

Language Skill Development in Non-Arabic Majors: A Longitudinal ICP Study

DOI: <https://doi.org/10.33806/ijaes1084>

Ibnu Rawandhy N. Hula
IAIN Sultan Amai Gorontalo, Indonesia

Ana Mariana
Universitas Muhammadiyah Gorontalo, Indonesia

Alimudin Rivai
IAIN Manado, Indonesia

Nicky B. Lasanggo
IAIN Sultan Amai Gorontalo, Indonesia

Received: 10.4.2025

Accepted: 2.3.2026

Published Online: 5.3.2026

Abstract: This study aimed to analyze the effectiveness of the International Class Program (ICP) in improving four Arabic language skills—listening, speaking, reading, and writing—among non-Arabic majors at IAIN Sultan Amai Gorontalo, Indonesia. Using a quantitative longitudinal design, data were collected from 50 students at two proficiency levels (*Tamhidi/A-1* and *Mubtadi'/A-2*) through pre-, mid-term, and post-tests. The results showed an initial score of 54.39 percent (inadequate) for *Tamhidi/A-1* and 61.45 percent (average) for *Mubtadi'/A-2*. Student scores increased significantly by the mid-semester assessment, reaching 74.48 percent and 77.21 percent for the *Tamhidi/A-1* and *Mubtadi'/A-2* groups, respectively. However, relative weaknesses remained in writing, reading, and speaking skills for the *Tamhidi/A-1* group, and in listening for the *Mubtadi'/A-2* group. Overall, test scores showed significant progress, increasing by 20.09 percent from the initial evaluation to the mid-term, and by another 10.90 percent from the mid-term to the final, for a total increase of 30.99 percent. These results demonstrate that the ICP effectively improved the students' language skills. Therefore, this study suggests the integration of technology, the participation of native speakers, and the development of an adaptive curriculum to further enhance the program's effectiveness.

Keywords: International Class Program (ICP), language skill development, longitudinal study, major students, non-Arabic

1. Introduction

The rapid advancement of globalization is requiring higher education institutions to develop innovative curricula with the intent of equipping students and preparing graduates with the global skills needed in the fields of expertise (Purwanto, Hartono and Wahyuni 2023; Abo-Khalil 2024). Among the most momentous global

competencies is proficiency in foreign languages, which is invaluable in strengthening alumni competitiveness in the workplace. This proficiency also facilitates access to better educational opportunities at universities, both domestically and abroad, and enhances the reputation of institutional accreditation and educational programs (Ke, Junfeng and Xiaojing 2022).

The acceptance of foreign languages, specifically Arabic, is becoming increasingly significant as relations between Arabic-speaking countries and the international community improve (Yaseen et al. 2009; Hussein 2024). However, a number of educational institutions face challenges in improving foreign language skills for students who are learning languages other than Arabic.

When individuals attempt to master Arabic, those who do not focus on the primary studies often face additional challenges that can hinder their learning progress. Weak foundational language competencies could also lead to hurdles with grammar comprehension, the precise formulation of sentences, achieving lucid phonetic articulation, and diminished confidence in using Arabic in diverse circumstances (Al Busaidi 2015; Ahkam et al. 2025). Therefore, effective programs must be designed to accommodate the varying initial abilities of students to gradually master the Arabic language and gain confidence in using it (Andrian and Yul 2023).

The International Class Program (ICP) has become a promising program to address foreign language learning challenges for non-native speakers (Irham, 2023). This is a lecture program designed to meet four primary indicators, namely (1) using a foreign language as the medium of instruction, (2) including guest lecturers (native speakers) or lecturers with foreign qualifications, (3) adopting an internationally standardized curriculum, modules, as well as teaching materials, and (4) providing adequate learning support, facilities, and infrastructure (Famularsih et al. 2022).

Although many institutions have offered foreign learning opportunities for students, the effectiveness of the programs in improving listening, speaking, reading, or writing language capabilities among non-Arabic major students is in doubt. Previous studies have presented that students in non-major programs have difficulties in practicing and enhancing proficiency in learning foreign languages, specifically where these programs are ill-conceived or do not respond to the learning needs (Kos 2023). Therefore, it is urgently necessary to consider how the ICP can contribute to addressing the issues and enhancing foreign language proficiency for non-Arab language major students.

In Indonesia, the implementation of ICP is still hampered in eastern regions such as Sulawesi, Kalimantan, Nusa Tenggara, and Papua (Mudzakkir, Sukoco and Suwignjo 2022). As an illustration, IAIN Sultan Amai Gorontalo, which has been established for 27 years (1997–2024), has never seriously submitted an ICP program. This leads to low foreign language proficiency among students, specifically those majoring in languages such as Arabic and English Education. The curriculum is also still lacking in the implementation of international teaching

models, making it difficult for students to compete at the national and international levels in terms of language skills (Hilmi and Abdul Hamid 2025).

To address the issue, decision makers at IAIN Sultan Amai Gorontalo have started designing and implementing the ICP to improve the quality of foreign language teaching. The program has been running for a year, applying a “one for all” curriculum designed to meet the needs of all students, regardless of their background or initial abilities (Astuti, Shalawati and Hadijah 2022). The policy has received a positive response from students, as it provides significant benefits, such as improving the university’s ranking, enhancing graduates’ competencies, and stimulating the four language skills among non-Arab language major students (Alazemi 2020).

Although there have been many studies evaluating the effectiveness of foreign language learning programs, the majority have focused on students majoring in languages or in educational institutions in Arabic-speaking regions. For example, analysis conducted by Sourani et al. (2023) explored the extent to which the CLT method was effective in Arabic-speaking countries, including Lebanon and Jordan, by emphasizing the ongoing discussion and criticism surrounding the method, similar publication by Chen (2024) also emphasized the importance of adaptable programs, but the main focus was only on students with a formal linguistic education background. This situation created a significant study gap regarding methods to improve programs such as ICP for students who did not major in Arabic and who generally had lower language proficiency levels at the outset.

In addition, information regarding the implementation of ICP in Islamic universities in Indonesia is still minimal. The publication conducted by Mudzakkir et al. (2022) showed that the implementation of ICP in State Islamic Religious Universities (PTKIN) in eastern Indonesia, such as in Sulawesi, Kalimantan, and Papua was almost non-existent. Even in Java, the ICP program focuses more on improving English language skills than Arabic. This study aims to fill the gap by providing empirical data on the effectiveness of ICP in improving Arabic language skills for students who do not major in Arabic at IAIN Sultan Amai Gorontalo.

Previous studies have also neglected the longitudinal dimension when evaluating language teaching programs. For example, Izzi et al. (2022), conducted a study on the implementation of the ICP at Surabaya State University. Similarly, Kern (2006) emphasized the significance of integrating technology into foreign language teaching and did not assess the long-term impact. To address the limitations, this study adopts a longitudinal method that tracks student progress from pre-to post-test, thereby offering a more comprehensive understanding of the effectiveness of the ICP over time (Herpen 2024).

Another gap addressed in this study is the lack of focus on specific obstacles experienced by students in non-Arabic language programs. Furthermore, this study will investigate methods in which the presence of native-speaking guest lecturers and learning materials meeting international standards can contribute to overcoming the obstacles.

The originality of this study can be observed from the longitudinal method applied to examine the effectiveness of the Integrated Curriculum Program (ICP) in improving language skills among students who did not choose Arabic as the major related to foreign language skills. Based on a series of tests that have been conducted, such as initial, mid-, and final testing, this study will conduct an in-depth review of the subjects.

Furthermore, this study aims to provide useful insights for policymakers and education program managers with valuable insights for designing curricula better suited to students' needs, as well as concrete suggestions for creating more efficient ICP programs in the future. To regularly investigate the effectiveness of the ICP in enhancing Arabic language proficiency among non-Arabic major students, the analysis is guided by the following questions:

- (1) To what extent does the ICP effectively enhance the four language skills, namely, listening, speaking, reading, and writing, among non-Arabic major students at IAIN Sultan Amai Gorontalo?
- (2) What is the curve of students' skill development across the three assessment points (pre-test, midterm, and post-test)?
- (3) Which definite language skills determine the most significant improvement, and also continue to present challenges for participants throughout the program?
- (4) What aspect, as disclosed through data analysis, adds to the examined patterns of increase and persistent weaknesses in language acquisition?

These questions are designed to contribute an extensive and specific understanding of the ICP's impact as well as to inform eventual curriculum development and program authority.

