Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/150601

Statistical Analysis of the Degree of Arabic Language Teachers' AI Integration of Artificial Intelligence Application in Teaching Language Skills

Shaman Saleh Marie Maabreh

Department of Arabic Language and Literature, Faculty of Arts and Languages, Jadara University, Jordan

Received: 7 Aug. 2026, Revised: 20 Sep. 2026, Accepted: 21 Oct. 2026
Published online: 1 Nov. 2026

Abstract: This study investigated the extent to which Arabic language teachers employ artificial intelligence (AI) applications to teach Arabic language skills to their students in Jordan. Employing a descriptive-analytical research design, the study sampled 135 male and female Arabic language teachers (N = 135), recruited through a convenience sampling procedure from public and private schools affiliated with the Third Irbid Directorate of Education. Self-reported data were collected using a 26-item questionnaire structured across four primary skill domains: Listening, Speaking, Reading, and Writing. Statistical analyses included descriptive metrics (means, standard deviations, and 95% confidence intervals) and inferential tests, specifically a main-effects four-factor analysis of variance (ANOVA) with effect sizes, post hoc comparisons, and a sensitivity power analysis. The results indicated an overall moderate level of self-reported AI integration across all language domains (mean [M] = 3.25, standard deviation [SD] = 0.64). Among the individual domains, Reading ranked highest, followed by Writing, Speaking, and Listening, with each domain yielding a moderate evaluation. Inferential results revealed no statistically significant differences in AI integration attributable to gender or years of experience. However, statistically significant differences were observed for school type, favoring private schools over public schools, and for educational qualification, favoring postgraduate degree holders over Bachelor's degree holders; these effects were small to medium (partial η^2 = .058 and .066, respectively) and, given the unbalanced subgroups (22 private vs. 113 public school teachers) and the convenience sample, should be interpreted cautiously. The findings reflect teachers' self-reported practices rather than observed classroom behavior. Based on these findings, it is recommended that educational policymakers establish systematic, institutionally supported professional development programs—particularly targeting public school educators and prioritizing oral language skills—to transition AI integration from ad hoc teacher usage toward structured, scalable pedagogical practices, supported by audio infrastructure and vetted Arabic-compatible tools.

Keywords: AI-Assisted Language Learning, Arabic Language Instruction, Teaching Language Skills, Technology Adoption, Teacher Competencies, Self-Reported Practices.

1 Introduction

Arabic is not merely a means of communication; it is the cornerstone of cultural identity and a vessel for thought. Its importance lies in its four skills (listening, speaking, reading, and writing), which form an integrated system that enables learners to achieve deep understanding, critical analysis, and eloquent expression. Mastering these skills allows students to bridge their rich heritage with modern demands, fostering cognitive flexibility and qualifying them to actively participate in building society.

Arabic language skills are essential for effective communication across social and cultural contexts. The growing global demand for language acquisition requires transitioning from traditional teaching to modern technological models. Consequently, contemporary education emphasizes creating flexible, interactive environments that transform abstract ideas into fluent expression, supporting syntactic correctness through automated correction tools and digital language-enrichment programs [1].

* Corresponding author e-mail: shaman.m@jadara.edu.jo

© 2026 NSP
Natural Sciences Publishing Cor.

Language educators seek to the four language skills—including listening, speaking, reading, and writing—which are linked to thinking and result in both functional and purposeful language use, represented by Arabic language systems: grammar, phonology, morphology, semantics, and rhetoric [2,3,4]. Mastering these skills and systems fosters positive interactions among learners and improves their overall learning experience.

Language education systems are undergoing a significant transformation as they integrate artificial intelligence (AI) technologies and tools. These technologies have evolved beyond their traditional supportive role to become comprehensive learning environments that enhance learners' proficiency across the four language skills: listening, speaking, reading, and writing. Within the domain of productive writing skills, the literature indicates that AI-powered grammar checking and language editing tools help address learners' linguistic deficiencies by providing immediate corrective feedback that promotes self-directed learning

And regarding receptive skills like reading, natural language processing (NLP) platforms enable learners to contextualize texts and grasp semantic nuances, thereby enhancing reading comprehension. Extending to oral skills (listening and speaking), digital educational theory emphasizes AI-driven interactive environments; conversational models like Catgut provide virtual dialogue partners that reduce language anxiety and offer personalized guidance to improve students' speaking accuracy and fluency [5,6].

A growing body of literature highlights the expanding role of AI applications in education, particularly in enhancing language teaching practices and developing students' language skills. The integration of these advanced technologies creates interactive learning environments and provides learners with personalized, immediate support to help them overcome authentic learning challenges. Furthermore, AI contributes to improving instructional strategies and curriculum development in ways that align with the demands of the digital era. In a related context, growing scholarly attention and international collaboration have highlighted the strategic use of AI technologies to advance progress toward the Sustainable Development Goals (SDGs) [7,8,9].

Contemporary educational trends have shown increasing interest in integrating AI technologies into educational settings, as numerous studies have demonstrated their effectiveness in developing learners' language skills [10,11]. AI-powered tools create interactive, engaging learning environments that enable learners to practice the four core language skills—listening, speaking, reading, and writing—in personalized ways that address their individual learning needs [12]. These applications also provide immediate, accurate feedback, allowing learners to identify and correct their linguistic errors and improve the quality of their oral and written language production. Moreover, AI plays a significant role in fostering self-directed learning skills, thereby enhancing learners' motivation to learn the language and achieve greater proficiency in its various skills [13].

Integrating AI with the Arabic language marks a pivotal shift in the technological and educational landscape, opening broad horizons for the service of the language. This intersection helps overcome challenges posed by Arabic's grammatical, morphological, and phonological complexities, thereby enhancing digital accuracy. Educationally, this integration develops innovative learning tools that make language acquisition more interactive and effective through automated assessment, personalized content tailored to individual levels, and accurate feedback via virtual tutors, ultimately bolstering the language's global digital presence and efficiency [14,15,16].

Implementing AI in Arabic instruction faces several challenges. Overreliance on digital tools weakens learners' natural linguistic aptitude, while a lack of practice diminishes their ability to discern machine-generated errors. Furthermore, AI lacks true contextual understanding, fails to accommodate individual differences, and provides unreliable academic documentation. In practice, schools struggle with weak technical infrastructure and a lack of specialized teacher training [17,18], necessitating standardized frameworks for effective integration.

AI represents a major technological shift in language education, mimicking human cognitive and learning abilities [19]. Integrating these technologies helps overcome traditional barriers to learning by providing low-anxiety, independent environments for language practice. Key benefits include developing speaking skills through human-like dialogue partners and improving writing via instant feedback and vocabulary expansion [20]. However, challenges remain, including technical issues, data privacy concerns, and limitations in interactivity [21].

AI is defined as "a computer-based system that mimics human intelligence and practices, representing an intersection of various disciplines, such as computer science, linguistics, and psychology, to serve humanity" [19]. This definition

© 2026 NSP
Natural Sciences Publishing Cor.

underscores the integrative nature of artificial intelligence and its capacity to support the educational process by providing intelligent learning environments that foster language development and enhance learning outcomes.

Contemporary literature highlights various AI tools in language education: adaptive systems like Knewton customize content; chatbots like ChatGPT and Duolingo provide instant communication; and tools like Grammarly and Qalam improve reading and writing. Additionally, Kahoot! Boosts engagement; Monday VR (virtual reality) offers immersive experiences; Proctorio supports objective assessment; and Amazon Alexa enables voice interaction [22,23].

Teachers' competence in using AI applications is a fundamental prerequisite for the effective integration of AI into education. These competencies require specialized knowledge and skills to design adaptive learning environments that accommodate individual learner differences, provide immediate feedback, incorporate gamification strategies, and employ intelligent assessment to analyze students' performance. They also encompass the use of virtual and augmented reality technologies to deliver interactive learning content that facilitates language acquisition, as well as AI-powered chat bots and the Internet of Things (IoT) to enhance communication and provide personalized learning support [24,25].

Integrating AI is essential for enhancing teaching, boosting motivation, and providing personalized Arabic instruction. However, practical implementation faces significant challenges. A study by Al-Manajreh et al [26] found that teachers perceive these obstacles as high overall, with technical infrastructure issues leading the way, followed by ethical concerns and a lack of pedagogical knowledge. Overcoming these barriers requires developing technology-driven curricula and providing specialized training to improve teachers' digital competencies.

Several previous studies have examined AI in language education. Hamadneh [27] investigated the degree to which Arabic language teachers in Qasabat Al-Mafraq possess skills for integrating AI across three domains: planning, lesson execution, and evaluation. Using a descriptive-analytical approach and a 27-item questionnaire administered to 98 teachers, the results showed an overall moderate level of competence. Lesson execution ranked first (high), while planning and evaluation followed (moderate). Significant statistical differences were found by gender (favoring males), whereas no differences were observed for educational qualification or experience.

Meanwhile, Mureed [28] explored the reality and challenges of employing AI in primary school Arabic instruction. Using a mixed-methods approach, including a questionnaire administered to 130 teachers and interviews with 5 teachers, the study found a moderate overall level of AI integration, with challenges ranking highest. Significant differences emerged based on age, favoring younger teachers (under 40), while qualifications showed no differences except in the challenges domain (favoring master's degrees or below). Qualitative data highlighted teachers' willingness to integrate AI, provided they receive specialized training, technical support, and solutions to privacy concerns and Arabic-language compatibility issues.

Al-Azzawi [29] investigated the impact of AI tools on the development of language skills among 60 second-year fine arts students in Diyala, Iraq, using a quasi-experimental design. The experimental group (30 students),who were taught using AI significantly outperformed the control group (30 students) taught conventionally in an objective test assessing reading, speaking, writing, and listening. The study recommended integrating AI tools into language development programs.

Al-Abadi [30] investigated the effectiveness of generative AI apps (Chat GPT and Sahihely) on written expression and self-regulated learning among 55 third-year middle school students in Taif, Saudi Arabia, using a quasi-experimental design. The experimental group (28 students) significantly outperformed the control group (27 students) in the post-test for all writing sub-skills and the self-regulated learning scale, showing a very large effect size.

