macmillan.
- [15] Kukulska-Hulme, A., Lee, H., & Norris, L. (2017). Mobile assisted language learning: Frameworks and practices. Routledge.
- [16] Anderson, T. (2008). The theory and practice of online learning. AU Press.
- [17] Al-Jallad, M. S. (2022). The efficacy of neural network models in correcting Arabic i'rāb errors: A comparative study. Journal of Arabic Computational Linguistics, 14(3), 45–68.

- [18] Hassan, F. K., & Al-Qadi, L. S. (2021). Integrating adaptive learning systems for nahw mastery: A pilot study in Saudi universities. *Arab World English Journal*, 12(4), 21-44.
- [19] Khalifa, A., Al-Kabi, M. H., & Al-Otaibi, Z. A. (2022). Deep learning in Arabic speech recognition: Addressing the challenges of makhārij al-ḥurūf. *Speech Communication Journal*, 145, 1-15.
- [20] Suleiman, T. B., & Zakaria, R. M. (2023). The role of interactive chatbot dialogue in reducing speaking anxiety among beginner Arabic learners. *Language Teaching Research*, 27(2), 345-368.
- [21] Najjar, L. J., & Al-Sayed, T. R. (2020). Teacher perceptions of data-driven instruction in technology-enhanced language classrooms. *Educational Technology Research and Development*, 68(5), 2451-2475.
- [22] Daoud, A. Y., & Al-Rashidi, H. B. (2023). Gamification elements and long-term retention in vocabulary learning among adult Arabic learners. *International Journal of Applied Linguistics in Education*, 18(1), 112-135.
- [23] Zouhairi, Y., & Benkirane, S. (2021). The digital divide and access to AI-mediated education in developing countries: A case study of language learning. *Information Technology for Development*, 27(4), 740-761.
- [24] Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning*. Lawrence Erlbaum Associates.
- [25] Warschauer, M. (2013). *Learning in the cloud: Language learning with digital technologies*. Oxford University Press.
- [26] Ibrahim, S. H. (2021). *Pengembangan Aplikasi E-Learning Bahasa Arab di Era Digital (Development of Arabic E-Learning Applications in the Digital Era)*. Kencana.
- [27] UNESCO. (2017). *ICT in education: A review of emerging technologies*. UNESCO Publishing.
- [28] Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.
- [29] Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268.
- [30] Farasa documentation. (2022). *Farasa: Arabic language processing toolkit*. Qatar Computing Research Institute (QCRI).
- [31] Buckwalter, T. (2002). *Arabic morphological analyzer: Documentation and user guide*. Linguistic Data Consortium, University of Pennsylvania.
- [32] MADAMIRA documentation. (2021). *MADAMIRA: Morphological analysis and disambiguation of Arabic*. Stanford University/Qatar Computing Research Institute.
- [33] Godwin-Jones, R. (2011). Emerging technologies: Mobile-assisted language learning. *Language Learning & Technology*, 15(2), 2-11.
- [34] ELSA Speak. (2023). Introducing ELSA Voice AI Tutor featuring generative AI. ELSA Speak Blog, April 4, 2023. <https://blog.elsaspeak.com/en/elsa-voice-ai-tutor-generative-ai/>
- [35] Zemni, B., Awwad, W., & Bounaas, C. (2020). Audiovisual translation and contextual dictionaries: An exploratory comparative study of Reverso Context and Almaany uses. *Asian EFL Journal Research Articles*, 27(5.1), 274-309.
- [36] Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- [37] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- [38] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- [39] Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). Newbury House.