2. Literature review

2.1 Language learning in a longitudinal framework

The analysis was grounded in basic theories of foreign language acquisition, emphasizing the gradual, dynamic, and context-sensitive nature of language development and characteristics that could only be fully captured through longitudinal observation (Foster et al. 2025). Therefore, this study adopted a longitudinal framework as a methodological choice and a theoretical necessity that correlated with the core assumptions of the guiding theories and was empirically supported by recent longitudinal analysis in second language acquisition (Guirado-Moreno et al. 2025; Sun, Alhowail and Beckmann 2025). The following theoretical frameworks provided the conceptual backbone of this study.

(1) Input-interaction-output hypothesis

The Input-Interaction-Output Hypothesis posited that language acquisition was facilitated by comprehensible input, important social interaction, and convenience for output (Lu et al. 2022). In the ICP context, bilingual situations included students (Indonesian–Arabic/English), who adopted

international curricula and engaged directly with native-speaking instructors, which were conditions designed to maximize linguistic input and interaction (Lorenzo, Granados and Ávila 2019; Bani Younes et al. 2023). However, the actual impact of these elements evolved only over time. Longitudinal studies such as those by Renaud, Jauvin and Béliveau (2025) established that continued and repeated exposure, not isolated instances, led to measurable gains in proficiency. This validated the ICP's immersive design and forced the use of pre-, mid-, and post-tests to track incremental development correlated with behaviourist principles of observable advance (Novawan et al. 2020).

(2) Dynamic systems theory (DST) in language development

DST provided a groundwork for understanding that progress in language learning was non-linear, complex, and influenced by assorted internal and external factors such as motivation, learning strategies, and direct experiences. In this study, DST served as the basis for analyzing participants' language proficiency development during the implementation of the ICP, along with fluctuations and patterns of improvement at each evaluation stage (pre-test, midterm, post-test) (Banihashem et al. 2022).

(3) Skill acquisition theory

This theory describes the renewal process from declarative ability (language knowledge) into procedural proficiency (the fluent and automatic use of language). In the longitudinal design, this theory helped explain how intensive practice, positive reinforcement, and the use of technology within the ICP were able to increase students' practical ability to use the Arabic language effectively (Fuss and Navarro 2023). The longitudinal perspective further enabled evaluation of the long-term effectiveness of Communicative Language Teaching (CLT) (Robat et al. 2021). In an equal context of content-and-language unified learning (CLIL), a recent longitudinal investigation showed that students' tracks reported higher levels of positive emotions and motivation than those in non-CLIL classes. However, similar effects were mediated by the initial vocabulary knowledge (Mettewie et al. 2024).

2.2 ICP

The ICP was a government initiative program in Indonesia aimed at enhancing the global competitiveness of higher education institutions. It was designed to produce graduates proficient in foreign languages through elite or international classes that incorporated both national and international curricula (Winarti 2017). There were three variants of international courses, namely (1) Internationally Cooperative, (2) Internationally Active, and (3) Internationally Operative.

In Indonesia, the ICP predominantly adopted foreign curricula. For English language instruction, many institutions implemented the Cambridge International

Examinations (CIE) curriculum. Nufus et al. (2022) argued that there were five indicators of the implementation of the ICP, namely (1) using Arabic/bilingual or complete Arabic language introduction, (2) using foreign language teaching materials and materials, (3) taught by native speaker guest lecturers both offline and online, (4) using complete learning facilities and infrastructure, in the form of computers and laboratories, and (5) having tests and evaluations of four language skills (Nufus et al. 2022).

In the implementation stage, students participating in the program must undergo a placement test with four detailed level labels. At the end of the program, participants must meet the ICP graduation indicators by taking a series of oral and written tests using a Computer-Based Test (CBT) following the Test of Arabic as a Foreign Language (TOAFL) pattern with a minimum score of 500 (Pimada et al. 2023).

To strengthen the direct relevance to ICP practices in Indonesia, recent empirical evidence showed the following: (a) student experiences with TOAFL (Rahman and Supriani 2023) and (b) the implementation of CLT in Indonesia's Kurikulum Merdeka (Ramadhani, Nidi and Zahra 2024).

3. Method

3.1 Types of studies

This study used a quantitative longitudinal design, which was particularly suited for tracking innovation in language proficiency over time (Jackson 2024). The longitudinal method was preferred not only for the methodological rigor but also because it correlated with the theoretical base of DST and Skill Acquisition Theory. This postulated that language development was a time-sensitive and non-linear process requiring repeated amounts to capture authentic trajectories (Xu and Yang, 2024). Quantitative data were used to assess the program's effectiveness through numerical values, enabling statistical correlation across three time points, namely pre-test, midterm, and post-test.

3.2 Sample

This study consisted of 50 students from the Islamic Education Study Program who participated in the ICP, representing 30 percent of the total population (176 students). These 50 students were divided into two levels (*Tamhidi*/A-1 and *Mubtadi*'/A-2). The analysis was conducted at the Language Development Center of IAIN Sultan Amai Gorontalo for up to 10 days during the ICP.

The location was selected due to the following reasons. (1) The ICP was integrated and implemented at the Language Development Center, which was a pilot program of IAIN Sultan Amai Gorontalo starting in 2024. (2) The ICP had clearly established policies and standardized operational guidelines that guided its implementation. (3) The ICP at the Language Development Center was supported by IT facilities, including 50 computer units, two internationally-standardized textbooks, pre-test, midterm, and final test instruments, as well as oral test devices

and CBT. (4) The ICP was designed to be a flagship program that could enhance accreditation rankings and Key Performance Indicators in the field of language, and (5) The ICP was conducted with adequate funding support, bringing in presenters from overseas, making it essential to assess the effectiveness in improving students' language proficiency. To ensure sample representativeness and curtail selection bias, participants were selected using stratified random sampling based on the initial placement test scores, which further ensured proportional representation from both proficiency levels (Fkih 2022; Nnadozie and Ngoni 2023).

3.3 Data collection procedure

Data were collected using a single type of data elicitation tool, namely quantitative (test). A regulated quantitative instrument was also used for data collection, which was a language proficiency test administered at three intervals, namely pre-test (baseline), midterm (after 5 days), and post-test (after 10 days).

The test consisted of a set of written questions prepared for quantitative measurement, including a pre-test, a midterm, and a final, conducted over four days. The formats comprised oral tests and CBT. The materials were designed and developed from the ICP teaching instruments, covering the four language skills.

Table 1. Assessment of foreign language learning effectiveness

Test	Quality of Foreign Language Learning	Number of Participants	Number of Questions	Format
Pre-test	Listening/ <i>Istima'</i>	50	10	Writing
	Speaking/ <i>Kalam</i>		5-6	Oral
	Reading/ <i>Qira'ah</i>		5-6	Oral
	Writing/ <i>Kitabah</i>		5-6	Writing
Midterm	Listening/ <i>Istima'</i>	50	30	Writing
	Speaking/ <i>Kalam</i>		30	Oral
	Reading/ <i>Qira'ah</i>		30	Oral
	Writing/ <i>Kitabah</i>		30	Writing
Final Tes	Listening/ <i>Istima'</i>	50	50	CBT
	Speaking/ <i>Kalam</i>		50	Oral
	Reading/ <i>Qira'ah</i>		50	CBT
	Writing/ <i>Kitabah</i>		50	CBT

The test results were processed using a Likert scale and the university grading system, which led to Table 2.

Table 2. ICP university grading system

Scale	Quality	Grade Description	US Grade
7	91 – 100	Excellent	A
6	81 – 90	Very Good	A -

5	71 – 80	Good	B
4	61 – 70	Average	B -
3	51 – 60	Deficient	C
2	41 – 50	Unsatisfactory	C -
1	< – 40	Fail	D

3.4 Data analysis procedure

To measure the effectiveness of the ICP program implementation, this study analyzed pre-test, midterm, and final test/CBT scores, as well as oral test results for the four language skills. The data were then processed for descriptive statistical analysis to obtain a numerical or visual representation of the characteristics of a population or sample. The results of the data analysis were subsequently synthesized by integrating quantitative and qualitative findings on the outcomes of the foreign language improvement program for non-Arabic students, including its impacts, benefits, and anticipated changes. The analysis also identified areas and opportunities that required follow-up action by policymakers in future program planning and implementation.