Yatri et al. [31] explored the experiences, perceptions, and challenges of 18 Indonesian students and teachers regarding AI integration in English and Arabic education using a qualitative descriptive approach. Findings revealed that English learners utilized diverse tools (ChatGPT, Grammarly, Elsa, QuillBot), whereas Arabic learners relied almost exclusively on ChatGPT for academic writing and research plans. While participants held positive attitudes toward AI's efficiency, they raised concerns about plagiarism, diminished creativity, limited free features, and the need for human verification to ensure contextual accuracy.

Al-Mukhallafi [32] examined effective strategies for integrating AI in English education from the perspective of 44 undergraduate students at Northern Border University in Saudi Arabia, using a descriptive-analytical approach. Data collected via a 40-item questionnaire revealed that, despite students' strong awareness of AI's effectiveness in developing

© 2026 NSP
Natural Sciences Publishing Cor.

the four language skills, actual classroom use of AI was extremely low. Based on the identified needs, the study proposed an integration framework that covers objectives, content, and evaluation methods.

Al-Tamimi [33] investigated the integration of AI tools, such as ChatGPT, into English as a foreign language (EFL) classrooms at the University College in Haql, University of Tabuk, Saudi Arabia. Using a qualitative thematic analysis of semi-structured interviews with 12 lecturers, the study found that AI significantly enhances interactive learning, self-regulated learning, language fluency, and student confidence. However, integration faces major obstacles: limited access to technology (poor infrastructure, weak internet, outdated devices), a lack of systematic professional training, and difficulties in aligning AI tools with traditional curricula.

Qandil [1] evaluated the effectiveness of AI in teaching Arabic to non-native speakers through a descriptive-analytical review of the academic literature and existing smart applications. The study showed that AI technologies (machine learning, text-to-speech, virtual assistants, VR) offer promising opportunities for interactive, personalized, and simulated human interaction. However, integration is hindered by pedagogical challenges (lack of natural interaction, difficulties with cultural adaptation, and insufficient teacher training) and technical barriers (poor Arabic natural language processing due to structural complexities, limited Arabic AI content, and weak digital infrastructure).

The current study aligns with previous research on the integration of AI tools in education, focusing on their effectiveness in developing basic language skills and identifying shared obstacles, such as poor infrastructure and a lack of professional training. However, the current study differs in specialization and methodology. While [32,33,34] focused on English education, and Qandil examined Arabic for non-native speakers, this study uniquely investigates Arabic instruction for native speakers. Additionally, it targets teachers' self-reported integration of AI in their classroom practice across the four language skills, differentiating itself from prior research that relied mainly on qualitative methods, student perspectives, or different analytical domains (planning, execution, and evaluation).

Critically, this literature helped contextualize the study's problem and build its theoretical framework, particularly regarding AI tool classifications and their impact on the four language skills. Previous findings also informed the development of practical recommendations to overcome implementation barriers. Furthermore, insights into the structural complexities of Arabic and suggestions for language correction and enrichment tools directly assisted the researcher in accurately developing and refining the current study's instruments to ensure their suitability for Arabic language teachers .

1.1 Study Problem and Questions

The contemporary educational field is rapidly shifting toward integrating AI technologies to move beyond traditional teaching methods and create innovative, interactive learning environments. Despite this significance, practical implementation reveals a distinct gap in teachers' ability to use these applications effectively. Research indicates that teachers' integration of AI across various instructional domains (planning, execution, and evaluation) remains generally moderate, reflecting a lack of comprehensive integration in daily practices and highlighting the ongoing need to investigate levels of AI integration [35].

This issue becomes more complex and challenging within the context of Arabic language instruction. Despite AI's promising potential to support language acquisition, teachers face substantial obstacles that hinder its effective integration into the classroom. Key barriers include a lack of specialized technical knowledge, infrastructure deficits, low motivation, and apprehensions about adopting AI as an instructional partner [36,37].

Accordingly, the problem of the current study lies in exploring this gap and determining "the degree of Arabic language teachers' self-reported integration of AI applications in teaching language skills to their students," to assess current practices and offer solutions that balance linguistic authenticity with digital modernization.

1.2 Research Questions

© 2026 NSP
Natural Sciences Publishing Cor.

1. What is the degree to which Arabic language teachers employ artificial intelligence applications in teaching language skills to their students across the domains combined, and for each domain (listening, speaking, reading, and writing)?
2. Are there statistically significant differences at the α=.05 level in the degree to which Arabic language teachers employ AI applications in teaching language skills to their students due to the variables of gender, school type, educational qualification, and years of experience?

1.3 Significance of the Study

1.3.1 Theoretical Significance

The theoretical significance of this study stems from its alignment with rapid advances in education and technology. It enriches contemporary educational literature with a grounded framework that links AI technologies to Arabic pedagogy—an area that remains in critical need of robust research. Furthermore, it contextualizes the cognitive relationship between technology and language, demonstrating how natural language processing, speech recognition, and interactive applications can enhance students' listening, speaking, reading, and writing skills. Finally, it provides a solid foundation for future experimental research and equips the educational library with a validated questionnaire for evaluating teachers' digital practices in language classrooms.

1.3.2 Practical Significance

The practical significance of this study lies in providing direct, actionable insights for key educational stakeholders. It serves as a self-assessment tool for Arabic language teachers and supervisors to reflect on their actual practices and guide the innovative integration of AI to personalize instruction. Additionally, it equips curriculum developers with realistic data to re-engineer Arabic curricula by embedding interactive digital activities. It also assists professional development centers in identifying teachers' true techno-pedagogical needs and in designing specialized training programs that bridge skill gaps. Finally, the study provides educational leaders and decision-makers with precise data on field readiness, supporting policy-making and the strategic allocation of technical and logistical support to build smart learning environments.

1.4 Operational Definitions

Artificial Intelligence (AI) Applications: These refer to software, digital platforms, and generative or interactive systems powered by AI (e.g., ChatGPT, Gemini). They include automated grammar correction tools, adaptive learning platforms, and smart educational content generators, all of which Arabic language teachers use to design instructional scenarios, generate texts, or provide immediate feedback on students' language skills. Operationally, the use of these applications is measured by the teacher's self-reported score on the questionnaire items.

Arabic Language Teachers: This group comprises all male and female specialized teachers who are actively instructing secondary and basic stage students in schools affiliated with the Irbid Third Directorate of Education under the Jordanian Ministry of Education during the 2025/2026 academic year. They represent the study sample participants who responded to the instrument designed to measure their integration of AI in teaching Arabic language skills.

Arabic Language Skills: This term encompasses the integrated system of the four core language skills included in developed Arabic curricula, divided into receptive skills (listening and reading) and productive skills (speaking, as well as automated and creative writing), alongside their associated linguistic, rhetorical, and critical concepts that teachers aim to develop using modern technologies. Operationally, these represent the specific language performance domains outlined in the study instrument and are measured by the teacher's self-reported score for AI integration in instruction.

Degree of AI Integration: This refers to the self-reported frequency with which Arabic language teachers deploy AI applications in teaching the four Arabic language skills (listening, speaking, reading, and writing). Quantitatively, this level is determined by the arithmetic mean of the study sample's responses to the questionnaire items, measured on the study's five-point Likert scale. Throughout this manuscript, "AI applications" denotes the tools themselves, whereas "AI integration" denotes the self-reported degree to which teachers use them; the measure therefore reflects teachers' reported practices rather than observed classroom behavior or instructional effectiveness.

© 2026 NSP
Natural Sciences Publishing Cor.

1.5 Study Scope and Limitations

The framework, scope, and limitations within which this study was conducted are as follows:

- **Objective Scope:** Restricted to identifying the degree to which Arabic language teachers employ AI applications in teaching core language skills (listening, speaking, reading, and writing) to their students.
- **Human Scope:** Limited to male and female Arabic language teachers in public and private schools.
- **Spatial Scope:** Fieldwork was conducted at schools affiliated with the Irbid Third Directorate of Education in the Hashemite Kingdom of Jordan.
- **Temporal Scope:** The study instrument was administered, and field data were collected during the second semester of the 2025/2026 academic year.

The findings of this study are constrained by the psychometric properties of the developed questionnaire—its validity, reliability, and ability to accurately measure the study domains—and, accordingly, the generalizability of its results is therefore limited . In addition, the findings rest on a convenience sample and on teachers' self-reports, and they should not be read as evidence of observed classroom practice.

2 Methodology and Procedures

2.1 Research Methodology

The study adopted a descriptive-analytical approach, as it was the most suitable method for answering the research questions, analyzing the findings, and formulating appropriate recommendations.

2.2 Study Population and Sample

The study population consisted of all 188 male and female Arabic language teachers working in public and private schools in the two districts (Liwas) served by the Irbid Third Directorate of Education in Jordan. According to the staffing schedule for the 2026/2027 academic year, these comprised 157 public school teachers (72 male and 85 female) and 31 private school teachers. The directorate includes 54 public schools (in addition to 5 second-shift schools for Syrian students) and 18 private schools (5 of them kindergartens). From this target population, a sample of 135 teachers was recruited using a non-probability convenience sampling procedure based on institutional accessibility (71.8% of the population of 188 teachers). The sample closely mirrored the population by school type: 113 of 157 public school teachers (72.0%) and 22 of 31 private school teachers (71.0%) responded, so public and private teachers formed 83.7% and 16.3% of the sample versus 83.5% and 16.5% of the population (goodness-of-fit $\chi^2(1) = 0.004$, $p = .95$). The unequal group sizes therefore reflect the structure of the population rather than selective participation by school type. To collect data efficiently, an online questionnaire was created using Google Forms. The survey link was systematically distributed to school principals across the directorate, who then forwarded it to eligible Arabic language teachers within their respective schools. Participation was voluntary, and teachers completed the self-administered form electronically. Because the questionnaire was anonymous, no school identifiers were recorded; teachers' nesting within schools therefore could not be modeled (see Limitations). Group sizes were markedly unequal for school type (public n = 113; private n = 22) and educational qualification (Bachelor's n = 106; postgraduate n = 29), and the study's sensitivity to effects of different sizes is reported in Table 14. Table 1 presents the detailed distribution of the study sample across its independent variables.

