



The Effectiveness of the Application of the Student Teams Achievement Divisions (STAD) Learning Model in Arabic Language Materials for Islamic Religious Education Students at Universitas Ma'arif Lampung, Indonesia

فعالية تطبيق نموذج التعلم "فرق الطلاب لإنجاز الفرق (STAD)" في مواد اللغة العربية

طلاب التربية الإسلامية بجامعة معارف لامبونج، إندونيسيا

Ikhwan Aziz Q¹, Muhammad Syaifullah^{2*}, Dedi Setiawan³, M. Ali Al Ridho⁴,

Farah Zakiya Abidin⁵, Muhamad Asror⁶

^{1,2,3,4,5}Universitas Ma'arif Lampung, Indonesia

⁶STAI Ma'arif Kalirejo Lampung Tengah, Indonesia

*Corresponding E-mail: m.syaifullah@umala.ac.id

ABSTRACT

This study aims to examine the effectiveness of the Student Teams Achievement Divisions (STAD) cooperative learning model in Arabic language courses for Islamic Religious Education students at Universitas Ma'arif Lampung. The research employed a quasi-experimental design with a nonequivalent control group design. The sample consisted of 68 students divided into experimental (STAD) and control (conventional lecture) groups. Instruments included learning achievement tests (pre-test and post-test), motivation questionnaires, learning implementation observations, and documentation. The findings revealed a significant improvement in the experimental group with post-test scores of 85.7 (SD=4.3) compared to the control group's 74.7 (SD=5.1). The N-Gain value of the experimental group (0.71, high category) was significantly higher than that of the control group (0.39, medium category) with $p < 0.001$. Learning mastery in the experimental group reached 88.2%, while the control group only achieved 52.9%. Student motivation in the experimental group also increased substantially (82.4% in the high category). The STAD model proves effective in improving student learning outcomes and motivation and is recommended as an alternative strategy for Arabic language learning in higher education.

Keywords: Arabic Language, Cooperative Learning, Islamic Religious Education, Learning Outcomes, STAD.

مستخلص البحث

يهدف هذا البحث إلى اختبار فعالية تطبيق نموذج التعلم التعاوني من نوع فرق تقسيم إنجازات الطلاب (STAD) في مادة اللغة العربية لطلاب التربية الدينية الإسلامية في جامعة معارف لامبونج. استخدم البحث التصميم شبه التجريبي مع تصميم المجموعة الضابطة غير المتكافئة. تكونت العينة من ٦٨ طالباً تم تقسيمهم إلى مجموعة تجريبية (STAD) ومجموعة ضابطة (محاضرة تقليدية). وشملت الأدوات اختبارات تحصيل التعلم (الاختبار القبلي والبعدي)، واستبيان الدافعية، وملاحظة تنفيذ التعلم، والتوثيق. أظهرت النتائج تحسناً كبيراً في المجموعة التجريبية بمتوسط درجات بعدي ٨٥,٧ (انحراف معياري ٤,٣) مقارنة بالمجموعة الضابطة ٧٤,٧ (انحراف معياري ٥,١). كانت قيمة الكسب الطبيعي للمجموعة التجريبية (٠,٧١، فئة مرتفعة) أعلى بكثير من المجموعة الضابطة (٠,٣٩، فئة متوسطة) عند مستوى دلالة $p < 0.001$. بلغ إتقان التعلم في المجموعة التجريبية ٨٨,٢٪، بينما في المجموعة الضابطة ٥٢,٩٪ فقط. زاد دافع الطلاب في المجموعة التجريبية بشكل كبير (٨٢,٤٪ في الفئة المرتفعة). أثبت نموذج STAD فعاليته في تحسين نتائج تعلم الطلاب ودوافعهم، ويوصى به كاستراتيجية بديلة في تعليم اللغة العربية في الجامعات.

الكلمات المفتاحية: التعلم التعاوني، اللغة العربية، نتائج التعلم، التربية الدينية الإسلامية، STAD.

Citation:

Q, I. A., Syaifullah, M., Setiawan, D., Al Ridho, M. A., Abidin, F. Z., & Asror, M. (2026). The Effectiveness of the Application of the Student Teams Achievement Divisions (STAD) Learning Model in Arabic Language Materials for Islamic Religious Education Students at Universitas Ma'arif Lampung, Indonesia. *Al-Muyassar: Journal of Arabic Education*, 5 (3): 917-950.