Data analysis was conducted in two phases. Phase 1: Descriptive Statistics were computed for each of the four language skills (listening, speaking, reading, writing) across all three assessment points (pre-test, midterm, post-test) and for both participant groups (*Tamhidi/A-1* and *Mubtadi'/A-2*). For each skill and time point, the analyst calculated the mean to show central tendency, the standard deviation to reflect score variability among participants, and the percentage gain using the formula: $[(\text{Post-test Score} - \text{Pre-test Score}) / \text{Pre-test Score}] \times 100$ percent (Crehan Koehler and Slakter 2023). This formula was used to quantify the magnitude of improvement from pre-test to midterm, midterm to post-test, and cumulatively from pre-test to post-test. Participant performance was further categorized using the institutional grading scale (Table 3: 7-point scale ranging from “Fail” to “Excellent”) to determine persistent strengths and weaknesses across skill domains systematically. Phase 2: Inferential Statistics (Added for Robustness). Paired Samples t-tests were conducted to determine whether improvements from pre-test to midterm, and midterm to post-test, were statistically significant ($p < 0.05$) (Pallant 2020). All statistical analyses were accomplished using SPSS version 26.

3.5 Reliability and validity

The credibility of the results depended on the reliability and validity of the measurement tools used, namely the pre-test, midterm, and final test. To provide content validity, the test instruments were meticulously developed from the ICP’s teaching materials (Madadzadeh and Bahariniya 2023). These adjustments ensured that the tests accurately measure the knowledge and skills taught throughout the program, particularly the four language skills. To establish construct validity, the exam format followed the widely recognized TOAFL pattern, which was a standard

model known for its proven ability to comprehensively assess Arabic language proficiency (Alzahrani 2024).

The study was conducted on a separate group of 20 students to ensure the quality and consistency of the instrument. The results of the test showed the validity of all test items, with each item suggesting a correlation coefficient (r) above the critical value of 0.4438. This suggested that every question effectively measured the intended construct. The critical r value at $\alpha = 0.05$ was 0.444 with $n = 20$ and $df = 18$, confirming that all items exceeded the minimum validity threshold. Items with $r \geq 0.30$ were also considered to have acceptable discriminative power. Furthermore, the instruments showed high reliability. The internal consistency of the test was confirmed by a Cronbach's Alpha coefficient of 0.87, which exceeded the recommended threshold of 0.70 (Zakariya 2022), showing strong internal consistency and reliability across the test items. The test items also showed high discriminatory power, confirming the ability to distinguish between students with different levels of proficiency. The structured and standardized method in the design and delivery of this test minimized potential variations and enhanced the overall credibility of the results (Cheung et al. 2024).

3.6 Ethical considerations

This study had obtained official ethical approval from the Research Ethics Committee of IAIN Sultan Amai Gorontalo, Indonesia (Reference Number: IX/67.b/2024), ensuring full compliance with international ethical standards for research, including human participants. All 50 academic participants provided written consent after receiving a detailed explanation of the objectives, methodology, and the entitlement to disengage at any moment without scholarly repercussions or additional ramifications. To maintain participant confidentiality, all data were anonymized using numerical coding (P01–P50), and raw data were stored in password-secured digital files accessible only to the principal team. Participation was entirely voluntary with no incentives or academic penalties associated with the study.

4. Results

4.1 Limited trial pre-test results for *Tamhidi/A-1* and *Mubtadi'/A-2* classes

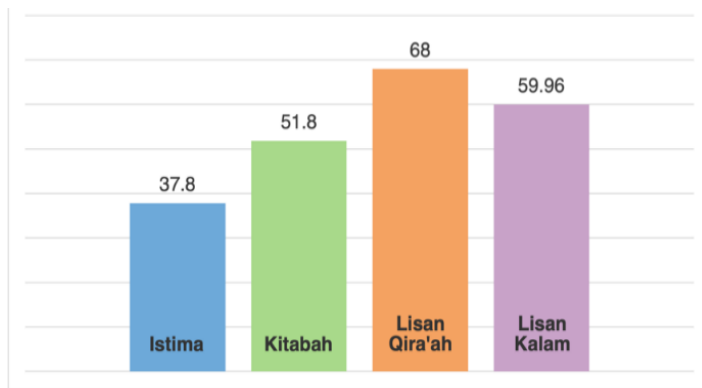


Figure 1. Average pre-test results for class A-1 trial

Each of the 25 participants who took the pre-test in the *Tamhidi/A-1* class of the ICP program was assessed based on four components, namely *istima'*, *kitabah*, *qira'ah*, and *kalam*. The average score across all four components was 54.39, indicating an inadequate assessment.

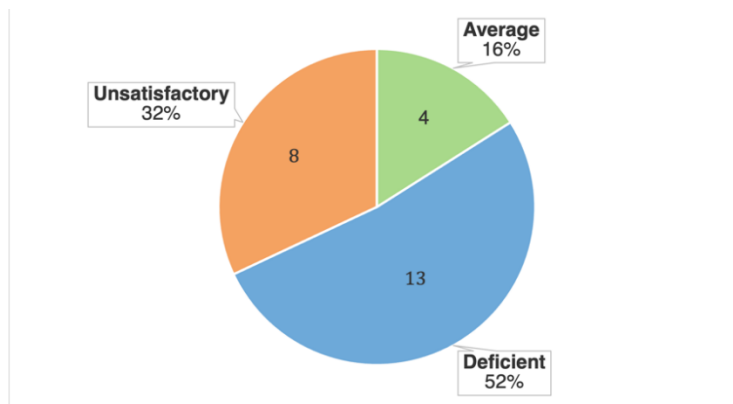


Figure 2. Percentage of pre-test results for ICP class A-1.

As shown in Figure 2, the 25 participants' ICP pre-test scores fall into three primary categories.

- (1) Average 16 percent: This category included four participants from *Tamhidi/A-1* who have a sufficient basic understanding. The students required additional attention to improve their language skills, specifically in speaking and advancing to a higher level.

- (2) Deficient 52 percent: This category consisted of 13 participants with poor results. The participants required intensive support in mastering listening and writing skills.
- (3) Unsatisfactory 32 percent: Participants in this category also performed poorly, requiring intensive support in mastering the four language skills, mainly listening, writing, and speaking.

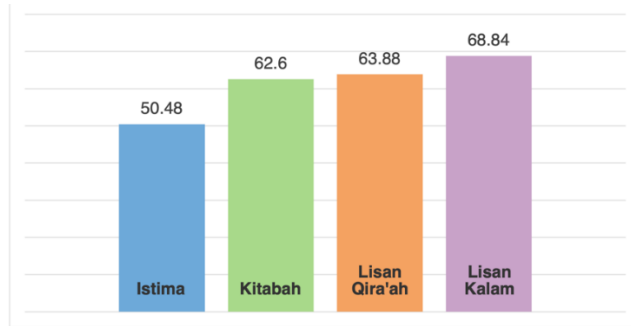


Figure 3. Average pre-test results for class A-2

The data in Figure 3 consisted of 25 participants who took the pre-test in the *Mubtadi'/A-2* class of the ICP program. Each participant was assessed based on four components, namely *istima'*, *kitabah*, *qira'ah*, and *kalam*. The average score across all four components was 67.45, which was classified as average.

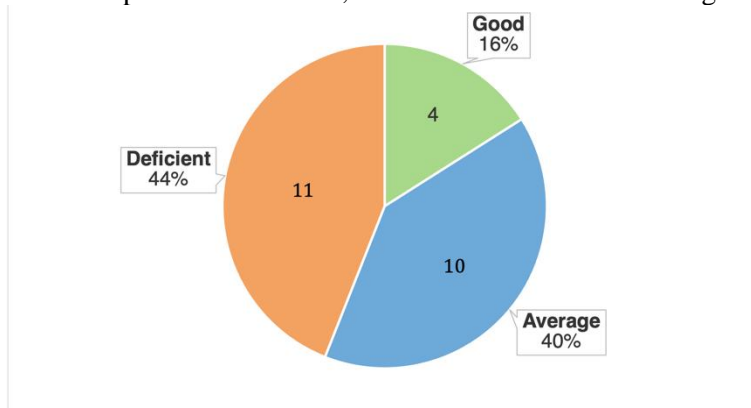


Figure 4. Percentage of pretest grades for ICP class A-2

Figure 4 showed that of the 25 participants who took the ICP pre-test, the results were divided into three categories as follows.

- (1) Good 16 percent: Participants in this category understood the material well. The students could proceed to a higher level with further reinforcement in several aspects.