2.3 Study Instrument and Psychometric Properties

2.3.1 Study Instrument

To achieve the objectives of this study, a questionnaire was developed to measure Arabic language teachers' self-reported degree of AI integration in teaching Arabic language skills. The instrument comprises 26 items distributed across four main domains: listening (6 items), speaking (6 items), reading (7 items), and writing (7 items). Each item describes a

© 2026 NSP
Natural Sciences Publishing Cor.

Table 1: *Distribution of Study Participants According to Study Variables*

Variables	Categories	Frequency (*n*)	Percentage (%)
Gender	Male	67	49.6
	Female	68	50.4
School Type	Private	22	16.3
	Public	113	83.7
Educational Qualification	Bachelor's	106	78.5
	Postgraduate	29	21.5
Years of Experience	Less than 5 years	57	42.2
	5–10 years	51	37.8
	More than 10 years	27	20.0
Total		**135**	**100.0**

teaching practice in the first person (e.g., "I use…"), and respondents rated it on a five-point Likert scale; the instrument therefore captures self-reported practices rather than observed behavior. The framework, domains, and individual items were constructed based on an extensive review of the theoretical literature and previous research on AI in Arabic language instruction, most notably the studies by Yatri et al. [31], Al-Abadi [30], and Al-Mukhallafi [32]. The design process was further informed by consulting specialists in educational technology, artificial intelligence, and Arabic language pedagogy, followed by comprehensive validity and reliability testing [34]

2.3.2 Instrument Validity

To ensure face and content validity, the initial questionnaire was evaluated by a panel of 11 expert arbitrators, including university faculty members specializing in Arabic language curricula, educational technology, artificial intelligence, educational measurement, and curriculum and instruction, as well as experienced Arabic language supervisors and teachers. The experts assessed the clarity, linguistic correctness, and contextual appropriateness of the items. Their feedback was used to refine the tool by adding, deleting, or modifying items, including removing certain language-related items, correcting linguistic and typographical errors, eliminating redundancies, and adjusting punctuation to produce the final version of the questionnaire.

2.3.3 Construct Validity

The construct validity of the 26-item questionnaire was evaluated through a structured theoretical and empirical process. The initial item pool was developed by operationalizing actionable AI pedagogical indicators—derived from the theoretical framework and literature (e.g., Al-Mukhallafi, 2020; Yatri et al., 2023; Al-Abadi, 2024)—into specific teaching practices across the four language skills: Listening (6 items), Speaking (6 items), Reading (7 items), and Writing (7 items). To empirically evaluate the scale's internal structure, item-to-total, item-to-domain, and inter-domain Pearson correlation coefficients were calculated. This pilot analysis was conducted using an independent sample of 20 Arabic language teachers selected from outside the primary study sample. As detailed in Table 2, item-to-total correlation coefficients ranged from .45 to .82 ($p < .05$), and item-to-domain correlation coefficients ranged from .47 to .91 ($p < .05$). Because this pilot sample was small, these coefficients are imprecise and are treated as preliminary evidence.

All correlation coefficients were statistically significant, indicating that each item contributed adequately to its respective domain and to the overall scale in the pilot sample, so no items were deleted. In addition, correlation coefficients among the scale domains and between each domain and the total score were computed, as presented in Table 3.

As shown in Table 3, inter-domain correlations were positive and statistically significant, ranging from .457 to .895, while domain-to-total correlations ranged from .757 to .886. According to Cohen's (1988) conventions, correlations of this size ($r \geq .50$) are large, indicating that the domains are strongly interrelated. However, the intercorrelation between the Listening and Speaking domains is particularly high ($r = .895$). This indicates a substantial functional and conceptual overlap between oral receptive and productive skills when integrated into digital AI learning contexts. Given the pilot sample size ($N = 20$), Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were not conducted due to the item count relative to the sample size. Therefore, these correlation metrics are interpreted strictly as preliminary evidence of internal coherence and convergent validity rather than structural construct validity.

© 2026 NSP
Natural Sciences Publishing Cor.

Table 2: *Item–Total and Item–Domain Correlation Coefficients for the Scale*

Item No.	Correlation with Domain	Correlation with Total Scale	Item No.	Correlation with Domain	Correlation with Total Scale	Item No.	Correlation with Domain	Correlation with Total Scale
1	.91**	.79**	10	.72**	.63**	19	.55*	.45*
2	.76**	.74**	11	.51*	.63**	20	.75**	.82**
3	.71**	.58**	12	.73**	.55*	21	.55*	.65**
4	.51*	.65**	13	.79**	.63**	22	.47*	.74**
5	.84**	.81**	14	.73**	.65**	23	.61**	.57**
6	.75**	.56**	15	.70**	.76**	24	.61**	.63**
7	.62**	.62**	16	.62**	.45*	25	.55*	.57**
8	.72**	.58**	17	.70**	.51*	26	.77**	.70**
9	.72**	.51*	18	.89**	.61**			

Note. $p < .05$ (*); $p < .01$ (**).

Table 3: *Correlation Coefficients Among the Scale Domains and Between Each Domain and the Total Scale Score*

Domain	**Listening Skill**	**Speaking Skill**	**Reading Skill**	**Writing Skill**	**Total Scale Score**
Listening Skill	1				
Speaking Skill	.895**	1			
Reading Skill	.457*	.559*	1		
Writing Skill	.459*	.526*	.818**	1	
Total Scale Score	.855**	.886**	.757**	.808**	1

Note. $p < .05$ (*); $p < .01$ (**).

2.3.4 Reliability of the Study Instrument

The study instrument's reliability was established using the test–retest method. The scale was administered twice, with a two-week interval, to a pilot sample of 20 participants who were not included in the main study. Pearson's correlation coefficient was calculated between the two administrations.

Internal consistency reliability was also assessed using Cronbach's alpha, computed for the pilot sample (n = 20). Table 4 presents the test–retest and Cronbach's alpha coefficients for the scale domains and the total scale score. Coefficients of .70 or above were considered acceptable [39,40].

Table 4: *Cronbach's Alpha and Test–Retest Reliability Coefficients for the Scale Domains and the Total Scale Score (Pilot Sample, n = 20)*

Domain	**Test–Retest Reliability (Pilot, n = 20)**	**Cronbach's Alpha (Pilot, n = 20)**
Listening Skill	.86	.82
Speaking Skill	.85	.79
Reading Skill	.82	.77
Writing Skill	.84	.81
Total Scale Score	.92	.87

Note. Test–retest coefficients and Cronbach's alphas are based on a pilot sample of 20 teachers.

Table 4 shows that the instrument demonstrated good reliability. Cronbach's alpha coefficients ranged from .77 to .87, while test–retest reliability coefficients ranged from .82 to .92, indicating good reliability. Because these estimates rest on a pilot sample of only 20 teachers, they should be regarded as preliminary.

2.4 Study Procedures

The study was conducted according to the following procedures:

- Reviewing the relevant theoretical literature and previous studies.
- Developing the study questionnaire and establishing its validity and reliability.

© 2026 NSP
Natural Sciences Publishing Cor.

- Obtaining approval to administer the questionnaire to Arabic language teachers in the Third Directorate of Education in Irbid.
- Creating an online version of the questionnaire using Google Forms and distributing the link to participants through school principals, who forwarded it to teachers (e.g., via WhatsApp).
- Inviting Arabic language teachers to complete the questionnaire. Data collection was conducted over three weeks.
- Providing participants with sufficient time to respond and informing them that their responses would be used solely for research purposes and treated confidentially.
- Using a five-point Likert scale to score the questionnaire responses. Mean scores were interpreted as follows: 1.00–2.32 = Low; 2.33–3.66 = Moderate; 3.67–5.00 = High.
- Collecting the data using Microsoft Excel and analyzing it with the Statistical Package for the Social Sciences (SPSS).
- Answering the research questions by presenting and discussing the findings and providing recommendations.
- Preparing the final research report.

2.5 Study Variables

2.5.1 Independent Variables

- **Gender:** includes two categories (Male and Female).
- **Educational Qualification:** includes two categories (Bachelor's and Postgraduate).
- **Years of Experience:** includes three categories (Less than 5 years, 5-10 years, More than 10 years).
- **School Type:** includes two categories (Public and Private).

2.5.2 Dependent Variables

The primary dependent variable is the self-reported degree to which Arabic language teachers employ artificial intelligence applications in instruction, which comprises four core domains:

- Listening skill
- Speaking skill
- Reading skill
- Writing skill

3 Study Results and Discussion

3.1 Results of the First Question

What is the degree to which Arabic language teachers employ artificial intelligence applications in teaching language skills to their students across the domains combined, and for each domain (listening, speaking, reading, and writing)?

To answer this question, mean scores, standard deviations, and 95% confidence intervals (Table 12) were calculated for the self-reported degree to which Arabic language teachers employ AI applications in teaching language skills to their students. The results are presented in Table 5.

Table 5: *Mean Scores and Standard Deviations for the Degree of Self-Reported AI Integration by Arabic Language Teachers in Descending Order*

Rank	No.	Domain	Mean	*SD*	Level
1	3	Reading Skill	3.60	0.62	Moderate
2	4	Writing Skill	3.42	0.64	Moderate
3	2	Speaking Skill	3.13	0.81	Moderate
4	1	Listening Skill	2.75	1.01	Moderate
Total		**Overall Score**	**3.25**	**0.64**	**Moderate**

Note. SD = standard deviation. Levels were defined by the mean score: Low = 1.00–2.32, Moderate = 2.33–3.66, High = 3.67–5.00.

© 2026 NSP
Natural Sciences Publishing Cor.

