

VALIDITY AND RELIABILITY OF A GENERATIVE ARTIFICIAL INTELLIGENCE (GEN AI) INSTRUMENT FOR ARABIC LANGUAGE STUDENTS AT PUBLIC UNIVERSITIES IN THE KLANG VALLEY: A PILOT STUDY

Mohammad Najib Jaffariⁱ, Nurkhamimi Zainuddinⁱⁱ, Muhammad Hakim Kamalⁱⁱⁱ & Mohammad Imran Ahmad^{iv}

- ⁱ (*Corresponding author*). Associate Professor, Faculty of Major Language Studies, Universiti Sains Islam Malaysia, Negeri Sembilan, Malaysia. najib@usim.edu.my
- ⁱⁱ Associate Professor, Faculty of Major Language Studies, Universiti Sains Islam Malaysia, Negeri Sembilan, Malaysia. khamimi@usim.edu.my
- ⁱⁱⁱ Lecturer, Kolej Permata Insan, Universiti Sains Islam Malaysia, Negeri Sembilan, Malaysia. hakimkamal@usim.edu.my
- ^{iv} Senior Lecturer, Faculty of Islamic Civilization Studies, Universiti Islam Selangor, Selangor, Malaysia. imranahmad@uis.edu.my

Article Progress

Received: 15. April 2026

Revised: 17 May 2026

Accepted: 30 May 2026

Abstract	<i>This study examines the validity and reliability of a research instrument designed to investigate the use of Generative Artificial Intelligence (GenAI) among Arabic language students at public universities in the Klang Valley, Malaysia. This study aligns with the best practices in quantitative research, where instruments must undergo validation and verification processes prior to actual data collection. The content validity was established through evaluation by three field experts from various public universities from Universiti Teknologi MARA Shah Alam, Universiti Pendidikan Sultan Idris, and National University of Malaysia (UKM) using the Kappa coefficient analysis, yielding an overall Kappa value of 0.72, indicating a good level of agreement. Subsequently, a pilot study involving 36 respondents was conducted to assess the reliability of the instrument using Cronbach's Alpha. All constructs, namely Lacks, Necessities, and Wants, demonstrated Cronbach's Alpha values of 0.871, 0.954, and 0.957, respectively, confirming strong internal consistency. These findings confirm that the instrument is valid, reliable, and suitable for use in the main study.</i> Keywords: Arabic, Language, Learning, Generative Artificial Intelligence, Students.
-----------------	--

INTRODUCTION

The rapid advancement of Artificial Intelligence (AI), particularly Generative Artificial Intelligence (GenAI), has brought significant transformations across various sectors, including language education. Within the landscape of higher education, the utilisation of AI is becoming increasingly pervasive and is expected to become an integral component of the teaching and learning process (Ahmad & Kassim, 2025).

In the context of Arabic language acquisition, the integration of Generative Artificial Intelligence (GenAI) has opened new opportunities for autonomous learning, instant feedback, and advanced writing support. However, the Arabic language presents unique linguistic challenges, including its rich morphology, complex syntactic structures, and

dialectal diversity, which complicate the development of accurate and effective AI applications compared to many other languages (Alayba, 2025; Moustafa et al., 2026)

In Malaysia, public universities particularly those located in the Klang Valley, such as Universiti Kebangsaan Malaysia (UKM), Universiti Putra Malaysia (UPM) and Universiti Teknologi MARA (UiTM) offer Arabic language programmes to students from diverse educational backgrounds. Despite the rapid growth of AI technology, there remains a notable dearth of empirical research examining how Arabic language learners utilise, perceive, and adopt GenAI technologies in their learning journeys (Yatri et al., 2023).

Consequently, this study is conducted to undergo a pilot study before analysing the needs of Arabic language learners regarding GenAI usage, focusing on the dimensions of lacks, necessities, and wants within the framework of language learning.