- (2) Average 40 percent: Participants in this category have a sufficient basic understanding but need more attention to improve language skills for progress to a higher level.
- (3) Deficient 44 percent: Participants in this category showed less than satisfactory results. The students required additional intervention and support to improve their language skills. Remedial or supplementary programs were necessary to assist the participants.

4.2 Results of the midterm ICP trial

The ICP midterm test was an evaluation conducted to measure the understanding and skills of 50 participants in various topics and subtopics taught over 10 days with native speakers. The main objective was to assess participants' progress in mastering the material and to identify strengths and weaknesses in the four language skills, namely *qira'ah*, *kalam*, *istima'*, and *kitabah*. Each level (A-1 and A-2) was tested based on the learning modules studied, using an exam format consisting of 60 diverse questions, including multiple-choice, essays, and structured assignments that covered all aspects of language proficiency.

The midterm exam results were used to develop more effective teaching strategies and provide feedback to students on areas for improvement. In accordance with the examination of outcomes, educators could modify the pedagogical strategies to accommodate the particular requirements of learners in correlation with the achievement classifications, namely Outstanding, Commendable, Satisfactory, Mediocre, Inadequate, or Unsuccessful. Students with poor performance received targeted support before advancing. Therefore, the midterm evaluations served as an important tool for improving educational quality and fostering student development.

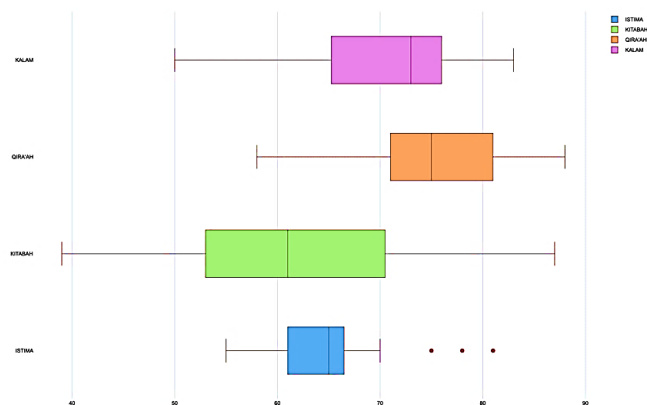


Figure 5. Midterm results for ICP class A-1

Figure 5 showed that students' foreign language skills increased after attending the ICP for 10 days. The Class *Tamhidi/A-1* midterm assessment, including 25 participants, yielded an overall total score of 6,852 across all four language skills with an average score of 68.5, classified as Average. Based on skill level, the average scores were 65.2 for *istima'*, 62.04 for *kitabah*, 76.2 for *qira'ah*, and 70.6 for *kalam*. These results showed uneven but generally positive development, with *qira'ah* performing the strongest and *kitabah* the weakest.

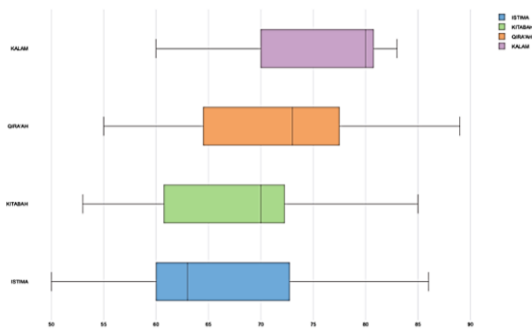


Figure 6. Midterm results for ICP class A-2

In terms of score distribution, three participants 12 percent achieved Very Good scores (81–90), five participants 20 percent were categorized as Good (71–80), and 17 participants 68 percent fell within the Average category (61–70). No participants were classified as Deficient, Unsatisfactory, or Fail at the intermediate stage. Overall, the midterm results showed that participation in the ICP contributed to measurable improvement in foreign language proficiency after 10 days of instruction. The results also showed the need for continued and targeted pedagogical support, particularly in writing and listening, to promote more balanced skill development across all language domains.

The Figure showed that students' foreign language proficiency improved after participating in the ICP for 10 days. The midterm assessment of class *Mubtadi'/A-2*, including 25 participants, yielded an overall total score of 7,102 across the four language skills with an average score of 71, which fell into the Good category. By skill, the average scores were 66.2 for *istima'*, 69.0 for *kitabah*, 72.5 for *qira'ah*, and 76.4 for *kalam*. These results showed relatively balanced development with *kalam* and *qira'ah* evolving as the strongest areas. Regarding score distribution, four participants 16 percent achieved Very Good scores (81–90), eight participants 32 percent were classified as Good (71–80), and 13 participants 52 percent fell into the Average category (61–70). No participants were categorized as Deficient, Unsatisfactory, or Fail at the midterm stage.

Overall, the midterm results showed that participation in the ICP contributed to measurable improvement in foreign language proficiency after 10 days of instruction. However, the results also outlined the need for ongoing and targeted

pedagogical support, particularly in writing and listening, to foster more balanced skill development across all language domains.

4.3 Results of the post-test for the ICP

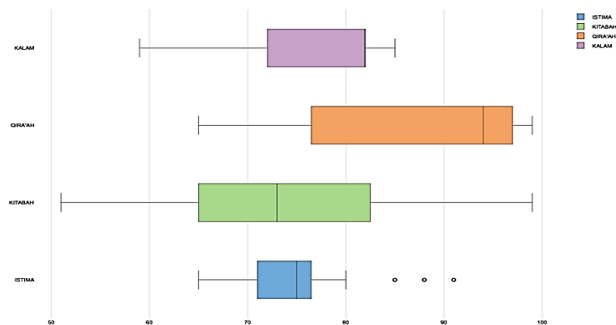


Figure 7. Final test results for ICP class A-1

The data from the post-test in Figure 7 consisted of the assessment results of the four language skills, namely *istima'* (50 questions), *kitabah* (50 questions), *qira'ah* (50 questions), and *kalam* (50 questions). These four skills were assessed over two days using oral and CBT formats. The post-test was taken by 25 *Tamhidi/A1* class participants who had participated in the ICP activities for 10 days, producing the following data.

Average Scores were *istima'*: 75.2, *kitabah*: 74.04, *qira'ah*: 87.62, and *kalam*: 77.44. Standard Deviations showed *istima'*: 6.04, *kitabah*: 13.35, *qira'ah*: 11.32, and *kalam*: 7.24. Standard deviation was a statistical measure of the variation or spread of scores in a data set. In the context of post-test data analysis, the explanation was as follows.

Interpretation: (a) A low standard deviation implied that participants' scores tended to be close to the average, showing consistency in language proficiency. (b) A high standard deviation showed that participants' scores were further from the average, indicating greater variation in foreign language proficiency.

The standard deviations provided insights into the consistency and variation in each skill. *Istima'* and *kalam* skills showed good consistency, while *kitabah* and *qira'ah* showed a need for improvement, particularly in *kitabah* and *kalam*. Looking at the distribution of scores and the classifications, the analysis showed that Very Good (80-89) recorded 12 participants, Good (70-79) indicated 9, and Average (60-69) suggested 4.

The post-test results for class *Tamhidi/A-1* showed excellence in *qira'ah*, which was the highest average score, suggesting excellent reading abilities. The participants were able to understand and analyze texts effectively. Similarly, in the

kalam skill, the average score was good, showing that students had a strong understanding with relatively small variation among participants.

Weaknesses were perceived in the skill of *istima'*, where the average score was relatively low, showing a need to focus on improving writing skills. The presence of an upper outlier suggested that some participants performed better than most others. Furthermore, *kitabah* skill showed the lowest average among these aspects, and the highest standard deviation indicated significant variation among participants, with some experiencing difficulties writing well.

Post-test results for 25 participants in *Tamhidi/A-1* class showed strong skills in *qira'ah* and *kalam*, with speaking as the primary strength. However, *kitabah* showed a significant need for improvement. *Istima'* also showed low variability, with some participants performing well while others needed reinforcement. Participants could improve overall language skills with the right method and a focused learning program.