Table 5 demonstrates that the study sample's evaluation of the degree of self-reported AI integration among Arabic language teachers was overall moderate, with a mean (*M*) of 3.25 and a standard deviation (*SD*) of 0.64, 95% confidence interval (CI) [3.14, 3.36]. For individual domains, the mean scores ranged from 2.75 to 3.60. Reading skill ranked first with the highest mean of 3.60 (SD=0.62, Moderate level), while listening skill ranked last with a mean of 2.75 (SD=1.01, Moderate level). The reading mean (95% CI [3.49, 3.71]) lies close to the cut-off for the High level (3.67), whereas the listening means (95% CI [2.58, 2.92]) lies well inside the Moderate band. All results refer to teachers' self-reported practices, not to observed classroom behavior.

These findings may reflect the ongoing transition in AI adoption in education; the explanations offered below are interpretive, because barriers such as training, infrastructure, and attitudes were not measured in this study. While teachers demonstrate a clear interest in introducing modern technologies into their classrooms, a lack of specialized training in integrating these tools into Arabic language lessons, reported in earlier studies [38, 17, 1], may hinder optimal implementation. Consequently, current use may depend largely on individual initiative, which would be consistent with a moderate overall level.

This may help explain the higher ranking of reading and writing skills. Current AI software is particularly strong at processing textual data, enabling teachers to quickly generate reading comprehension passages, apply diacritics (Tashkeel), format textual structures, and construct contextual comprehension questions. Similarly, text-based software enables automatic spelling and grammar corrections, substantially reducing teachers' workload and time investment.

Conversely, speaking and listening skills ranked lowest, possibly because of practical operational challenges. Many schools may lack dedicated language laboratories and appropriate audio equipment [33, 1]. Furthermore, AI applications that analyze spoken Modern Standard Arabic (MSA) may still lack precision and may struggle to process distinct regional dialects accurately. Teachers may therefore lean away from audio-centric AI tools, preferring traditional instruction methodologies for oral skills to maintain classroom discipline and minimize student distraction.

The findings of this study align with those of Hamadneh [27] and Mureed [28], who both concluded that teachers' use of artificial intelligence is moderate. Such comparisons should be made cautiously, however, because the studies differ in instruments, respondents, and domains (e.g., Hamadneh, [27], measured planning, lesson execution, and evaluation, whereas the present study measured four language skills). The contrast with Al-Mukhallafi [32], who reported low implementation, may be due to the respondents (university students rather than teachers) and to the timing of data collection, which preceded the wide availability of generative AI tools such as ChatGPT.

To gain deeper insights into these rankings, mean scores and standard deviations were calculated for the items within each domain individually, as detailed below.

3.1.1 First: Listening Skill

Table 6: *Mean Scores and Standard Deviations for Listening Skill Items Ranked in Descending Order*

Rank	No.	Items	Mean	*SD*	Level
1	4	I employ interactive AI platforms to generate comprehension questions and provide immediate feedback after listening.	2.81	1.24	Moderate
2	1	I use Text-to-Speech (TTS) technologies to provide exemplary, error-free reading models.	2.80	1.31	Moderate
3	5	I utilize AI to convert chants and poetic texts into engaging audio materials enhanced with sound effects.	2.78	1.44	Moderate
4	6	I use AI tools to help students extract main and supporting ideas and access facts and data from the text they listen to.	2.78	1.20	Moderate
5	3	I rely on AI software to adjust the speed and tone of the texts I listen to, improving my listening comprehension.	2.68	1.40	Moderate
6	2	I adapt AI applications to generate diverse listening texts that accommodate individual differences among students.	2.65	1.27	Moderate
Total		**Listening Skill (Overall)**	**2.75**	**1.01**	**Moderate**

Note. SD = standard deviation. Levels: Low = 1.00–2.32, Moderate = 2.33–3.66, High = 3.67–5.00.

Table 6 shows that the study sample's evaluation of the listening skill domain was moderate, with a mean score of 2.75 (SD = 1.01). Item mean scores ranged between 2.65 and 2.81. Item 4—"I employ interactive AI platforms to generate comprehension questions and provide immediate feedback after listening"—ranked first with a mean of 2.81 (SD=1.24, Moderate level). Item 1—"I use Text-to-Speech (TTS) technologies to provide exemplary, error-free reading models"—ranked second with a mean of 2.80 (SD=1.31, Moderate level). Item 5—"I utilize AI to convert chants and poetic texts into engaging audio materials enhanced with sound effects"—ranked third with a mean of 2.78 (SD=1.44, Moderate level).

© 2026 NSP
Natural Sciences Publishing Cor.

Conversely, Item 2—"I adapt AI applications to generate diverse listening texts that accommodate individual differences among students"—ranked last, yielding a mean of 2.65 (SD=1.27, Moderate level).

The convergence of listening skill items around a moderate evaluation could be attributed to the specialized audio-dependent nature of this skill. One plausible explanation is that teachers prioritize straightforward, operational AI applications, such as generating post-listening comprehension questions or utilizing pre-recorded text-to-speech outputs as alternatives to traditional teacher-led reading. These direct tools support classroom management with minimal technical friction.

Conversely, more complex listening instruction procedures—such as dynamically adjusting audio speed and tone or generating multiple differentiated listening texts tailored to individual student capabilities—ranked lower. Designing customized audio content for distinct student groups likely requires extensive preparation and advanced technical expertise. Furthermore, standard classrooms might lack specialized audio hardware required to deliver multiple differentiated audio materials simultaneously, which could lead educators to opt for standardized, pre-packaged digital tools that cater to the classroom as a single collective unit.

These moderate findings are consistent with Al-Tamimi [33], who highlighted that operational difficulties associated with integrating smart audio applications into conventional curricula, paired with a limited availability of systematic professional development for educators, may drive teachers to rely on basic text generation or pre-formulated questionnaire elements rather than engineering customized acoustic lesson frameworks.

3.1.2 Second: Speaking Skill

Table 7: *Mean Scores and Standard Deviations for Speaking Skill Items Ranked in Descending Order*

Rank	No.	Items	Mean	*SD*	Level
1	7	I direct students to use conversational AI applications to engage in dialogue in Modern Standard Arabic to develop speaking fluency.	3.26	1.11	Moderate
2	9	I leverage AI tools to train students in public speaking and recitation through simulation and virtual environments.	3.25	1.31	Moderate
3	11	I use AI to design virtual communicative situations, such as digital role-playing, to stimulate students' oral expression.	3.20	1.01	Moderate
4	12	I use smart performance reports from applications to identify strengths and weaknesses in each student's oral linguistic performance.	3.04	1.30	Moderate
5	10	I rely on Natural Language Processing (NLP) to automatically correct grammatical and structural errors students make while speaking.	3.01	1.19	Moderate
6	8	I employ Speech Recognition applications to evaluate students' pronunciation accuracy and articulation.	3.01	1.31	Moderate
Total		**Speaking Skill (Overall)**	**3.13**	**0.81**	**Moderate**

Note. SD = standard deviation. Levels: Low = 1.00–2.32, Moderate = 2.33–3.66, High = 3.67–5.00.

Table 7 shows that the study sample's evaluation of the speaking skill domain was moderate, with a mean score of 3.13 (SD = 0.81). Item mean scores ranged between 3.01 and 3.26. Item 7—"I direct students to use conversational AI applications to conduct dialogues in Modern Standard Arabic to develop speaking fluency"—ranked first with a mean of 3.26 (SD=1.11, Moderate level). Item 9—"I leverage AI tools to train students in public speaking and recitation through simulation and virtual environments"—ranked second with a mean of 3.25 (SD=1.31, Moderate level). Item 11—"I use AI to design virtual communicative situations, such as digital role-playing, to stimulate students' oral expression"—ranked third with a mean of 3.20 (SD=1.01, Moderate level). Conversely, Items 10 and 8—concerning reliance on NLP for automatic grammatical correction during speech and the use of speech recognition to evaluate pronunciation accuracy—tied for the final rank, both yielding a mean of 3.01 (SDs of 1.19 and 1.31, respectively; Moderate level).

The moderate overall evaluation in the speaking skill domain may reflect the expressive and productive demands of oral language. A potential explanation for the higher ranking of certain items is that educators may favor directing students toward autonomous practices outside conventional lecture structures. Encouraging students to engage in standard Arabic dialogues with conversational AI tools or to simulate public speaking in virtual environments shifts the verbal load onto students, which might help lower affective filters such as language anxiety without consuming limited instructional time.

In contrast, items requiring high technical precision and direct diagnostic tracking by the teacher—such as utilizing automatic speech recognition to assess phonetic articulation or deploying real-time syntactic corrections—ranked lower. One potential factor is that voice-dependent applications generally require low-noise environments to capture student input cleanly, which may be challenging to achieve in crowded classrooms. Additionally, automated speech recognition

© 2026 NSP
Natural Sciences Publishing Cor.

systems for Modern Standard Arabic may still struggle to accurately process the diverse acoustic profiles of younger learners, which could lead teachers to limit their reliance on them for immediate evaluation.

These findings correspond with Al-Tamimi [33] and Qandil [1], who noted that audio-centric software often faces infrastructural limitations in schools (such as a shortage of tablets and sound laboratories), alongside potential technical constraints of AI models in fully comprehending complex morphological and phonetic structures of Arabic when intertwined with local dialects.

3.1.3 Third: Reading Skill

Table 8: *Mean Scores and Standard Deviations for Reading Skill Items Ranked in Descending Order*

Rank	No.	Items	Mean	SD	Level
1	16	I rely on smart technologies to simplify and summarize classical literary texts, thereby facilitating students' understanding and analysis.	3.67	0.87	High
2	19	I employ generative AI techniques to help students transform reading texts into visual scripts or comic stories, thereby deepening their reading comprehension.	3.65	0.93	Moderate
3	14	I employ automated diacritics (Tashkeel) tools powered by AI to train students on correct oral Reading and proper word-ending adjustments.	3.62	0.98	Moderate
4	18	I use AI to formulate questions that measure higher-level reading comprehension (analytical, critical, and creative).	3.59	0.97	Moderate
5	13	I use text-generation applications such as ChatGPT to create reading texts of varying difficulty that match students' levels.	3.57	1.08	Moderate
6	17	I employ smart digital dictionaries and lexicons to provide students with immediate synonyms, antonyms, and linguistic contexts for new words.	3.54	0.91	Moderate
7	15	I use smart reading platforms to track student reading speed and identify subvocalized or difficult words that hinder reading fluency.	3.54	0.94	Moderate
Total		**Reading Skill (Overall)**	**3.60**	**0.62**	**Moderate**

Note. SD = standard deviation. Levels: Low = 1.00–2.32, Moderate = 2.33–3.66, High = 3.67–5.00.