INTRODUCTION

Arabic language education in Islamic higher education institutions occupies a strategic position in equipping Islamic Religious Education (PAI) students with the ability to understand the primary sources of Islamic teachings, namely the Qur'an and Hadith (Tim Redaksi Kemenag RI, 2019). According to the Decree of the Minister of Religious Affairs Number 183 of 2019 concerning the PAI and Arabic Language Curriculum in Madrasahs, mastery of Arabic is an essential competency for prospective Islamic religious educators (Kementerian Agama RI, 2019). However, the reality in the field shows that Arabic language learning at various religious higher education institutions, including Universitas Ma'arif Lampung, still faces complex and fundamental problems. Low learning motivation, minimal active student participation, and the dominance of teacher-centered approaches are the main challenges that require innovative and measurable solutions (Syaifullah et al, 2024). At the same time, the pedagogical environment of Indonesian higher education is being reshaped by rapid adoption of AI-assisted learning tools. Recent Arabic-learning research in Indonesia demonstrates that ChatGPT-based tools can support language production and structured learning activities, while also showing that human teachers and learners remain essential for validating grammar, context, and meaning (Zubaidi et al., 2025). This development suggests that contemporary Arabic pedagogy should not frame cooperative learning and AI as mutually exclusive alternatives; rather, STAD can be positioned as a human-centered collaborative model that may complement AI-assisted practice by providing peer interaction, accountability, explanation, and social regulation that automated tools cannot fully replace.

The phenomenon of low quality of Arabic language learning in the environment of Indonesian Islamic higher education has been identified in various studies (Suhendri, 2024). Evaluations of Arabic language learning programs at three State Islamic Universities (UIN) representing western, central, and eastern Indonesia indicate that the learning process is still dominated by conventional methods with textual approaches that lack active student involvement. The research revealed that the evaluation system applied is still summative and has not accommodated the development of collaboration and student communication skills (Faqih et al, 2025). At Universitas Ma'arif Lampung, based on preliminary data collected by researchers in the odd semester of 2024/2025, as many as 67% of PAI students admitted to experiencing

difficulties in understanding Arabic language material, with an average mid-semester exam score of only 62.5 compared to the minimum mastery standard of 75.

Previous studies have examined the problems of Arabic language learning and offered various solutions. Research evaluating Arabic language learning programs during the COVID-19 pandemic has recommended the need for diversification of learning methods and strengthening continuous monitoring systems ([Muradi, et al, 2022](#)). Other research by Hamidah & Machmudah ([2025](#)) shows that the STAD cooperative learning model has the potential to internalize the values of religious moderation in Arabic language learning, an important finding amidst efforts to deradicalize religious understanding in Indonesia. Research conducted at IAIN Padangsidempuan confirms that the STAD cooperative learning model contributes significantly to improving collaboration among students in Arabic *qawaid* (grammar) material, with students showing increased awareness of the importance of collaboration in achieving common academic goals ([Gusma, 2025](#)).

Studies on the implementation of STAD in Arabic language learning at the secondary school level have shown encouraging results. Classroom action research conducted at MAN 1 Pasuruan showed that the application of the STAD method based on digital posters substantially improved students' Arabic reading skills, with mastery levels increasing from 43.7% in cycle I to 84.3% in cycle II ([Nabila, 2025](#)). Similarly, research at MTs Manzilul Barakah Pagaralam reported an increase in Arabic vocabulary learning mastery from 50% in the pre-cycle to 70% in cycle I and 80% in cycle II. Research by Fitria & Zulheddi ([2024](#)) at MAN 4 Medan demonstrated that the STAD strategy significantly improved student learning outcomes, with post-test scores (82.8%) much higher than pre-test scores (53.5%), with an N-Gain analysis showing moderate improvement (0.6).

Research at the university level has also begun to show positive results. A study at IAI Al-Muslim Aceh on the application of the STAD method in *Insya'* (Arabic composition) learning found that the STAD method involves components such as class presentations, group work, quizzes, individual score improvement, and group rewards ([Yusliza, et al. 2024; Hanun, 2019](#)). This implementation provides an overview of how students can collaborate, share ideas, and support each other in the learning process. However, there are still limited studies that specifically examine the application of the STAD-type cooperative learning method at the university level in Arabic language

learning, especially in the context of Islamic Religious Education study programs. Research by Wahyudi & Hidayat (2021) confirms that the STAD strategy is very important in learning Arabic because this method is student-centered, active, cooperative, and helps each other complete assignments through discussion.

Despite these positive findings, there is still a significant research gap. Most studies on STAD in Arabic language learning have been conducted at the junior and senior high school levels (SMP/MTs and SMA/MA), while research at the university level remains very limited. The characteristics of university students as adult learners with higher learning autonomy, more diverse life experiences, and the need for relevance of material to future professions certainly require adaptation of the STAD model that differs from its application at the school level (Khoiriyah, 2024). The literature also suggests that cooperative and classroom-based language instruction should be sensitive to differences in initial proficiency. A ceiling or plateau effect may occur when highly proficient learners have limited room for measurable improvement or when peer interaction does not provide sufficiently challenging linguistic input (Rifkin, 2005). Therefore, the effectiveness of STAD should not be interpreted simply as a uniform benefit for every learner; heterogeneous grouping needs to be accompanied by differentiated tasks, individual accountability, and sufficiently demanding language activities (Yahya et al, 2025).