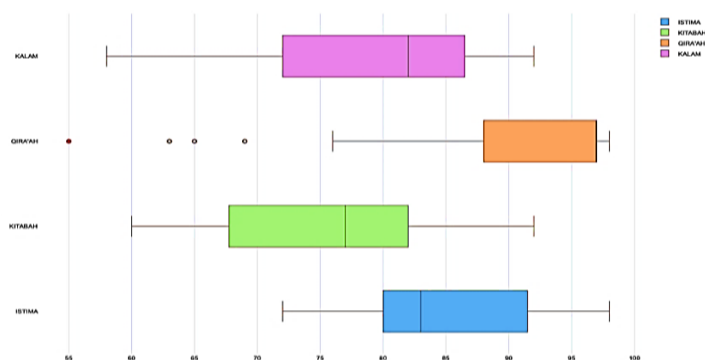


Figure 8. Results of the post-test for ICP class A-2

The data from the post-test consisted of the results of the assessment of four language skills, namely *istima'* (50 questions), *kitabah* (50 questions), *qira'ah* (50 questions), and *kalam* (50 questions). These four skills were assessed over two days using both oral test models and CBT. The post-test was attended by 25 participants from the *Mubtadi'/A-2* class who had participated in the ICP activities for 10 days, leading to the following data.

Average Scores in *istima'*: 85.2, *kitabah*: 76.44, *qira'ah*: 89.72, and *kalam*: 79.88. Standard Deviations for Each Skill recorded (1) *Istima'*: 8.04, (2) *Kitabah*: 9.53, (3) *Qira'ah*: 13.40, and (4) *Kalam*: 8.84. The standard deviation represented a statistical measure of the variation of scores within a data set. In the context of post-test data analysis, the following was an explanation:

Interpretation: (a) When the standard deviation was low, then the participants' scores tended to be close to the mean, showing consistency in language ability. (b)

When the standard deviation was high, the participants' scores were more spread out than the mean, showing more significant variation in foreign language ability.

The distribution of scores and the classifications showed that 5 participants were Excellent (90 and above), 13 were Very Good (80-89), 6 were Good (70-79), and 1 was Average (60-69). The post-test results showed mastery of *qira'ah* material, where the highest average score showed excellent reading skills. The participants could understand and analyze texts well. However, some participants needed more practice reading diverse long texts to improve reading comprehension. In the aspect of *istima'*, although the average score was lower compared to *qira'ah*, participants showed good listening comprehension. The students needed more practice listening to varied audio materials to help become accustomed to different accents and intonations.

Weaknesses were evident in *kitabah* skills, as the lowest average showed a need to improve writing skills. There was significant variation among participants, showing that some were experiencing difficulties. In the *kalam* aspect, although the average was not bad, there was variation among participants, with some struggling to express the ideas verbally, often sounding stiff and afraid to make mistakes. Therefore, additional practice in public speaking and group discussions was necessary to enhance confidence and speaking abilities among participants.

The post-test results of 25 participants from class *Mubtadi' A-2* showed that participants have good skills in *qira'ah* and *istima'*. However, the *kitabah* skills showed a significant need for improvement. In contrast, *kalam* showed variation, with some participants performing well and others requiring more practice to master vocabulary, style, and expression that could facilitate interaction with peers and native speakers. Figure 9 presents the comparative data of the pre-test, midterm, and post-test results.

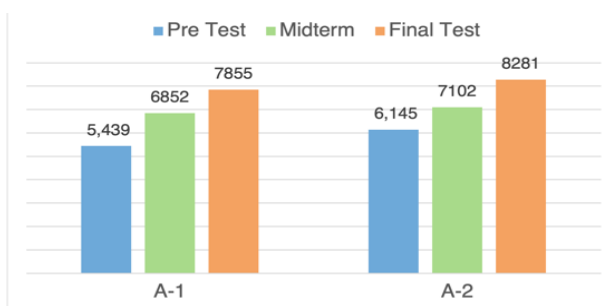


Figure 9. ICP participant evaluation result score

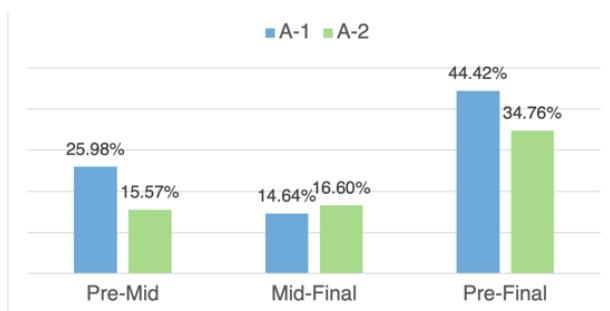


Figure 10. Percentage increase in evaluation scores of ICP participants

The data in Figure 10 showed a significant increase between the pre-test, midterm, and post-test as follows. (1) *Tamhidi*/A-1 class: pre-test to midterm: 25.98 percent, midterm to post-test: 14.64 percent, and pre-test to post-test: 44.42 percent. (2) *Mubtadi*'/A-2 class: Pre-Test to midterm: 15.57 percent, midterm to post-test: 16.60 percent, and pre-test to post-test: 34.76 percent. The data showed that both groups of participants suggested significant improvement from the pre-test to the post-test, with both *Mubtadi*'/A-2 and *Tamhidi*/A-1 classes indicating progress in higher categories. This showed that the design and implementation of the ICP were effective in improving participants' foreign language skills.

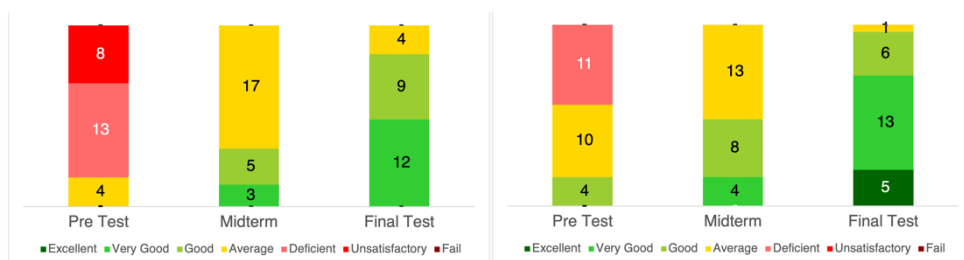


Figure 11. ICP test score predicate A-1 and A-2

None of the 50 participants who took the pre-test achieved a very good classification. After participating in the ICP program, the post-test results showed that 5 participants achieved very good scores. Initially, no participants were in the Very Good category, which later increased to 25 participants.

5. Discussion

The aims of the ICP are to enhance foreign language acquisition with an appropriate focus on Arabic. The publication shows significant progress in Arabic learners among non-Arabic majors. Additional evaluations can be conducted at three distinct testing phases, namely pre-test, mid-test, and post-test.

5.1 The increase from pre-test to midterm

Based on data, there is an increase of 25.98 percent for the *Tamhidi/A-1* class and 15.57 percent for the *Mubtadi'/A-2* class from pre-test to midterm. This significant increase shows that the teaching methods implemented in the ICP program effectively provide a foundational understanding of the Arabic language to students who previously had no language background. The facilitative factors in this regard include the use of appropriate materials and interactive teaching methodology with the help of a highly qualified instructor. The better rise in the *Tamhidi/A-1* class would also show that beginner students are more receptive to learning methods clearly brought out for the requirements. Therefore, these results answer Question 1 directly by showing that the ICP significantly enhances all four skills at higher levels among beginner-level students, *Tamhidi/A-1*.

The pattern strongly correlates with Skill Acquisition Theory (Jackson, 2024), holding that declarative knowledge (vocabulary knowledge, grammar rules) is gained most quickly during the early phases of learning. The 'all-in-one' strategy of ICP's integrated model combines native speakers, bilingual instruction, and standardized learning materials to provide an ideal learning environment for the 'knowledge uptake' phase. This study confirms the pattern in learning Arabic for non-majors in Eastern Indonesia, a vastly underrepresented demographic in the literature on second language acquisition (Alzahrani, 2024).

The results were consistent with Mettewie et al. (2024), who reported that CLIL-based programs promoted stronger initial gains and higher motivation among learners with lower baseline proficiency. This correlation shows that the ICP effectively uses accelerated learning in the early stages among novice students through intensive and structured instruction.

5.2 The increase from midterm to post-test

The increase from midterm to post-test shows lower results, where 14.64 percent for *Tamhidi/A-1* and 16.60 percent for *Mubtadi'/A-2*. This deceleration is not a failure as it is a hallmark of authentic language development, as predicted by DST. As Sh. Zhang, H. Zhang and C. Zhang (2022) and Xu and Yang (2024) determine in longitudinal EFL writing studies, language acquisition is naturally non-linear, where rapid early gains are often followed by plateaus as learners consolidate knowledge and face more complex challenges. The data showed an identical pattern, confirming that the ICP longitudinal design successfully suggested undetectable developmental trajectories by snapshot evaluations. This developmental pattern responds to Question 2 by showing that language growth is fastest at the beginning and then stabilizes as learners consolidate the knowledge across successive evaluation stages.