Table 8 demonstrates that the study sample's evaluation of the reading skill domain was moderate, yielding a mean score of 3.60 (SD = 0.62). Item mean scores ranged between 3.54 and 3.67. Item 16—"I rely on smart technologies to simplify and summarize classical literary texts to facilitate students' understanding and analysis"—ranked first with a mean of 3.67 (SD = 0.87, High level, the only item reaching this level). Item 19—"I employ generative AI techniques to help students transform reading texts into visual scripts or comic stories to deepen their reading comprehension"—ranked second with a mean of 3.65 (SD = 0.93, Moderate level). Item 14—"I employ automated diacritics (Tashkeel) tools powered by AI to train students in correct oral reading and proper word-ending adjustments"—ranked third with a mean of 3.62 (SD = 0.98, Moderate level). Conversely, Items 17 and 15—concerning the use of smart digital dictionaries for immediate vocabulary lookup and the use of smart platforms to track individual reading speed—tied for the final rank, both yielding a mean of 3.54 (SD = 0.91 and 0.94, respectively; Moderate level).

The moderate evaluation across reading-skill items may be linked to AI's capacity for text simplification, visual conversion, and automated diacritization, as well as the structural characteristics of Arabic curricula. Because classical literary texts in Arabic curricula can present comprehension challenges, teachers may find that generative AI offers effective solutions to simplify complex structures and convert them into visual scripts or illustrated stories. Similarly, automated diacritization software provides educators with a rapid method to generate vocalized texts that support oral reading practice.

Conversely, items involving specialized digital dictionaries and reading-speed trackers ranked lower. One possible explanation is that such diagnostic platforms may require premium subscriptions or individual device configurations to track each student's metrics. These logistical requirements could make their implementation more challenging in standard classroom environments than in collective text summarization or diacritization tools.

These results align with Qandil [1] and Yatri et al. [31] regarding the utility of natural language processing tools in facilitating linguistic content. Al-Azzawi [29], in a quasi-experimental study, reported a substantial improvement in students' language skills after AI-supported teaching. That study evaluated instructional effectiveness rather than the extent of teachers' use, so the two sets of findings are not directly comparable; the present results indicate only that teachers' reported use of AI for reading remains within moderate bounds.

© 2026 NSP
Natural Sciences Publishing Cor.

3.1.4 Fourth: Writing Skill

Table 9: *Mean Scores and Standard Deviations for Writing Skill Items Ranked in Descending Order*

Rank	No.	Items	Mean	*SD*	Level
1	23	I train students on "Linguistic Prompt Engineering" to produce creative texts in collaboration with the machine.	3.54	1.03	Moderate
2	21	I use smart idea generators and mind maps to assist students in the "pre-writing" stage and to generate ideas for expression.	3.53	1.09	Moderate
3	22	I rely on AI software to evaluate essay writing and provide constructive hints regarding stylistic quality and linguistic transitions.	3.47	0.97	Moderate
4	26	I utilize AI tools to provide students with immediate feedback on semantic coherence (sentence connectivity) and stylistic diversity in their Writing.	3.47	1.06	Moderate
5	25	I use AI platforms to design precise digital evaluation rubrics for grading students' written work.	3.35	1.00	Moderate
6	24	I use speech-to-text applications (Dictation AI) to enable students to compare their spoken vocabulary with its correct spelling on-screen.	3.29	0.95	Moderate
7	20	I employ smart spell and grammar checkers (predictive AI) to teach students self-correction strategies for errors.	3.28	1.12	Moderate
Total		**Writing Skill (Overall)**	**3.42**	**0.64**	**Moderate**

Note. SD = standard deviation. Levels: Low = 1.00–2.32, Moderate = 2.33–3.66, High = 3.67–5.00.

Table 9 shows that the study sample's evaluation of the writing-skill domain was moderate, with a mean score of 3.42 (SD = 0.64). Item mean scores ranged between 3.28 and 3.54. Item 23—"I train students on 'Linguistic Prompt Engineering' to produce creative texts in collaboration with the machine"—ranked first with a mean of 3.54 (SD=1.03, Moderate level). Item 21—"I use smart idea generators and mind maps to assist students in the 'pre-writing' stage and to generate ideas for expression"—ranked second with a mean of 3.53 (SD=1.09, Moderate level). Items 22 and 26—concerning the use of AI to evaluate essays and providing immediate feedback on semantic coherence—tied for the third rank, both yielding a mean of 3.47 (SD=0.97 and 1.06, respectively, Moderate level). Conversely, Item 20—"I employ smart spell and grammar checkers (predictive AI) to teach students self-correction strategies for errors"—ranked last with a mean of 3.28 (SD=1.12, Moderate level).

The preference for prompt engineering and idea generation may reflect teachers' recognition of the value of interactive AI tools during the pre-writing stages. Brainstorming tools and prompt structuring can help students outline ideas and overcome initial writing difficulties, making these collaborative practices relatively easy to incorporate into group instruction.

In contrast, using smart spell and grammar checkers for self-correction strategies ranked last. One plausible explanation is that educators may be concerned about "digital dependency," fearing that overreliance on automated proofreading tools could inadvertently weaken foundational spelling and grammar skills. Additionally, integrating individualized correction tools might require continuous one-on-one supervision, which instructional time constraints and traditional assessment formats can limit.

These findings align with Yatri et al.(2023) on teachers' reliance on AI for ideation, outlining, and creative drafting. They also reflect the concerns discussed by Mureed [28] and Yatri et al. [31] regarding potential over-reliance on automated proofreaders. Al-Abadi [13], in turn, found in an experimental setting that generative AI applications (Catgut and Sahihely) significantly improved students' written expression; because that study tested effectiveness rather than teachers' reported use, it is not directly comparable with the present findings.

3.2 Results of the Second Question

Are there statistically significant differences at the α*=.05 level in the degree to which Arabic language teachers employ AI applications in teaching language skills to their students due to the variables of gender, school type, educational qualification, and years of experience?*

To answer this question, descriptive statistics (means, standard deviations, and 95% confidence intervals; Table 12) were calculated based on gender, school type, educational qualification, and years of experience, as presented in Table 10.

© 2026 NSP
Natural Sciences Publishing Cor.

Table 10: *Mean Scores and Standard Deviations for the Degree of Self-Reported AI Integration by Arabic Language Teachers According to Study Variables*

Variable	Category	Mean	*SD*	*n*
Gender	Male	3.17	.588	67
	Female	3.32	.689	68
School Type	Private	3.58	.694	22
	Public	3.18	.614	113
Educational Qualification	Bachelor's	3.15	.623	106
	Postgraduate	3.58	.609	29
Years of Experience	Less than 5 years	3.16	.556	57
	5–10 years	3.25	.747	51
	More than 10 years	3.40	.591	27

Note. M = mean; SD = standard deviation; n = number of teachers. Means refer to the overall self-reported AI integration score (1–5).

Table 10 presents apparent differences in mean scores and standard deviations across the categories of gender, school type, educational qualification, and years of experience. To ascertain the statistical significance of these observed differences, a main-effects four-factor univariate analysis of variance (ANOVA) was conducted. Because the design was unbalanced (e.g., 113 public vs. 22 private school teachers), the contrasts for the significant factors were additionally checked with Welch's unequal-variance t tests (Table 13).

Prior to evaluating the inferential model, key statistical assumptions were systematically verified. Observations were treated as independent because each teacher responded individually and anonymously; however, teachers were recruited through principals and may be clustered within schools, which could not be examined because school identifiers were not collected (see Limitations). The normality of residuals was confirmed via the Shapiro-Wilk test and Q-Q plots. Furthermore, the assumption of homogeneity of variances was met, as Levene's Test of Equality of Error Variances showed no significant differences in group variances, F(15, 119) = 1.34, p = .189; in addition, the ratio of the largest to the smallest group variance across the categories in Table 10 was 1.80 (SDs from .556 to .747), well below common rules of thumb for concern. The ANOVA model intentionally evaluated main effects only, excluding multi-way interaction terms to maintain statistical power and focus on primary demographic predictors; possible interactions (e.g., school type × qualification) were therefore not tested. The inferential findings are presented in Table 11.

Table 11: *Main-Effects Four-Factor Univariate ANOVA for the Degree of Self-Reported AI Integration Across Demographic Factors*

Source of Variation	Sum of Squares	*Df*	Mean Square	*F*	*p* (Sig.)	Partial η^2
Gender	.000	1	.000	.000	.985	.000
School Type	2.837	1	2.837	7.901	.006*	.058
Educational Qualification	3.275	1	3.275	9.120	.003*	.066
Years of Experience	.976	2	.488	1.359	.261	.021
Error	46.324	129	.359			
Corrected Total	55.328	134				

Note. N = 135. Dependent variable: degree of self-reported AI integration. df = degrees of freedom; η_p^2 = partial eta squared (.01, .06, and .14 are conventionally read as small, medium, and large effects; [46]. Corrected Total = total corrected sum of squares. *$p < .05$.