PROBLEM STATEMENT

While the emergence of GenAI has shifted the paradigm of language education, its application in Arabic language learning remains in its nascent stages and is not yet profoundly understood. Although AI holds the potential to enhance pedagogical effectiveness, several critical issues warrant attention.

First, there is a lack of comprehensive empirical evidence regarding the readiness and acceptance of Generative Artificial Intelligence (GenAI) among Arabic language learners in Malaysian public universities. Existing studies have primarily focused on English language education or broader higher education contexts, leaving a significant research gap concerning Arabic language learning within the Malaysian context (Lee et al., 2025; Moustafa et al., 2026).

Second, the structural complexity, rich morphology, and dialectal diversity of the Arabic language pose substantial challenges for developing accurate and effective AI applications, particularly in natural language processing and language generation (Mohd et al., 2020; Moustafa et al., 2026). This complexity directly impacts the efficacy of technology integration in mastering the language.

Third, despite the development of Arabic-centric AI models such as Jais, their effectiveness and user acceptance within the Malaysian higher education framework remain ambiguous (Al-Onaizan, 2024).

RESEARCH OBJECTIVES

This study seeks to evaluate the validity and reliability of a research instrument used to study the use of Generative Artificial Intelligence (GenAI) by Arabic language students in public universities in Klang Valley, Malaysia. In particular, the study aims to: (1) validate the content of the instrument based on expert review using the Kappa coefficient analysis, (2) provide internal consistency reliability of the instrument through a pilot study with 36 respondents by Cronbach's Alpha, and (3) determine the suitability of the instrument to be administered in the main study based on the results of the instrument content and internal consistency reliability tests.

LITERATURE REVIEW

The use of GenAI in higher education shows a shift toward its role as a research assistant, aiding students in information retrieval, academic writing, and data analysis, which impacts their learning and research processes (Okulicz-Kozaryn, 2026). Furthermore, it has been observed that GenAI can enhance students' productivity and creativity; but it also poses challenges of authenticity of text, dependency, and academic honesty (Yagmurlugil & Karaca; Everson, 2025; Nagpal, 2024).

Research also shows that there needs to be guidance to ensure that students use AI in an ethical way. In some specialties like gifted education and language learning, the assessment of the competency of the learners requires rigour and context sensitivity (Muhammad Hakim, 2026; Kamal 2026a).

Pedagogically, GenAI provides adaptive and personalised learning support tailored to learners' needs, thereby fostering learner autonomy and improving language learning outcomes (Lee et al., 2025; Saarela et al., 2026). Building on this, another systematic review of 49 empirical studies shows that the use of GenAI in language classrooms is mainly located in higher education, reinforcing its increasing salience in tertiary language teaching (Lee et al., 2025).

This technology can also enhance creativity and learning experiences when integrated with immersive technologies such as virtual reality and game-based learning, which have been shown to increase students' engagement, motivation, and learning outcomes (Zawacki-Richter et al., 2019; Lee et al., 2025). However, this research study also highlights the pros and cons of GenAI in the field of education, such as the potential for misinformation, algorithmic bias, and the difficulty in ensuring a deep conceptual understanding for students (Zeynep & Yesilyurt, 2025).

In addition, several factors influence students' acceptance of GenAI, including perceived usefulness, perceived ease of use, trust in technology, as well as ethical and privacy concerns (Kasneci et al., 2023). The study also reveals that personality traits are significant factors influencing the perception and use of AI tools among learners, with some personality traits being positively correlated with being receptive to learning environments facilitated by AI (Linur et al., 2025).

The above indicate that the adoption of GenAI in education is not solely reliant on the technical aspects but also on the attitudes of users, their psychological makeups, and preparedness. Accordingly, a modern Arabic language learning environment is being used in this study, where the elements of GenAI are added to the Hutchinson and Waters model to gain an understanding of students' needs.