Although the amount of increase slowed from midterm to post-test, there was no regression in either group, showing that learning progress was sustained. This supports the view that the integrated and structured ICP instructional model provides sufficient support to sustain learning progress even as cognitive demands increase. In connection with CLIL publication, the structured programs help

stabilize performance and prevent skill decline during the consolidation phase, specifically with lower-proficiency learners (Mettewie et al. 2024).

5.3 The increase from pre-test to post-test

The study shows that the ICP improves Arabic language skills in a qualitative manner, with learners transforming from hesitant learners to confident users, as evidenced by the total increase of 44.42 percent at the *Tamhidi*/A-1 and 34.76 percent at the *Mubtadi*'/A-2. However, this phenomenon can be explained more comprehensively through modern perspectives that focus on second language acquisition. Regarding the ICP, continuous exposure to native-speaker input—through conversation sessions and listening materials used in CBT—together with controlled output activities such as speaking tests and writing exercises, increased learners' willingness to communicate (WTC). This trend is evident in increased kalam performance in both levels (Fathi et al. 2024; Lee and Liu 2024; Lambert, 2025).

The results address the well-documented gap in the context of learning Arabic as a second language (L2), where learners often show theoretical knowledge but lack oral communication skills. In correlation with the latest evidence, the results show that targeted instructional design and multimedia-supported interactive practices can significantly improve learners' speaking performance and engagement (Almelhes 2024).

In addition, the results challenge the assumption that ICP models developed in Java or for English can be directly applied to Arabic instruction in Eastern Indonesia. Although Wahyuni et al. (2023) described ICP as “internationally applicable,” this study showed that success depended on contextual adaptation, particularly in addressing persistent weaknesses in *kitabah* and *istima'*, requiring targeted and not generic interventions.

5.4 Contribution to theory and practice: Filling the gaps

Collectively, these results make three key contributions to the broader literature on language learning and ICP effectiveness:

- (1) Empirical Validation of Longitudinal Frameworks in an Understudied Context. Although longitudinal designs are theoretically preferred (Lei et al. 2022), a limited number of studies have applied the methods in the Arabic ICP context, specifically in Eastern Indonesia. This study shows that the designs are not only feasible but also essential, indicating non-linear patterns (plateaus, skill-specific trajectories) invisible to cross-sectional evaluations.
- (2) Contextual Specificity of ICP Effectiveness. Contrary to the assumption that “one size fits all,” the data show that ICP effectiveness varies depending on skill and proficiency level. *Kitabah* remains a universal challenge consistent with the global SLA literature (Almelhes 2024), while *istima'* evolves as a major obstacle in this context, likely due to limited exposure to authentic Arabic input outside the classroom. The pattern

outlines the need for context-sensitive curriculum adjustments that emphasize high-quality listening input and targeted writing support, particularly in PTKINs in Eastern Indonesia. Integrating the localized adaptations into policy frameworks would enhance the contextual responsiveness of ICP implementation (Silaturrahmi et al. 2023).

- (3) Pedagogical Levers for Scaling Impact. These results prove that including native speakers, using CBT for objective assessment, and adopting “all-in-one” skill integration Verbeke and Simon (2023), are effective, but not sufficient when done alone. To sustain progress beyond the plateau, programs must include (a) Differentiated instruction by proficiency level (*Tamhidi*/A-1 vs *Mubtadi*'/A-2), (b) targeted writing workshops with genre-based feedback, (c) listening labs with accent-diverse audio materials, and (d) confidence-building speaking circuits to reduce anxiety (Peungcharoenkun and Waluyo 2023).

5.5 Limitations and methodological transparency

In spite of its strengths, this study has certain limitations that are acknowledged to provide academic integrity. This includes (1) Sample Size and Scope: The results may not apply to other regions or universities with only 50 participants from one institution. (2) Duration: The 10-day program was relatively short, as the long-term retention of the benefits gained has not been measured. (3) Instrument Limitations: Although CBT reduces rater bias, oral assessments still rely on human raters despite having high inter-rater reliability ($\kappa = 0.89$), which cannot be eliminated. (4) Lack of Control Group: The analysis cannot fully separate the effects of the program from other variables without a non-ICP control group (e.g., student motivation, self-study), and (5) Contextual Specificity: These results are closely related to the Gorontalo's socio-linguistic context replication in other eastern Indonesian provinces (Papua, Kalimantan) may yield different results. These limitations do not invalidate the results but contextualize the results as a necessary step for responsible scholarship.

5.6 Directions for future study

Further analysis is needed to perform the following. (1) Extend the longitudinal examination to 6–12 months after the program to determine skill retention. (2) Compare the effectiveness of ICP in different linguistic regions of Indonesia for local policy. (3) Associate the qualitative insights with proficiency data to explore affective factors such as motivation and anxiety. (4) Conducting trials of innovative tools such as AI-based writing support, adaptive grammar feedback, and listening exercises with diverse accents to strengthen writing and listening skills, and (5) using experimental designs with non-ICP control groups to separate program effects from confounding influences. These efforts will contribute significantly to the development of better theory and practice in ICP implementation.

6. Conclusion

In conclusion, this study clearly showed that the ICP effectively enhanced Arabic language proficiency in all four core skills, addressing RQ1. At the pre-test stage, the *Tamhidi/A-1* class was at an average of 54.39 percent (deficient), while *Mubtadi'/A-2* was 61.45 percent (average). During the midterm testing, both groups showed tremendous improvement. Although *Tamhidi/A-1* attained an average of 68.5 percent, *Mubtadi'/A-2* reached an average of 71 percent. These results verified the overall effectiveness of the ICP in enhancing students' language competence.

In relation to RQ2, the gradual gain from the pre-test to the midterm and finally to the post-test reflected a clear developmental trajectory in learners' language abilities. Students in both levels experienced the fastest improvement during the initial phase of learning, followed by steady consolidation toward the end of the program. This included a pattern that was natural in second-language acquisition.

The results equally shed light on RQ3, with respect to which skills have developed the most and remained challenging. The reading and speaking skills have shown the most progress in A-1 and A-2 classes, but *istima'* and *kitabah* were lower in performance with 66 percent in *Tamhidi/A-1* and were identified as a challenge in *Mubtadi'/A-2*.

To answer RQ4, several facilitating factors evolved initial differences in proficiency level at the start, exposure to authentic input, and the intensity of formal instruction in a structured manner within the ICP. It was further influenced by the fact that students had qualified instructors, interactive learning materials, and attractive class environments motivated participants toward better progress.

ICP determined notable effectiveness in improving the listening, speaking, reading, and writing skills of Arabic students. However, constant weaknesses in the listening and writing skills necessitate targeted programs. To handle these failures, educational interventions that use advanced learning methods, including technology-assisted learning, more practice-oriented activities, and native speakers, may bring about a change.

Acknowledgment

This study is funded by the Ministry of Religious Affairs of the Republic of Indonesia through the Research Scientific Publication, and Community Service Programme (LITAPDIMAS) as outlined in the Decree of the Chancellor of IAIN Sultan Amai Gorontalo No. IX: 67. a Year 2024. The authors would like to thank the Minister of Religious Affairs and the Rector for the support through policy and funding. The authors would also like to thank the Institute for Research and Community Service and the Head of the Language Development Centre of IAIN Sultan Amai Gorontalo for the invaluable support and facilitation, specifically in conducting the Computer-based ICP Test at the Language Lab.