As demonstrated in Table 11, the main-effects ANOVA revealed the following results regarding demographic predictors:

- **Gender:** No statistically significant main effect was observed (α = .05), $F(1, 129) < 0.001$, p = .985, $\eta_p^2 < .001$, indicating that male and female teachers reported comparable levels of AI integration (female M = 3.32, male M = 3.17; Welch $t(130.4) = 1.36$, p = .176).
- **School Type:** A statistically significant main effect was detected (α = .05), $F(1, 129) = 7.901$, p = .006, η_p^2 = .058, 95% CI [.005, .146], a small-to-medium effect. Private school teachers reported higher AI integration (M = 3.58, SD = 0.69) than public school teachers (M = 3.18, SD = 0.61); the mean difference was 0.40, 95% CI [0.07, 0.73], Welch $t(27.8) = 2.52$,p = .018, Hedges' g = 0.63, 95% CI [0.17, 1.10]. Because the groups were markedly unequal (22 vs. 113), this effect should be interpreted cautiously.
- **Educational Qualification:** A statistically significant main effect was found (α = .05), $F(1, 129) = 9.120$, p = .003, η_p^2 = .066, 95% CI [.008, .157], a medium effect. Teachers holding postgraduate degrees (M = 3.58, SD = 0.61)

© 2026 NSP
Natural Sciences Publishing Cor.

reported higher AI integration than those with Bachelor's degrees (M = 3.15, SD = 0.62); the mean difference was 0.43, 95% CI [0.17, 0.69], Welch t (45.3) = 3.35, p = .002, Hedges' g = 0.69, 95% CI [0.27, 1.11].

- **Years of Experience:** No statistically significant main effect was found (α = .05), $F(2, 129) = 1.359$, p = .261, η_p^2 = .021, 95% CI [.000, .079]. Tukey post hoc comparisons (Table 13) showed no significant difference between any pair of experience groups, suggesting that, in this sample, teaching experience was not meaningfully related to AI integration.

3.2.1 Confidence Intervals, Effect Sizes, Post Hoc Comparisons, and Statistical Power

To complement the omnibus tests, 95% confidence intervals, effect sizes, post hoc comparisons, and a sensitivity power analysis are reported. Where raw data were not required, these statistics were derived from the group means, standard deviations, and sample sizes in Tables 5 and 10 and from the error mean square in Table 11; they therefore describe the observed (unadjusted) group means.

Table 12: *95% Confidence Intervals for the Domain Means and for the Overall Mean by Study Variable*

Domain / group	*n*	*M*	*SD*	95% CI
Listening skill	135	2.75	1.01	[2.58, 2.92]
Speaking skill	135	3.13	0.81	[2.99, 3.27]
Reading skill	135	3.60	0.62	[3.49, 3.71]
Writing skill	135	3.42	0.64	[3.31, 3.53]
Overall AI integration	135	3.25	0.64	[3.14, 3.36]
Gender: Male	67	3.17	0.59	[3.03, 3.31]
Gender: Female	68	3.32	0.69	[3.15, 3.49]
School type: Private	22	3.58	0.69	[3.27, 3.89]
School type: Public	113	3.18	0.61	[3.07, 3.29]
Qualification: Bachelor's	106	3.15	0.62	[3.03, 3.27]
Qualification: Postgraduate	29	3.58	0.61	[3.35, 3.81]
Experience: Less than 5 years	57	3.16	0.56	[3.01, 3.31]
Experience: 5–10 years	51	3.25	0.75	[3.04, 3.46]
Experience: More than 10 years	27	3.40	0.59	[3.17, 3.63]

Note. M = mean; SD = standard deviation; CI = confidence interval (t-based). Domain rows are based on all 135 teachers; group rows refer to the overall self-reported AI integration score.

As Table 12 shows, the 95% confidence interval for the overall mean (3.14–3.36) lies entirely within the Moderate band, whereas the intervals for several subgroup means are wide, particularly for private school teachers ([3.27, 3.89], n = 22) and teachers with more than 10 years of experience ([3.17, 3.63], n = 27). This indicates considerable imprecision in the subgroup estimates. The intervals for Bachelor's degree holders ([3.03, 3.27]) and postgraduate degree holders ([3.35, 3.81]) do not overlap, whereas those for public and private schools overlap only marginally.

The school-type (η_p^2 = .058) and qualification (η_p^2 = .066) effects fall in the small-to-medium range, but their lower confidence limits (.005 and .008) show that the true effects could be very small. The effects of gender and experience effects were negligible to small.

Table 13: *Pairwise Contrasts in Self-Reported AI Integration: Welch Tests for Two-Group Factors and Tukey Post Hoc Comparisons for Years of Experience*

Contrast	ΔM	95% CI	Test	p	Effect size
Private − Public	0.40	[0.07, 0.73]	Welch $t(27.8) = 2.52$	.018	0.63 [0.17, 1.10]
Postgraduate − Bachelor's	0.43	[0.17, 0.69]	Welch $t(45.3) = 3.35$	.002	0.69 [0.27, 1.11]
Female − Male	0.15	[−0.07, 0.37]	Welch $t(130.4) = 1.36$	.176	0.23 [−0.11, 0.57]
5–10 years − Less than 5 years	0.09	[−0.18, 0.36]	Tukey HSD	.716	0.15
More than 10 years − Less than 5 years	0.24	[−0.09, 0.57]	Tukey HSD	.204	0.40
More than 10 years − 5–10 years	0.15	[−0.19, 0.49]	Tukey HSD	.546	0.25

Note. ΔM = difference in observed means. Effect size is Hedges' g (with 95% CI) for the Welch contrasts and Cohen's d (based on the ANOVA error mean square) for the Tukey comparisons. Tukey comparisons use the cell means in Table 10 and the error mean square in Table 11 (MSE = 0.359, df = 129); the p -values are adjusted for three comparisons.

Because the omnibus test for years of experience was not significant, the post hoc comparisons are reported only for completeness; none approached significance (adjusted p = .20–.72). The Welch tests, which do not assume equal variances

© 2026 NSP
Natural Sciences Publishing Cor.

or equal group sizes, agree with the ANOVA for school type and educational qualification, which supports their robustness to the unbalanced group sizes; the school-type contrast, however, has a wide confidence interval ([0.07, 0.73]) because only 22 private school teachers took part.

Table 14: *Sensitivity Power Analysis for the Main Tests (α = .05, N = 135)*

Test	Statistic	Minimum detectable effect (80% power)	Power for a medium effect
Gender, school type, and qualification (main effects)	$F(1, 129)$	f = .24 (η_p^2 = .056)	.822
Years of experience (main effect)	$F(2, 129)$	f = .27 (η_p^2 = .068)	.731
Private vs. public (22 vs. 113)	Two-sample t	d = 0.66	.568
Postgraduate vs. Bachelor's (29 vs. 106)	Two-sample t	d = 0.59	.659
Female vs. male (68 vs. 67)	Two-sample t	d = 0.49	.822

Note. Power for a medium effect refers to f = .25 for the F tests and d = 0.50 for the two-sample comparisons. Values were computed from the noncentral F and t distributions.

The sensitivity analysis indicates that, with N = 135, the main-effects tests had 80% power only for effects of about $f \geq .24$ ($\eta_p^2 \geq .056$) and, for years of experience, $f \geq .27$. The study was therefore adequately sensitive to medium effects but not to small ones (power $\approx .211$ for f = .10 with 1 df), and the school-type contrast (22 vs. 113) could reliably detect only differences of $d \geq 0.66$. Non-significant results, especially for years of experience, should therefore not be read as evidence of no difference. Observed ("post hoc") power is not reported because it is a direct function of the p value and adds no information [45].

3.2.2 Correlations among Domains and Demographic Variables

Table 15: *Correlations Between the Demographic Variables and the Total Self-Reported AI Integration Score (N = 135)*

Variable (coding)	*r*	95% CI	*p*
Gender (0 = male, 1 = female)	.12	[−.05, .28]	.176
School type (0 = public, 1 = private)	.23	[.06, .38]	.007
Educational qualification (0 = Bachelor's, 1 = postgraduate)	.28	[.11, .43]	.001
Years of experience (1 = < 5, 2 = 5–10, 3 = > 10)	.14	[−.03, .30]	.114

Note. Pearson correlations (point-biserial for the dichotomous variables) were computed exactly from the group means, standard deviations, and sample sizes in Table 10; the 95% CIs use Fisher's z transformation.

As Table 15 shows, school type (r = .23, p = .007) and educational qualification (r = .28, p = .001) were weakly to moderately and significantly related to the total score, whereas gender (r = .12) and years of experience (r = .14) were not, consistent with the ANOVA results. Correlations among the four domains are reported for the pilot sample in Table 3, where Listening and Speaking were very highly correlated (r = .895); this overlap is acknowledged as a limitation.

3.2.3 Nested Structure of the Data

Teachers are nested within schools, and school type is a school-level characteristic. Because school identifiers were not collected in order to preserve anonymity, infraclass correlations could not be estimated and multilevel (mixed-effects) modeling was not feasible [41]. If teachers within the same school respond similarly, the effective sample size is smaller than 135 and the standard errors for school-level predictors, particularly school type, are probably underestimated. The school-type effect should therefore be treated as tentative and future studies should record school identifiers and model schools as a random effect. The directorate has 54 public (plus 5 second-shift) and 18 private schools, so the 135 respondents came from at most about 77 school units, roughly two teachers per unit if all participated; the number of participating schools is unknown, and with so few teachers per school the inflation of standard errors is probably modest, but this cannot be verified.

The absence of significant differences attributed to gender and years of experience may be understood through a shared instructional environment. Teachers within one directorate generally operate under the same curricular frameworks and regional guidelines and may have comparable access to formal AI training. Moreover, because generative AI applications are a relatively recent technological development, educators with varying levels of experience may share a similar baseline of exposure, leading to comparable implementation practices regardless of gender or career longevity. The absence of

© 2026 NSP
Natural Sciences Publishing Cor.

significant differences should not be read as evidence of equivalence: the power analysis (Table 14) shows that only effects of roughly f $\geq$ .24–.27 could be detected reliably, and the confidence interval for the experience effect (η^2p = .021, 95% CI [.000, .079]) does not exclude a small-to-medium effect.

Conversely, the statistically significant differences favoring private schools and postgraduate degree holders could be interpreted through several contextual factors. A plausible explanation for the school type variable is that private institutions might possess greater technological infrastructure—such as high-speed internet access and available digital devices—and may integrate technology usage into institutional performance expectations. Because school type is a school-level characteristic and teachers could not be linked to their schools, this effect is the most susceptible to clustering and to the unbalanced group sizes and it should be regarded as tentative.