This adaptation is also guided by recent studies on creative teaching strategies in the field of Arabic language teaching such as flipping classrooms, which emphasizes the use of technology and student-centered teaching strategies (Kamal, 2026b). The necessities component includes the students' need to master Arabic language skills as well as the competency to use GenAI as a learning tool.

Lacks refer to gaps in language mastery and technological skills, including a lack of AI literacy and the ability to critically evaluate information. Meanwhile, wants involve the perceptions, interests, and levels of student acceptance regarding the use of GenAI in their learning. This approach allows the study to provide a more holistic overview of student needs, considering the interaction between linguistic, technological, and pedagogical factors.

Thus, this theoretical framework not only supports traditional learner needs analysis but also extends it to the evolving digital education context, thereby contributing to the development of more effective teaching strategies and technological integration in Arabic language learning at higher education institutions.

Generative Artificial Intelligence (GenAI) is increasingly being adopted in higher education to support teaching and learning activities. It assists both students and lecturers in generating learning materials, enhancing academic writing, and improving overall productivity (Yilmaz & Yilmaz, 2023). Furthermore, recent studies have shown that GenAI enhances student engagement, promotes personalised learning, and improves learning outcomes in language education, including Arabic language learning (Lee et al., 2025; Moustafa et al., 2026).

This development also paves the way for the implementation of adaptive and personalised learning, which is more responsive to individual student needs (Zawacki-Richter et al., 2024). Taken together, these findings highlight the transformative potential of GenAI in addressing the problems of traditional pedagogy, and indicate that further research and theory development are needed to evaluate its role in Arabic language instruction.

Nevertheless, Arabic language learning in public universities continues to face various challenges, including limited opportunities for authentic language use, difficulties

in mastering complex grammatical structures, and low learner motivation (Jaffar & Nazrol, 2025; Moustafa et al., 2026). In this regard, the implementation of Generative Artificial Intelligence (GenAI) has emerged as a promising innovation to address these challenges by enriching vocabulary acquisition, enhancing communicative competence, and supporting personalised and autonomous learning (Lee et al., 2025; Moustafa et al., 2026).

Systematic reviews have highlighted that technology-enhanced and learner-centred pedagogical approaches are essential for improving language learning effectiveness, particularly by promoting learner engagement, autonomy, and personalised learning experiences (Lee et al., 2025). Studies in Arabic language education likewise demonstrate that student-centred instructional approaches contribute significantly to improving learners' communicative competence and speaking proficiency (Yaacob & Abu Bakar, 2018).

Thus, the integration of AI into Arabic language learning not only improves learning outcomes but also strengthens overall pedagogical effectiveness. While the promises of using GenAI for higher education are numerous, there are key challenges to their implementation. They encompass unclear guidelines in the institutions, ethical issues, and concerns of students using technology the wrong way (Kasneci et al., 2023; Khalil & Er, 2023).

However, in the particular case of Arabic language education, the linguistic complexity of the language, the lack of datasets of high quality in Arabic, and the dilemma of the ethical issues of AI-generated content exacerbate these challenges (Moustafa et al., 2026). The challenges highlight the need for appropriate policy frameworks, awareness and pedagogical approaches to ensure that the benefits of GenAI for language learning are not compromised in terms of educational efficacy and integrity.

METHODOLOGY

This study adopts a quantitative approach with a descriptive-analytical design, aimed at systematically describing the phenomenon of GenAI usage among Arabic language students through a pilot study. This method is appropriate as it enables the researchers to identify patterns, trends, and levels of technology usage within the context of language learning (Creswell & Creswell, 2017).

Validity and Reliability of Instrument

Evaluation by three experts in the fields of Arabic Language, Educational Technology (Edu Tech) and Integration of Information Technology in Teaching and Learning is used to establish the content validity of the instrument. The research instrument is a structured questionnaire, which included five sections: Demographic (section A), Lacks (section B), Necessities (section C), Wants (section D) and Open-ended questions (section E).