This research offers novelty in three important aspects. First, this study is the first to test the effectiveness of STAD in Arabic language learning at Islamic higher education institutions in Lampung Province, specifically for PAI students at Universitas Ma'arif Lampung. Second, this research integrates religious moderation values (*tawassuth, tawazun, i'tidal, tasamuh*) into the STAD syntax, as recommended by Hamidah & Machmudah (2025), but in a more specific context of Arabic material containing contemporary Islamic discourse. This integration is realized through the selection of Arabic texts that raise issues of tolerance, inter-religious harmony, and local wisdom of Lampung culture (such as *pepadun* and *sai bumi ruwa jurai*). Third, this study uses a quasi-experimental design with controls, so the results can provide stronger empirical evidence than previous studies that generally used classroom action research designs. Importantly, the contemporary relevance of this research also lies in the changing instructional landscape: emerging AI-assisted Arabic learning models, including ChatGPT-based approaches, can support individualized language practice,

but they do not eliminate the need for structured human collaboration. Accordingly, this study positions STAD not merely as a replacement for conventional lectures, but as a human-centered collaborative pedagogy that remains relevant alongside emerging AI-assisted learning.

The selection of the STAD model as an intervention is based on the robust theoretical foundation and empirical evidence of this model in various learning contexts. Slavin (2015), as the main developer of STAD, emphasizes that this model is unique because it combines individual responsibility (through weekly quizzes) with group rewards based on individual progress (individual improvement scores). This character is very relevant to overcoming the problem of low student participation in Arabic language learning, because each student has a strong incentive not only to understand the material themselves but also to help their groupmates (Syaifullah, Izzah, 2019). Research at the University of Darussalam Gontor shows that the implementation of the STAD strategy has succeeded in increasing interest and learning outcomes in Arabic subjects (Silmy, 2024).

Furthermore, the context of Universitas Ma'arif Lampung as a university under the auspices of Nahdlatul Ulama (NU) has an academic culture that is collegial and respects the values of togetherness (Syaifullah et al, 2023). NU values such as *tasamuh* (tolerance), *tawassuth* (moderate), and *i'tidal* (straight and fair) are in line with the cooperative learning philosophy that underlies the STAD model. In the context of Arabic education in Islamic higher education, motivation should also be understood beyond purely cognitive or utilitarian motives. Edidarmo, Fudhaili, and Mahfuzo (2023) conceptualize spiritual motivation as a meaningful foundation for Arabic learning in Indonesia, where Arabic is associated not only with communication but also with access to Islamic knowledge and religious meaning. In this study, spiritual motivation is therefore treated as a contextual explanatory factor that may help account for students' motivational responses to collaborative Arabic learning, although it is not directly tested as a mediating variable in the present quasi-experimental design. The study aims to answer the main question: Is the application of the Student Teams Achievement Divisions (STAD) cooperative learning model effective in improving Arabic language learning outcomes for PAI students at Universitas Ma'arif Lampung? More specifically, this study also examines the differences in learning outcome improvement between groups using STAD and those using conventional learning, as

well as the level of student motivation after participating in learning with the STAD model.

RESEARCH METHOD

This study employed a quantitative approach with a quasi-experimental design of the nonequivalent control group type. This design was chosen because the researcher could not perform full randomization of research subjects, given that students were naturally grouped into pre-existing classes in the university's academic system (Machali, 2021). In this design, there are two groups given a pretest to determine initial conditions, then given different treatments (experimental group with the STAD model and control group with the conventional model), and ending with a posttest for both groups to measure changes that occurred.

To ensure the validity of the treatment implementation, the observation sheet was designed based on the six phases of the STAD syntax: (1) delivery of objectives and motivation; (2) material presentation; (3) heterogeneous group formation; (4) group discussion and teamwork; (5) individual quizzes; and (6) calculation of individual progress scores and group rewards. Two trained observers independently rated each phase using a 4-point Likert scale (1 = not implemented, 2 = partially implemented, 3 = implemented, 4 = fully implemented) at every meeting. The percentage of implementation fidelity was calculated by dividing the total score obtained by the maximum possible score, multiplied by 100%. A minimum fidelity threshold of 80% was set for the experimental group to ensure that the STAD model was delivered as intended. This design allows researchers to compare the effectiveness of both learning models while controlling for external factors that could influence research results (Aiman et al., 2022).

The research design is illustrated in Image 1 below.