Ibnu Rawandhy N. Hula (Associate Professor)- Corresponding Author
Arabic Language Education, Faculty of Education and Teacher Training, IAIN
Sultan Amai Gorontalo, Indonesia.
ORCID Number: 0000-0003-2569-3571
Email: ibnurawandi@iaingorontalo.ac.id

Ana Mariana (Lecturer)
English Literature, Faculty of Social Sciences,
Universitas Muhammadiyah Gorontalo, Indonesia
ORCID Number: 0000-0003-1886-7967
Email: anamariana@umgo.ac.id

Alimudin Rivai (Lecturer)
Arabic Language Education, Faculty of Education and Teacher Training,
IAIN Manado, Indonesia
ORCID Number: 0000-0001-6714-6301
Email: alimudin.rifai@iain-manado.ac.id

Nicky B. Lasanggo
Student, Arabic Language Education, Faculty of Education and Teacher Training
IAIN Sultan Amai Gorontalo, Indonesia
ORCID Number: 0009-0008-6396-8066
Email: nicky@iaingorontalo.ac.id

References

- Abo-Khalil, Ahmed G.** (2024). 'Integrating sustainability into higher education challenges and opportunities for universities worldwide'. *Heliyon*, 10(9): 1-13. <https://doi.org/10.1016/j.heliyon.2024.e29946>
- Ahkam, M. Ehiya, M.J Fathima Sujani, Mohamed Rishad and Singanayagam Umashankar.** (2025). 'Difficulties encountered in English language speaking skills among second language learners: A study based on school leavers at IPHS campus Akkaraipattu'. *Quality: Journal of Education, Arabic and Islamic Studies*, 2(4): 326-357. <https://doi.org/10.58355/qwt.v2i4.88>
- Alazemi, Abdullah.** (2020). 'The effect of the policy of English as a medium of instruction on Arabic in Kuwait'. In Salah Troudi (ed.), *Critical Issues in Teaching English and Language Education: International Research Perspectives*, (41–66). London: Palgrave Macmillan.
- Al Busaidi, Fatima.** (2015). 'Arabic in foreign language programmes: Difficulties and challenges'. *Journal of Educational and Psychological Studies*, 9(4): 701-717. <https://doi.org/10.24200/jeps.vol9iss4pp701-717>
https://doi.org/10.1007/978-3-030-53297-0_3
- Almelhes, Sultan.** (2024). 'Enhancing Arabic language acquisition: Effective strategies for addressing non-native learners' challenges'. *Education Sciences*, 14(10): 1-17. <https://doi.org/10.3390/educsci14101116>
- Alzahrani, Alaa.** (2024). 'LexArabic: A receptive vocabulary size test to estimate Arabic proficiency'. *Behavior Research Methods*, 56(6): 5529–5556. <https://doi.org/10.3758/s13428-023-02286-z>
- Andrian, Riko and Widiya Yul.** (2023). 'Arabic teaching efficacy model (ATEM): A language teaching model design'. *International Journal of Arabic-English Studies (IJAES)*, 23(2): 369–384. <https://doi.org/10.33806/ijaes.v23i2.468>
- Astuti, Missi T., Shalawati Shalawati and Sitti Hadijah.** (2022). 'Learning English as a foreign language in Indonesia: Senior high school students' motivation and challenges'. *AL-ISHLAH: Jurnal Pendidikan*, 14(4): 7001–7016. <https://doi.org/10.35445/alishlah.v14i4.2870>
- Banihashem, Sayyed K., Omid Noroozi, Stan V. Ginkel, Leah Macfadyen and Harm J. A. Biemans.** (2022). 'A systematic review of the role of learning analytics in enhancing feedback practices in higher education'. *Educational Research Review*, 37(11): 1-21. <https://doi.org/10.1016/j.edurev.2022.100489>
- Bani Younes, Mohammad, Alaa M. Smadi, Ali F. Abuseileek and Ghaleb Rabab'ah.** (2023). 'The influence of gender on EFL learners' discourse

functions in a CMC environment'. *International Journal of Arabic-English Studies (IJAES)*, 23(2): 195–210.

<https://doi.org/10.33806/ijaes.v23i2.460>

- Chen, Hua.** (2024). 'Innovative approaches in English language teaching: Integrating communicative methods and technology for enhanced proficiency'. *Communications in Humanities Research*, 32(1), 214–220. <https://doi.org/10.54254/2753-7064/32/20240075>
- Cheung, Gordon W., Helena D. Cooper-Thomas, Rebecca S. Lau and Linda C. Wang.** (2024). 'Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations'. *Asia Pacific Journal of Management*, 41(2): 745–783. <https://doi.org/10.1007/s10490-023-09871-y>
- Crehan, Kevin D., Rogher A. Koehler and Malcolm J. Slakter.** (2023). 'Longitudinal studies of test-wiseness'. *Journal of Educational Measurement*, 11 (3): 145–150. <https://doi.org/10.1111/j.1745-3984.1974.tb00992.x>
- Famularsih, Sari, Agus Nuryatin, Eko Handoyo and Sri W. Fitriati.** (2022). 'Curriculum development management of international class program (ICP) in islamic higher education in Indonesia'. *Eurasian Journal of Educational Research*, 100: 158–173. <https://doi.org/10.14689/ejer.2022.100.011>
- Fathi, Jalil, Masoud Rahimi and Ali Derakhshan.** (2024). 'Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions'. *System*, 121: 1–20. <https://doi.org/10.1016/j.system.2024.103254>
- Fkih, Fethi.** (2022). 'Similarity measures for collaborative filtering-based recommender systems: Review and experimental comparison'. *Journal of King Saud University - Computer and Information Sciences*, 34(9): 7645–7669. <https://doi.org/10.1016/j.jksuci.2021.09.014>
- Foster, Matthew E., Lisa M. López, Shaunacy Sutter, Karen Nylund-Gibson and Dina A.N. Arch.** (2025). 'Latine dual language learners' bilingual development in math and cognition: A longitudinal latent profile analysis'. *Cognitive Development*, 75: 101–110. <https://doi.org/10.1016/j.cogdev.2025.101584>
- Fuss, Ian G. and Daniel J. Navarro.** (2023). 'Open parallel cooperative and competitive decision processes: A potential provenance for quantum probability decision models'. *Topics in Cognitive Science*, 5(4): 818–843. <https://doi.org/10.1111/tops.12045>
- Guirado-Moreno, Jorge-Luis, Victor A. Sanchez-Azanza, Àngels Esteller-Cano, Eva Aguilar-Mediavilla and Daniel Adrover-Roig.** (2025). 'A longitudinal exploration of inhibitory control in children with developmental language disorder'. *Journal of Communication Disorders*,

117: 106-117. <https://doi.org/10.1016/j.jcomdis.2025.106561>

- Herpen, Sanne V., Femke Hilverda and Manja Vollmann.** (2024). ‘A longitudinal study on the impact of student-teacher and student-peer relationships on academic performance: The mediating effects of study effort and engagement’. *European Journal of Higher Education*, 1(20): 599-618. <https://doi.org/10.1080/21568235.2024.2414760>
- Hilmi, Danial and M. Abdul Hamid.** (2025). ‘Students language sensory patterns in online Arabic learning’. *International Journal of Arabic-English Studies (IJAES)*, 25(2): 551–566. <https://doi.org/10.33806/ijaes.v25i2.869>
- Hussein, Alaa A.** (2024). ‘The role of artificial intelligence applications in developing the quality of teaching: A case study in Iraq’. *International Conference on Computing and Machine Intelligence, ICMI 2024 - Proceedings*. <https://doi.org/10.1109/ICMI60790.2024.10586032>
- Irham, Irham.** (2023). ‘International outlook and mediocrity of higher education in Asia’. *Gragoatá*, 28(60): 1-7. <https://doi.org/10.22409/gragoata.v28i60.55727>
- Izzi, Muhamad N. A., Ismatul Maula Ramadhani, Rifdah Salsabila and Ikhwan Saifuddin.** (2022). ‘Implementasi international class program sebagai upaya terwujudnya world class university di Universitas Negeri Surabaya’. *Formosa Journal of Multidisciplinary Research (FJMR)*, 1(3): 543–558. <https://doi.org/10.55927/fjmr.v1i3.734>
- Jackson, Danial O.** (2024). ‘The longitudinal development of argumentative writing in an English for academic purposes course in Japan’. *System*, 126: 1-38. <https://doi.org/10.1016/j.system.2024.103482>
- Ke, Hu, Diao Junfeng and Li Xiaojing.** (2022). ‘International students’ university choice to study abroad in higher education and influencing factors’. *Frontiers in Psychology*, 13(10): 1–10. <https://doi.org/10.3389/fpsyg.2022.1036569>
- Kern, Richard.** (2006). ‘Perspectives on technology in learning and teaching languages. *TESOL Quarterly*, 40(1): 183–210. <https://doi.org/10.2307/40264516>
- Kos, Tomas.** (2023). ‘Exploring peer support among young learners during regular EFL classroom lessons’. *International Journal of Applied Linguistics*, 33(2): 169–189. <https://doi.org/10.1111/ijal.12456>
- Lambert, Craig.** (2025). ‘Neural correlates of task-specific willingness to communicate: Expanding the research agenda’. *TESOL Quarterly*, 1(10): 1–11. <https://doi.org/10.1002/tesq.3395>
- Lee, Ju S. and Linlin Liu.** (2024). ‘Dynamicity of EFL learners’ willingness to communicate in an online class’. *Journal of Multilingual and Multicultural Development*, 45(6): 2078–2096. <https://doi.org/10.1080/01434632.2022.2039677>