There are a total of 36 items for each instrument, which has three main constructs (Lacks with 12 items, Necessities with 12 items, Wants with 12 items) and demographic items and open-ended questions. The framework is based on the needs analysis model (Hutchinson and Waters, 1987), with dimensions of GenAI added in. This assessment guarantees that all the questions in the questionnaire are relevant and that each question is measuring the constructs as intended.

Uses the expertise to evaluate elements of the language used, its meaning, its appropriateness for the context, and its correspondence to the purpose of the study. An evaluation form for each expert was prepared and asked each one to evaluate each item as "Suitable" or "Needs Modification" and give written comments if it was thought that something needed to be modified. This is important to prevent ambiguity/misinterpretation amongst respondents.

The level of agreement among experts was analysed using the Kappa coefficient (Cohen, 1960). Landis and Koch (1977) reported that high Kappa values would indicate high levels of agreement and thus instrument validity. Data were analysed with IBM SPSS

Statistics Version 26.0, with descriptive statistics such as frequencies, percentages and mean scores used for the demographic profiling.

Next, an internal consistency Cronbach's Alpha was calculated for each construct to determine the internal consistency of each within the study. The suggestions for improvement by the experts are taken and further refined the instrument before entering next step.

PILOT STUDY

Following expert validation, a pilot study involving 36 Arabic language students at three public universities was conducted. These universities consisted of Universiti Kebangsaan Malaysia (UKM), Universiti Putra Malaysia (UPM) and Universiti Teknologi MARA (UiTM), Shah Alam. This sample size aligns with Browne (1995), who suggests that it is adequate for assessing the initial stability of an instrument.

The primary purpose of the pilot study is to evaluate the reliability of the instrument, particularly in terms of internal consistency among the items. Cronbach's Alpha is used as the main indicator to determine the extent to which items within each construct (lacks, necessities, wants) consistently measure the same concept. According to Hair et al. (2019), a high Cronbach's Alpha value indicates strong internal consistency, thereby enhancing the reliability of the data. In this study, the obtained alpha values exceed the recommended minimum threshold, indicating a high level of reliability.

This pilot study also serves as a preliminary mechanism to identify potential weaknesses in the instrument, such as confusing or difficult items, inappropriate terminology and sentence structures that may lead to misinterpretation. Feedback from respondents is valuable as it reflects how students interpret and respond to the items. Items with low correlation or those that negatively affect the overall alpha value are revised or removed before the actual study.

RESULTS

Content Validity's Result

Section A

The content validity of the instrument was assessed by three (3) experts with expertise in the fields of Arabic Language, Educational Technology, and the Integration of Information Technology in Teaching and Learning. The purpose of the expert evaluation is to ensure that each item in the instrument measures the intended construct, has clarity of language, and is appropriate for the research objectives and the target respondent level, namely Arabic students at the three public universities.

Each expert was asked to evaluate the items of the instrument based on two categories: "Suitable" and "Needs Modification." To analyse the level of agreement among the experts, this study used Cohen's Kappa coefficient, as this method is suitable for assessing inter-rater reliability and accounts for the possibility of agreement occurring by chance.

The Kappa coefficient was used to determine the level of consensus among the three experts in evaluating the suitability of the instrument items. The interpretation of the Kappa value is based on the recommendations of Landis and Koch (1977) as shown in Table 1.

Table 1: Interpretation of Kappa Coefficient Values

Kappa Value	Level of Agreement
< 0.20	Very Weak
0.21 – 0.40	Weak
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.81 – 1.00	Very Good

The analysis results showed that the overall Kappa coefficient value of the instrument was 0.72, indicating a good level of agreement among the three experts. This suggests that the items in the instrument generally have satisfactory content validity and were consistently rated by the experts.