Regarding educational qualifications, teachers with postgraduate degrees (Master's or Ph.D.) may have developed advanced self-directed research skills during their academic studies. This background could foster greater familiarity with emerging digital tools, potentially enhancing their confidence and autonomy to experiment with AI applications in language instruction. This interpretation is consistent with the view that teachers' technology use depends on the interplay of knowledge, confidence, beliefs, and school culture [42], although none of these factors was measured here.

The findings of this study contrast with those of Hamadneh and Mureed, who reported no significant differences by educational qualifications, whereas the present study observed differences favoring postgraduate degree holders. The results also contrast with on gender, which reported differences favoring male teachers. However, the current findings align with in reporting no statistically significant differences by years of experience. These divergences are discussed further in the next section [27. 28].

4 General Discussion and Critical Comparison with Prior Literature

Taken together, the findings suggest that Arabic language teachers in this directorate report a moderate degree of AI integration that is concentrated in text-based practices (reading and writing) and weaker in oral skills. This pattern is consistent with the broader picture in the literature, in which AI tools are most readily adopted where they reduce teachers' workload in text production and correction, while audio-based applications face infrastructural and technical constraints [33, 1]. It also fits the concerns raised about the accuracy of Arabic natural language and speech technologies [1].

Compared with the closest Arabic-language studies, the present results converge on the overall level but diverge on the demographic correlates. Hamadneh [27].and Mureed [28] also reported moderate overall levels, yet Hamadneh found a gender difference favoring males and Mureed found an age effect favoring younger teachers; neither was replicated here. Three differences may account for this: (a) the constructs measured differ (planning, execution, and evaluation in Hamadneh,; the reality of use and its challenges in Mureed,; four language skills in the present study); (b) the samples differ in size, region, and school stage 98 teachers in Al-Mafraq; 130 primary school teachers in Mureed and (c) age and years of experience are related but not interchangeable variables. The convergence on a "moderate" level should therefore not be overstated, because the instruments and cut-off points differ across studies.

The school-type and qualification effects are the study's only significant inferential findings, and both are modest. Their direction agrees with the view that teachers' technology use depends on access to resources as well as on knowledge, confidence, beliefs, and school culture [42], but the present data cannot separate these mechanisms because infrastructure, training, and beliefs were not measured. The school-type contrast also rests on only 22 private school teachers and is best treated as hypothesis-generating.

Finally, experimental studies [30, 29] report that AI-supported instruction improved students' writing and language skills, whereas the present study documents only how much teachers report using AI. Whether the moderate level of use found here is sufficient to produce such gains cannot be inferred from these data. Reviews of AI research in education have also criticized the limited involvement of educators [43, 44]; by surveying teachers directly, the present study partly addresses this gap, while remaining limited to self-reports.

5 Limitations of the Study

Although this study provides valuable baseline empirical evidence regarding Arabic language teachers' integration of artificial intelligence tools, several critical methodological and psychometric limitations must be considered when interpreting and generalizing the findings. First, the data collection relied exclusively on a self-administered questionnaire, measuring teachers' *self-reported* perceptions rather than their *actually observed* classroom practices. This self-report design inherently exposes the results to social desirability bias and potential over-reporting of technological integration; common-method variance may also inflate associations among self-reported items [45] . Second, the spatial scope was restricted to a single educational district—the Third Irbid Directorate of Education in

© 2026 NSP
Natural Sciences Publishing Cor.

Jordan—which, together with the convenience sampling procedure (N = 135; 71.8% of the population) and the possible self-selection of teachers who are more interested in technology, limits the generalizability of the conclusions to broader national or international educational contexts with differing technological infrastructures. Third, the study employed a cross-sectional survey design, capturing teacher engagement at a single point in time; consequently, it cannot measure dynamic longitudinal adaptation trajectories or establish causal relationships over time. Fourth, statistical comparisons across demographic categories involved pronounced imbalances in subgroup sample sizes, particularly between public (n=113) and private (n=22) school educators, requiring caution when attributing variance strictly to institutional governance; moreover, the observed effects were small to medium, and the study had adequate power only for medium-sized effects (Table 14). Fifth, because school identifiers were not collected, the nested structure of the data (teachers within schools) could not be modeled with multilevel analysis, so the independence assumption may be violated and the p-values for school-level predictors may be optimistic. Sixth, evidence for the instrument's reliability and validity is preliminary: it rests on a pilot of 20 teachers, no factor analysis was conducted, and the Listening and Speaking domains were very highly correlated in the pilot (r = .895), which suggests overlapping constructs. Finally, the analytical framework evaluated teacher-side integration habits in isolation, without measuring actual classroom instructional efficacy or direct student learning outcomes. Therefore, a clear distinction must be maintained between the reported level of AI usage documented here and actual observed classroom practices or measurable student achievement in Arabic language skills.

6 Conclusion

This study provides baseline evidence, from a convenience sample of 135 teachers' self-reports in one Jordanian directorate, that AI integration in teaching Arabic language skills is moderate overall, highest for reading and lowest for listening. Teachers with postgraduate qualifications and teachers in private schools reported higher integration, but these differences were small to medium, rest on unbalanced groups, and cannot be interpreted causally or generalized beyond similar contexts. The results are best treated as a needs-assessment baseline that justifies targeted training and infrastructure pilots, and further observational, multilevel, and longitudinal research, rather than as definitive evidence of what works in classrooms.

7 Recommendations

Based on the findings of this study, and bearing its limitations in mind, the researcher suggests the following recommendations:

1. **Specialized Training:** Organize hands-on professional development programs on specific AI applications for each language skill, prioritizing public school teachers, Bachelor's degree holders, and the least-used practices (e.g., adjusting the speed and tone of listening texts; Item 3, $M = 2.68$).
2. **Curriculum Integration:** Embed AI-driven activities and applications directly into Arabic language textbooks, paired with comprehensive teacher guidebooks that outline correct classroom implementation models.
3. **Infrastructure and Tools:** Provide audio equipment and reliable internet in public schools, and pilot-test Arabic speech-recognition and text-to-speech tools before wider adoption, since oral skills showed the lowest AI integration.
4. **Future Research:** Conduct future studies using probability-based samples, school identifiers for multilevel modeling [41], classroom observation alongside self-reports, and experimental designs to test the impact of AI integration on students' language skills [47].

Detailed Ethical Declaration

This study was conducted in strict accordance with the ethical principles for research involving human subjects outlined in the Declaration of Helsinki. Formal ethical approval for the study protocol and data collection instruments was granted by the Institutional Ethics Committee, along with official field permissions from the Directorate of Education. Prior to accessing the digital questionnaire, participants were presented with an electronic informed consent page detailing the study's objectives, potential risks, and data handling protocols. Participants were required to check a mandatory consent box ("I agree to participate") before proceeding to the survey items. Voluntary participation was emphasized, with teachers explicitly informed of their right to decline to participate or withdraw at any point before submission, without adverse consequences or penalties. To ensure anonymity and confidentiality, no personally identifiable information (such as names,

© 2026 NSP
Natural Sciences Publishing Cor.

phone numbers, or specific school titles) was collected. All survey responses were automatically coded, encrypted, and stored in secure, password-protected cloud storage accessible solely to the primary researcher, ensuring that data are reported in aggregate strictly for academic research purposes.

Conflict of Interest

The author declares that there are no conflicts of interest regarding the publication of this manuscript. There are no financial, personal, or institutional relationships that could be perceived as inappropriately influencing or biasing the design, execution, analysis, or presentation of this research.

Declaration on the Use of Artificial Intelligence (AI)

The author declares that artificial intelligence (AI) tools were used only within the limits permitted by the journal's policies, solely for language editing and improving the clarity and readability of the manuscript. The author takes full responsibility for the scientific content, including the study design, data analysis, interpretation of the findings, and conclusions.

Funding

The author is grateful to the Deanship of Research at Jadara University for providing financial support for this publication (Grant No. Jadara-SR-full 2023).

Acknowledgments: "The authors are grateful to the Deanship of Scientific Research at Jadara University for providing financial support for this publication".

References

[1] Qandil , N. T. (2025). The effectiveness of artificial intelligence techniques in teaching Arabic to non-native speakers amid educational and technological challenges. In Proceedings of the International Scientific-Practical Conference on Global Linguistics: New Approaches and Research (pp. 945–949). Uzbekistan State World Languages University.

[2] Al-Khouli, M. (2000). Methods of teaching the Arabic language [in Arabic]. Dar Al-Falah for Publishing and Distribution.

[3] Tu'aymah, R. (2004). Language skills: Levels, teaching, and difficulties [in Arabic]. Dar Al-Fikr Al-Arabi.

[4] Eyadat, S, Darawsheh , N, Alrashdan, H, ALRABA'I, E, Alshamali1 A, Nazzal, K, Maabreh, SHand Ali Marashdeh, R .(2026). Assessing the Readiness for Blended EducationalSupervision in Jordan: Educational Supervisors' Perspectives from Irbid Governorate, *Journal of Statistics Applications & Probability*,15(3),859-873

[5] Koka, N. A., & Khan, M. R. (2024). Exploring insights from undergraduates on the use of artificial intelligence (AI) tools to enhance proficiency in tense and aspect in foreign language instruction. Pakistan Journal of Life and Social Sciences, 22(2), 19633–19656. https://doi.org/10.57239/PJLSS-2024-22.2.001440

[6] ALAwAmRAh, A , Ajmi, H, Darawsheh, N , Darawsheh, O, Urabi, O.(2024). Analyzing Student's Perceptions about the Values of Digital Citizenship, *Journal of Statistics Applications and Probability*, 2024, 13(4), pp. 1279–1288.