- Lei, Lei, Ju Wen and Xiaohu Yang.** (2022). 'A large-scale longitudinal study of syntactic complexity development in EFL writing: A mixed-effects model approach'. *Journal of Second Language Writing*, 59: 1-27. <https://doi.org/10.1016/j.jslw.2022.100962>
- Lorenzo, Francisco, Adrián Granados and Inmaculada Ávila.** (2019). 'The development of cognitive academic language proficiency in multilingual education: Evidence of a longitudinal study on the language of history'. *Journal of English for Academic Purposes*, 41: 1-10. <https://doi.org/10.1016/j.jeap.2019.06.010>
- Lu, Hsin-Hui, Wei-Chun Che, Yu-Ju Lin and Jao-Shwann Liang.** (2022). 'Association of sibling presence with language development before early school age among children with developmental delays: A longitudinal study'. *Journal of the Formosan Medical Association*, 121(6): 1044–1052. <https://doi.org/10.1016/j.jfma.2021.07.026>
- Madadzadeh, Farzan and Sajjad Bahariniya.** (2023). 'Tutorial on how to calculate the content validity of scales in medical research'. *Perioperative Care and Operating Room Management*, 31: 10-25.. <https://doi.org/10.1016/j.pcorm.2023.100315>
- Mettewie, Laurence, Luk V. Mensel, Audrey De Smet and Benoît Galand.** (2024). 'A longitudinal test of the impact of CLIL on language emotions and learning motivation'. *System*, 123: 1-12. <https://doi.org/10.1016/j.system.2024.103355>
- Mudzakkir, Mohammad F., Badri Munir Sukoco and Patdono Suwignjo.** (2022). 'World-class universities: Past and future'. *International Journal of Educational Management*, 36(3): 277–295. <https://doi.org/10.1108/IJEM-07-2021-0290>
- Nnadozie, Chika F. and Nandipha Ngoni.** (2023). 'A method of recovering the very low concentration of pathogens in river water by combining centrifugation and membrane filtration'. *MethodsX*, 11: 1-4. <https://doi.org/10.1016/j.mex.2023.102291>
- Novawan, Adriadi, Siti Aisyiyah, Alfi Hidayatu Miqawati and Fitri Wijayanti.** (2020). 'Exploring the teachers' perspective on morality in an English as a foreign language pedagogy'. *Journal of ELT Research: The Academic Journal of Studies in English Language Teaching and Learning*, 5(1): 80–93. <https://doi.org/10.22236/JER>
- Nufus, Hayati, Erlina Alkoderi, Koderi Koderi, Mahendra U. C Ramadhan and Nopiyanti Nopiyanti.** (2022). 'Development of *Tarkib* teaching materials based on motion graphic in islamic junior high school'. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 14(1): 40–57. <https://doi.org/10.24042/albayan.v14i1.7145>
- Pallant, Julie.** (2020). *SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using IBM SPSS*. London: Routledge.

<https://doi.org/10.4324/9781003117452>

- Peungcharoenkun, Tipaya and Budi Waluyo.** (2023). 'Implementing process-genre approach, feedback, and technology in L2 writing in higher education'. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(1): 1-23. <https://doi.org/10.1186/s40862-023-00211-7>
- Pimada, Luluk H., Nurul Aini, Zahrotul I. Marroh and Wildana Wargadinata.** (2023). 'Teaching Arabic language (extra lessons) to the international class at the state Islamic University of Indonesia'. *Al-Ta'rib : Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya*, 11(1): 65–80. <https://doi.org/10.23971/altarib.v11i1.6167>
- Purwanto, M. Bambang, Rudi Hartono and Sri Wahyuni.** (2023). 'Essential skills challenges for the 21st century graduates: Creating a generation of high-level competence in the industrial revolution 4.0 era'. *Asian Journal of Applied Education (AJAE)*, 2(3): 279–292. <https://doi.org/10.55927/ajae.v2i3.3972>
- Rahman, Aulia and Ria R. Supriani.** (2023). 'Al-tajârîb wa al-tahâdiyyât 'inda muta'allimi al-lughah al-'Arabîyyah fî âdâi ikhtîbar al-lughah al-Mî'yârîy (TOAFL)'. *Jurnal Alfazuna : Jurnal Pembelajaran Bahasa Arab Dan Kebahasaaraban*, 7(2): 120–133. <https://doi.org/10.15642/alfazuna.v7i2.2721>
- Ramadhani, Gina, Imelda Verista Nidi and Lailatu Zahra.** (2024). 'The role of communicative language teaching (CLT) in teaching English to Indonesian elementary school students within the kurikulum merdeka'. *STAIRS: English Language Education Journal*, 5(2): 143-151. <https://doi.org/10.21009/stairs.5.2.7>
- Renaud, Florence, Karine Jauvin and Marie-Julie Bêliveau.** (2025). 'Longitudinal perspective on nonverbal intelligence development in young children with developmental language disorder'. *Research in Developmental Disabilities*, 160(3): 1-14. <https://doi.org/10.1016/j.ridd.2025.104963>
- Robat, Esmaeel S., Hossein Khodabakhshzadeh, Hamid Ashraf and Majid E. Shirvan.** (2021). 'The effectiveness of English language teaching methods: A meta-analysis in pursuit for the best method'. *Journal of Teaching Language Skills (JTLS)*, 40(1): 187–220. <https://doi.org/10.22099/jtls.2021.39199.2920>
- Silaturrehmi, Putri, Uswatun Hasanah and Muhammad Abdul Aziz.** (2023). 'Implementation of the islamization curriculum in Islamic religious higher education (PTKI): Development in Indonesia as an effort to integrate science in the perspective of maqasid al-shari'ah'. *AL-WIJDÂN Journal of Islamic Education Studies*, 8(3): 343–358. <https://doi.org/10.58788/alwijdn.v8i3.2748>
- Sourani, Maha, Faisal Abdulfattah and Inaad Sayer.** (2023). 'Exploring the

challenges of the communicative language teaching approach in the EFL context of Arab learners'. *Teaching English Language*, 17(2): 361–389. <https://doi.org/10.22132/tel.2024.427424.1535>

Sun, Qian, Abdulla Alhowail and Nadin Beckmann. (2025). 'Is foreign language anxiety malleable? A systematic review of longitudinal observation studies'. *System*, 133: 1-20.

<https://doi.org/10.1016/j.system.2025.103782>

Verbeke, Gil and Simon Ellen. (2023). 'Listening to accents: Comprehensibility, accentedness, and intelligibility of native and non-native English speech'. *Lingua*, 292: 1-25. <https://doi.org/10.1016/j.lingua.2023.103572>

Wahyuni, Sri, Mahyudin Ritonga and Windy Afrianti. (2023). 'Systematic review of learning methods for teaching Arabic listening and speaking skills'. *Al-Hayat: Journal of Islamic Education*, 7(1): 1-12.

<https://doi.org/10.35723/ajie.v7i1.321>

Winarti. (2017). 'Classroom interaction: Teacher and student talk in international class program (ICP)'. *KnE Social Sciences*, 1(3): 220-226.

<https://doi.org/10.18502/kss.v1i3.742>

Xu, Jinfen and Baoqi Yang. (2024). 'Dynamic development in L2 written complexity of chinese EFL learners'. *Chinese Journal of Applied Linguistics*, 47(1): 101–120. <https://doi.org/10.1515/CJAL-2024-0107>

Yaseen, Mustafa, Bente Maegaard, Kholid Choukri and Steven Krauwer. (2009). 'Towards networking efforts to build an Arabic cooperation roadmap'. *First International Conference on Networked Digital Technologies*, 441–446. <https://doi.org/10.1109/NDT.2009.5272131>

Zakariya, Yusuf F. (2022). 'Cronbach's alpha in mathematics education research: Its appropriateness, overuse, and alternatives in estimating scale reliability'. *Frontiers in Psychology*, 13(12): 1–6.

<https://doi.org/10.3389/fpsyg.2022.1074430>

Zhang, Shuang, Huiping Zhang and Cun Zhang. (2022). 'A dynamic systems study on complexity, accuracy, and fluency in English writing development by Chinese university students'. *Frontiers in Psychology*, 13(5): 1–18.

<https://doi.org/10.3389/fpsyg.2022.787710>