Section B

Kappa Analysis by Instrument Section

Kappa coefficient analysis was also conducted by section of the instrument as follows:

Table 2: Kappa Coefficient Analysis by Instrument Section

Instrument Section	Kappa Value	Interpretation
Section A: Demographics	0.80	Good
Section B: Lacks	0.68	Good
Section C: Necessities	0.70	Good
Section D: Wants	0.75	Good
Section E: Open Questions	0.82	Very Good

These findings indicate that all sections of the instrument recorded Kappa values exceeding 0.60, thereby proving that there is a good to very good level of agreement among experts in assessing the suitability of the research instrument items.

Although the Kappa coefficient value indicates a good level of agreement, the experts also suggested several improvements to the structure and phrasing of the items, particularly regarding the alignment of items with the research objectives, reclassification of the constructs of Lacks, Necessities, and Wants and improvements in the clarity of language and context of the items.

In this regard, the authors updated and improved the items of the instrument based on the comments and suggestions from the experts before the instrument was used in the pilot study. These improvements aim to further enhance the content validity and ensure that the instrument is clearer and easier to understand for the target respondents.

Overall, the Kappa coefficient analysis indicates that the instrument has a good and acceptable level of content validity, thus making it suitable for the next phase of the study, which is the pilot study to test the reliability of the instrument using Cronbach's Alpha analysis.

PILOT RELIABILITY'S RESULT

This section presents the findings of the pilot study conducted to evaluate the reliability of the research instrument prior to the actual data collection. The pilot study aimed to assess the internal consistency of the questionnaire items and to ensure that the instrument was suitable for use in the main study.

Total of 36 respondents, consisting of Arabic language students from public universities in Klang Valley, participated in the pilot study. The sample size was considered adequate for a pilot reliability test. The questionnaire items were measured using a five-point Likert scale, ranging from Strongly Disagree (1) to Strongly Agree (5).

Analysis of Demographic

This section details the background of the 36 respondents involved in the pilot phase, representing Arabic language students from three public universities in the Klang Valley.

Table 3: Demographic Profile of Pilot Study Respondents

Profile Category	Sub-Category	Frequency (f)	Percentage (%)
Gender	Female	29	80.6%
	Male	7	19.4%
University	UiTM / UPM / UKM	36	100%

Year of Study	Year 1 / Year 4	36	100%
Proficiency	Basic / Intermediate	36	100%
AI Experience	Yes (Experienced)	27	75.0%

Reliability Analysis Using Cronbach's Alpha

The reliability of the instrument was examined using Cronbach's Alpha coefficient, which is commonly used to measure internal consistency reliability in quantitative research. Cronbach's Alpha values range from 0 to 1, with higher values indicating greater reliability.

In this study, a Cronbach's Alpha value of 0.80 or above was considered indicative of good to excellent reliability, consistent with recommendations in educational and social science research.

Cronbach's Alpha Results by Construct

Cronbach's Alpha analysis was conducted for each construct in the instrument, namely Lacks, Necessities, and Wants. The results are presented in Table 4.

Table 4: Cronbach's Alpha Values for Each Construct (Pilot Study)

Construct	Cronbach's Alpha	Interpretation
Lacks	≥ 0.80 (0.871)	Good
Necessities	≥ 0.80 (0.954)	Very Good
Wants	≥ 0.80 (0.957)	Very Good
Overall Instrument	≥ 0.80 (0.927, Mean)	Excellent

Item Analysis and Refinement

In addition to overall reliability, item-level analysis was conducted to identify items that might reduce the internal consistency of the instrument. Two key indicators were examined; Corrected Item–Total Correlation, and Cronbach's Alpha if Item Deleted.

Items with low item–total correlation values were reviewed, and potential revisions were considered. However, the analysis indicated that the items demonstrated acceptable correlations and contributed positively to the reliability of their respective constructs. Therefore, all items were retained for the main study.

To sum up, the findings of the pilot study indicate that the research instrument exhibits good internal consistency reliability, with all constructs achieving Cronbach's Alpha values at or above the recommended threshold of 0.80. This suggests that the items within each construct consistently measure the same underlying concept. The results further confirm that the instrument was stable, reliable, and appropriate for investigating the necessities, lacks, and wants related to the use of Generative Artificial Intelligence (GenAI) in Arabic language learning.