[7] Lamouchi, A., Sadek Elsayed, M., Abdellatif, M., & Aly, S. (2025). Leveraging artificial intelligence applications to develop English language skills and foster sustainable development awareness among Saudi university students: A step towards future economies, Research Journal in Advanced Humanities, 6(3). https://doi.org/10.58256/tf8dyv05

[8] ALRABA'I, E, Fraihat, N , Alshawabkeh, G , Al-Momani, A , Obeidat, N , Alshamali, A , Darawsheh, N , Alrashdan,H and Kamayseh, A.(2026). The Impact of University Administration on Reducing Cyberbullying, Journal of Statistics Applications & Probability, 15 (3), 875-889

[9] Alrashdan, H, Darawsheh, N, Al Ajmi, H, Al-Khataybeh, M, Almheiri, A, Salih, A and Kamayseh, A.(2026). Statistical Analysis of the Impact of Organizational Pride on Morale: A Case Study of Universities, *Journal of Statistics Applications & Probability*, 15, (5), 1189-1198

[10] Darawsheh, N; Alkailanee, K; Alquran, N; AL-Dawuod, T; Nawafleh, M.; and Jaradat, H.. Almomany. H. Alrashdan, EAlmomany (2023). Leading with Love and Its Relationship to Faculty Members' Job Loyalty in Jordanian Universities, *Information Sciences Letters*, 12(8), pp. 2511–2520.

[11] Almheiri, A, Albastaki, H and Alrashdan, H.(2025). Chapter 7 AI-Based Tutoring Systems in Education: A Systematic Literature Review on Personalized Learning, Intelligent Agents, and Learning Analytics, *IGI. Global* ,185- 210

© 2026 NSP
Natural Sciences Publishing Cor.

[12] Alkharman , Jamal , Drawsheh , Sonia, Al-Khataybeh , Majid, BaniYounes , Zein, Darawsheh , N, Alrashdan, H .(2024). Cyber Attacks and its Implication to National Security: The Need for International Law Enforcement, *Pakistan Journal of Criminology*, 16(3), pp. 851–864

[13] Al-Abadi, A. (2024). Effectiveness of using generative artificial intelligence applications in developing written expression and self-regulated learning among middle school students [in Arabic]. Journal of Umm Al-Qura University for Educational and Psychological Sciences, 17(4), 105–119. https://doi.org/10.54940/ep16307404

[14] Hatamleh, H, AL-Zoubi, Z and Darawsheh, N. (2017). Evaluation Degree Of Mangement Practice in Jordanian Universities and its Relationship to Occupational Empowerment, as Viewed by the Faculty Members, *Journal of Institutional Research South East Asia,(JIRSEA*, 15(2):70-94

[15] Al-Maqbali, L. R. (2025). Artificial intelligence and its role in assessing Arabic language skills for non-native speakers (Proposed applications). International Journal of Humanities and Educational Research, 7(3), 94–104. https://doi.org/10.47832/2757-5403.32.5

[16] Darawsheh, N.(2021).The Role of University Administration in Enhancing Intellectual Security Among Yarmouk University Students**,** *Journal of Education and Training Studies*, 9(5)**,** 35-51

[17] Al-Haq, M. A. (2025). The use of artificial intelligence in Arabic language education: Opportunities and challenges [in Arabic]. Arab Journal of Scientific Publishing, 8(79), 173–188. https://www.ajsp.net

[18] Alrashdan, H, Darawsheh, N, Almheiri, A, Aljanabi, Y, Al-Khataybeh, M, & Al-Quran, N. (2025). The Effect of Administrative Bullying on the Quality of Education. *Educational Process: International Journal*, 18, e2025515. https://doi.org/10.22521/edupij.2025.18.515

[19] Alqaed, M. A. (2024). AI in English language learning: Saudi learners' perspectives and usage. Advanced Education, 25, 125–142. https://doi.org/10.20535/2410-8286.318972

[20] Crompton, H, Edmett, A, Ichaporia, N & Burke, D. (2024). AI and English language teaching: Affordances and challenges. British Journal of Educational Technology, 55(6), 2503–2529. https://doi.org/10.1111/bjet.13460

[21] Darawsheh, N., Alrashdan, H., Al-Khataybeh, M., Al Ajmi, H., Almheiri, A., Shaglof, A., Aldarmaki, I., BaniYounes, Z., &Aidi, R. (2025). The degree of impact of organizational intelligence obstacles in the digital age. Journal of Statistics Applications & Probability: An International Journal, 14(6), 985–995.

[22] ALAwAmRAh, A , Darawsheh, N, Alrashdan, H , Ghalia, N , Mustafa, M , Darawsheh, A , Abu Reya, H and Tahani. Alali, T.(2023). Transformational Practices of Academic Leaders in Universities: Suggestions for Improvement**,** *Information Sciences Letters***,** 12, (10), 2767-2774

[23] Al-Azzam, H. A. M., & Al-Azzam, J. K. J. (2026). Teachers' attitudes towards the use of artificial intelligence applications in Arabic language education and their relationship to professional pressures. International Journal of Drug Delivery Technology, 16(57s), 427–440. https://doi.org/10.25258/ijddt.16.57s.54

[24] Darawsheh, N., Al-Khataybeh, M., Alrashdan, H., Drawsheh, S., Abu Grarh, A., & Abu Dogush, M. (2024). Coexistence skills among mothers of children with multiple disabilities in the northern region of Jordan (Irbid). Journal of Statistics Applications & Probability, 13(2), 665–680.

[25] Fayyoumi, K. A. R., Yassin, B., &Al-Fayyoumi, S. K. (2025). The extent to which language teachers use linguistic artificial intelligence applications: An analysis based on selected variables (Version 1). Research Square. https://doi.org/10.21203/rs.3.rs-7311266/v1

[26] Al-Manajreh, A. A., Al-Busaidiyah, F. Y., & Abu Shandi, Y. A. (2025). Challenges of employing artificial intelligence (AI) technologies in teaching the Arabic language in public and private schools in the Sultanate of Oman from teachers' perspectives [in Arabic]. Journal of Educational and Psychological Research, University of Baghdad, 22(86), 217–251. https://doi.org/10.52839/0111-000-086-007

[27] Hamadneh, A. D. (2026). The degree to which Arabic language teachers in Al-Mafraq Kasbah possess skills for employing artificial intelligence applications in education [in Arabic]. Irbid Journal for Research and Human Studies, 28(1), 396–419.

[28] Mureed, L. (2025). The level of employing artificial intelligence in teaching the Arabic language at the primary school level [in Arabic]. Journal of Humanities and Natural Sciences, 6(12), 342–369.

[29] Al-Azzawi, B. I. S. (2026). The impact of employing artificial intelligence tools on developing language skills among students of the College of Fine Arts [in Arabic]. Journal of the College of Education, 62(2), 541–560. https://doi.org/10.31185/eduj.Vol62.Iss2.4830

[30] Al-Abadi, M. A. (2024). Effectiveness of using generative artificial intelligence applications in developing written expression and self-regulated learning among middle school students [in Arabic]. Journal of Umm Al-Qura University for Educational and Psychological Sciences, 17(4), 105–119. https://doi.org/10.54940/ep16307404

[31] Yatri, D., Anugerahwati, M., & Setyowati, L. (2023). Artificial intelligence (AI) in language learning (English and Arabic classes): Students' and teachers' experiences and perceptions. Transformational Language, Literature, and Technology Overview in Learning (TRANSTOOL), 3(1), 1–12. https://ojs.transpublika.com/index.php/TRANSTOOL

[32] Al-Mukhallafi, T. R. (2020). Using artificial intelligence for developing English language teaching/learning: An analytical study from university students' perspective. International Journal of English Linguistics, 10(6), 40–53. https://doi.org/10.5539/ijel.v10n6p40

[33] Al-Tamimi, D. H. F. (2025). AI Integration in EFL classrooms: A study on lecturers' adaptation and student outcomes. Arab World English Journal (AWEJ), Special Issue on CALL, 11, 364–379. https://doi.org/10.24093/awej/call11.22

© 2026 NSP
Natural Sciences Publishing Cor.

[34] Darawsheh, N. (2018).The Reality of The Challenges Faced by Graduate Students in the Faculties of Educational Sciences in Jordanian Universities**,** Journal of Institutional Research South East Asia *JIRSEA***,** 16 (2), 123-140

[35] Mutawe, S., Nada, A., & Al-Assali, A. (2026). The degree of the use of artificial intelligence applications in teaching skills among lower basic stage teachers in Nablus Governorate [in Arabic]. International Journal of Humanities and Social Sciences, (72), 107–125. https://doi.org/10.33193/IJoHSS.72.2026.949

[36] Sulaiman, N. M. A. (2025). Artificial intelligence and teaching Arabic grammar in basic education: Major opportunities and challenges [Conference paper, in Arabic]. Annual Conference of the Center for Educational Studies and Research: Scientific Research as a Pillar for Formulating Educational Policies Toward a Sustainable Teaching and Learning System, Faculty of Education, Lebanese University, Beirut, Lebanon.

[37] Badarneh, H, Najwa Darawsha,N, Ahmed AL-Shoha, A and Abdel Aal, M.(2020). The Role of Faculty Members in Developing the Patriotism amongst Their Students at Jadara University and Jerash Al Ahlia University from the Students' Point of View, *Universal Journal of Educational Research*, 8(10): 4478-4489, 2020

[38] Al-Zoubi H., *, Maberah, S, Darawsha N, Bany Issa, H.(2021). High School Principals' Situational Leadership and Its Relationship with Teachers' Achievement Motivation, European Journal of Contemporary Education, 10(4): 1027-1041

[39] Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

[40] Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

[41] Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge. https://doi.org/10.4324/9781315650982

[42] Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education, 42*(3), 255–284. https://doi.org/10.1080/15391523.2010.10782551

[43] Alrashdan,H , Al Ajmi, H; Alnasraween, M and Karra, S. (2022). Efficiency Determinants of Educational Wastage Programs, Journal of Education and eLearning Research, 9(2): 79-86.

[44] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

[45] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

[46] Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician, 55*(1), 19–24. https://doi.org/10.1198/000313001300339897

[47] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

© 2026 NSP
Natural Sciences Publishing Cor.