DISCUSSION

This study provides a comprehensive discussion of the research findings by integrating the results of expert content validity evaluation, pilot reliability testing conducted on the developed instrument. The purpose of this discussion is not only to summarise the findings, but also to critically evaluate the extent to which the research objectives have been achieved, particularly in addressing the needs for the use of Generative Artificial Intelligence (GenAI) among Arabic language students in public universities.

Furthermore, this study situates the findings within the broader context of current issues in digital education, especially in relation to the integration of emerging technologies into language learning environments. Such an approach ensures that the findings are not merely descriptive, but also interpretative and meaningful in informing current educational practices.

The robustness of this study is supported by a systematic and multi-layered methodological approach, consisting of two main stages:

1) Content Validity

The research instrument underwent evaluation by three field experts to ensure that each item accurately represented the constructs being measured. The use of Kappa analysis enabled the level of agreement among experts to be quantified, thereby enhancing transparency and methodological rigour in assessing the instrument's validity. This step ensured that the instrument aligns closely with the study objectives and theoretical framework.

2) Pilot Study

A pilot study involving 36 students was conducted to assess the reliability of the instrument using Cronbach's Alpha coefficient. This phase was crucial in determining the internal consistency of the items and identifying any weaknesses before the instrument was administered in the main study. It also helped refine the clarity and structure of the questionnaire.

Overall, this two-stage approach demonstrates strong methodological rigour, as the instrument underwent both validation and reliability verification processes prior to its use in future empirical investigations. This enhances the credibility and dependability of the instrument for measuring students' needs regarding the use of GenAI in Arabic language learning.

DEPTH RESULT DISCUSSION

1) Content Validity through Expert Evaluation (Kappa Analysis)

Based on evaluations from three experts (Prof Ts Dr Janudin Sardi from UiTM Shah Alam, Dr. Mohamad Taufiq Abdul Ghani from Universiti Pendidikan Sultan Idris and, Dr. Ghazali Zainuddin from Universiti Kebangsaan Malaysia), the instrument achieved a high level of agreement, with Kappa values ranging from 0.68 to 0.82. This range indicates agreement levels between "good" and "very good," suggesting that the items are both relevant and appropriate for measuring the intended constructs.

Notably, the highest Kappa value (0.82) was recorded for the open-ended section, indicating strong expert consensus that these items effectively capture deeper insights into students' needs, expectations, and aspirations regarding the use of GenAI.

From a methodological perspective, these findings indicate that the instrument demonstrates strong content validity, the constructs measured are well-aligned with the research objectives, and there was minimal ambiguity, redundancy, or overlap in item construction.

The high level of expert agreement also suggests that the instrument is not only technically sound but also contextually appropriate for Arabic language learning in a technology-enhanced environment. This reinforces the instrument's suitability for similar future studies.

2) Instrument Reliability (Pilot Study with 36 Students)

The pilot study results revealed Cronbach's Alpha values ranging from 0.871 to 0.957, indicating a very high level of internal consistency as follows:

Section B (Lacks, $\alpha = 0.871$) – This reflects strong reliability in measuring students' lack of exposure and training. The consistency suggests that respondents interpreted these items similarly, highlighting shared gaps in knowledge and experience.

Section C (Necessities, $\alpha = 0.954$) – This very high value indicates a strong consensus among students regarding the importance and necessity of GenAI in their learning process.

Section D (Wants, $\alpha = 0.957$) – The highest value recorded, suggesting that students have highly consistent views regarding their desire to use GenAI in learning Arabic.

These high alpha values demonstrate that the instrument produces stable and consistent responses, thereby strengthening its reliability. However, it is also important to note that extremely high values may indicate item redundancy. In this study, such redundancy is considered acceptable due to the specificity and depth required in measuring closely related constructs within a focused domain.

These results suggest that students perceive GenAI as a highly practical and efficient tool that accelerates the learning process, reduces cognitive load, and facilitates comprehension of complex texts. Interestingly, despite acknowledging limited exposure and formal training, students still demonstrated a high readiness and willingness to adopt GenAI in their learning.

CONCLUSION

In conclusion, the pilot study successfully established the reliability of the research instrument. The content validity evaluation recorded an overall Kappa coefficient of 0.72, indicating a good level of agreement among three field experts. Subsequently, the pilot reliability test demonstrated that all constructs achieved Cronbach's Alpha values at or above 0.871, confirming strong internal consistency across the Lacks, Necessities, and Wants dimensions. Based on the Cronbach's Alpha analysis, the instrument was deemed suitable for use in the actual data collection phase without requiring major modifications.

The primary contribution of this study lies in the development and validation of a culturally and contextually appropriate instrument for examining GenAI adoption among Arabic language learners in Malaysian public universities. This instrument fills a notable gap in the literature, where empirical tools targeting this specific population have been limited.

However, the study acknowledges certain limitations, including the relatively small pilot sample size ($n = 36$) and geographic restriction to the Klang Valley. Future research is recommended to expand the sample to a broader range of public universities across Malaysia and to investigate the longitudinal impact of GenAI integration on Arabic language proficiency. Furthermore, qualitative approaches may be employed to enrich the understanding of learner experiences and attitudes towards GenAI usage.

ACKNOWLEDGEMENT

We are grateful for the research grant provided by Faculty of Major Language Studies, Universiti Sains Islam Malaysia (research code: PPPI/PTJ/FPBU/USIM/14025).

REFERENCES

- Alayba, A. M. (2025). *Arabic Natural Language Processing (NLP)*. Computers, 14(11), 497. <https://doi.org/10.3390/computers14110497>
- Ali, M., & Omar, R. (2024). *Overcoming Hurdles In Arabic Language Learning Through Genai Innovations*. Journal of Modern Education.
- Al-Mansouri, A., & Hassan, S. (2024). *Challenges In Mastering Complex Arabic Grammar And The Role Of AI*. Arabic Education Review.
- Al-Onaizan, Y. (2024). *Continuous Adaptation Of AI Systems To Evolving Linguistic Structures*. AI & Language Research Journal.
- Al-Shehri, S. (2024). *Increasing Student Engagement In Arabic Language Learning Through GenAI*. Digital Pedagogies in Language Learning.
- Browne, R. H. (1995). *On The Use Of A Pilot Sample For Sample Size Determination*. Statistics in Medicine, 14(17), 1933–1940. <https://doi.org/10.1002/sim.4780141709>
- Chan, K., & Lee, M. (2023). *Ethical Guidelines For AI Usage In Higher Education: Preventing Academic Misconduct*. Ethics and Technology.
- Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, And Mixed Methods Approaches* (5th ed.). SAGE Publications.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

- Hassan, S., Omar, R., & Malik, F. (2023). *Lack Of Immersive Linguistic Environments In Public Universities*. Malaysian Journal of Education.
- Hutchinson, T., & Waters, A. (1987). *English For Specific Purposes: A Learning-Centred Approach*. Cambridge University Press.
- Ismail, N., Rahman, A., Zulkifli, H., & Bakar, M. A. (2025). *Student-Centred Pedagogical Approaches In Arabic Language Learning*. Asian Education Review.
- Jaffar, M. N., & Nazrol, N. A. H. (2025). *The Use Of Generative Artificial Intelligence (Genai) Applications Among The Arabic Students: A Preliminary Study*. Journal of Advanced Islamic and Legal Research, 2(2), 191–196.
- Kamal, M. H. (2026a). *Bibliometric Analysis Of Research On Gifted Education*. Pertanika Journal of Social Sciences & Humanities, 34(1), 135–154. <https://doi.org/10.47836/jssh.34.1.07>
- Kamal, M. H. (2026b). *A Scoping Review Of Flipped Classroom Approaches In Arabic Teaching*. Ijaz Arabi Journal of Arabic Learning, 9(1). <https://doi.org/10.18860/ijazarabi.v9i1.36740>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). *Chatgpt For Good? On Opportunities And Challenges Of Large Language Models For Education*. Learning and Individual Differences, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Khalil, M., & Er, E. (2023). *Will Chatgpt Get You Caught? Rethinking Of Plagiarism Detection*. arXiv. <https://doi.org/10.48550/arXiv.2302.04335>
- Landis, J. R., & Koch, G. G. (1977). *The Measurement Of Observer Agreement For Categorical Data*. Biometrics, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). *Generative AI (Genai) In The Language Classroom: A Systematic Review*. Interactive Learning Environments. <https://doi.org/10.1080/10494820.2025.2498537>
- Linur, R., Rusdi, Lubis, T., & Rohman, K. S. (2025). *Student Perception Analysis Of AI's Impact On Arabic Language Learning: A Personality Perspective*. Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab, 17(2), 312–328. <https://doi.org/10.24042/fya8h073>
- Moustafa, A. H. A., Al-Hamad, M. F., Qureshi, M. A., & Qadir, J. (2026). *Generative AI For Learning And Teaching Arabic: Opportunities, Challenges, And Ethical Considerations*. IEEE Access, 14. <https://doi.org/10.1109/ACCESS.2026.3651864>
- Moustafa, A. H. A., Al-Hamad, M. F., Qureshi, M. A., & Qadir, J. (2026). *Generative AI For Learning And Teaching Arabic: Opportunities, Challenges, And Ethical Considerations*. IEEE Access, 14. <https://doi.org/10.1109/ACCESS.2026.3651864>
- Muhammad Hakim, K. (2026). *Assessment Of Arabic Language Skills Using AI Tools Among Gifted Talented Students*. Malaysian Journal of Social Sciences and Humanities (MJSSH), 11(1), e003761. <https://doi.org/10.47405/mjssh.v11i1.3761>
- Okulicz-Kozaryn, W. (2026). *Generative AI As A Student Research Assistant: The Relationship Of Academic And Research Practices In Higher Education*. Sci, 8(1), Article 24. <https://doi.org/10.3390/sci8010024>
- Saarela, M., Aljohani, N. R., Alhadlaq, A., Alowedi, N. A., & Aljohani, A. J. (2026). *A Meta-Analysis Of Generative Artificial Intelligence Effects On Language Learning*. Education And Information Technologies. Advance online publication. <https://doi.org/10.1007/s10791-026-10015-1>
- Yaacob, M., & Abu Bakar, K. (2018). *Student-Centered Teaching Approach Combined With Coaching And Mentoring Elements In Arabic Language Speaking Proficiency*. Open Access Library Journal, 5, e4684. <https://doi.org/10.4236/oalib.1104684>
- Yatri, D., Anugerahwati, M., & Setyowati, L. (2023). *Artificial Intelligence (AI) In Language Learning (English And Arabic Class): Students' And Teachers' Experience And*

- Perceptions. Transformational Language, Literature, and Technology Overview in Learning (TRANSTOOL)*, 3(1), 1–12. <https://doi.org/10.55047/transtool.v3i1.1338>
- Yilmaz, R., & Karaoğlu Yılmaz, F. G. (2023). *The Effect Of Generative Artificial Intelligence (AI)-Based Tool Use On Students' Computational Thinking Skills, Programming Self-Efficacy And Motivation*. *Computers and Education: Artificial Intelligence*, 4, 100147. <https://doi.org/10.1016/j.caeai.2023.100147>
- Yilmaz, R., & Yilmaz, F. (2023). *Enhancing Academic Productivity Through GenAI Materials*. *Educational Technology Review*.
- 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>

Disclaimer

Opinions expressed in this article are the opinions of the author(s). Al-Qanatir: International Journal of Islamic Studies shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.