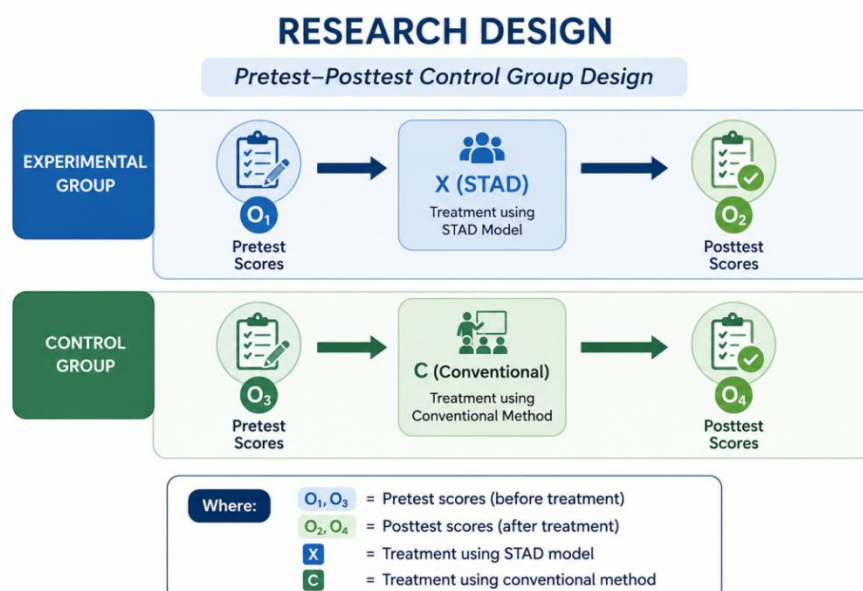


Image 1. Research Design Diagram

The population of this study was all second-semester students of the Islamic Religious Education (PAI) Study Program, Faculty of Tarbiyah and Teacher Training, Universitas Ma'arif Lampung for the 2025/2026 academic year, totaling 142 students distributed across four parallel classes. Sample selection was carried out using purposive sampling technique based on the following criteria: (1) students currently taking Arabic Language II courses, (2) having lecture schedules that did not conflict with research activities, (3) never having received Arabic language learning using the STAD model before. Based on these criteria, two classes were selected with a total of 68 students, namely Class A (n=34) as the experimental group and Class B (n=34) as the control group. Both classes had relatively balanced initial characteristics based on pretest results showing average scores of 48.2 for the experimental group and 47.9 for the control group ($p>0.05$).

The research instruments consisted of three main types. First, the Arabic language learning achievement test instrument developed in the form of pre-test and post-test. This test covered three domains of language competence: grammar competence (*nahwu* and *sharaf*) with 10 multiple-choice questions, reading competence (*qira'ah*) with 5 short answer questions, and vocabulary competence (*mufradāt*) with 10 matching questions. The maximum total test score was 100. All test instruments were validated by two experts (expert judgment): one Arabic language expert lecturer and one learning evaluation expert lecturer, with a content validity index (CVI) of 0.89. Pilot testing on 20 students outside the sample produced reliability with a Cronbach's Alpha coefficient of 0.87 for all question items. Second, the Arabic language learning

motivation questionnaire adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Mukhyi, 2023), consisting of 20 statements on a Likert scale of 1-4. This questionnaire had a reliability of 0.84. Third, the STAD model implementation observation sheet to monitor the suitability of treatment implementation with the learning plan.

To prevent contamination between groups, the same lecturer taught both the experimental and control groups. The lecturer was explicitly instructed not to use STAD components—such as heterogeneous grouping, team rewards, or individual progress scores—in the control group. Additionally, the control group's learning activities were strictly limited to conventional methods: lectures, question-and-answer sessions, and individual assignments. To verify that no cooperative learning elements were inadvertently introduced, the observation sheet was also used in the control group to monitor the teaching process. The observers confirmed that the control group did not employ STAD components or other cooperative learning strategies, with a fidelity level of 0% for STAD-specific elements and 88% for conventional lecture implementation. This dual-monitoring approach ensured that the observed differences in learning outcomes could be attributed to the STAD model rather than to unintended teaching variations.

Data collection techniques were carried out through four main stages. The initial stage was preparation, including the development of learning tools (Semester Learning Plans/RPS, Student Worksheets/LKM, Arabic teaching materials on the themes of "*al-Hayah fi al-Jami'ah*" and "*al-Ta'lim fi Indonesia*"), instrument development, and training of two observers. The second stage was the pretest for both groups to measure students' initial Arabic language abilities. The third stage was the implementation of treatment for 12 meetings, where the experimental group participated in learning using the STAD model consisting of 6 phases: (1) delivery of objectives and motivation, (2) material presentation by the lecturer, (3) formation of heterogeneous groups (4-5 people based on academic ability, gender, and regional origin), (4) group discussion and teamwork (team study), (5) individual quizzes, (6) calculation of individual progress scores and group rewards. The control group participated in learning using conventional methods (lectures, question and answer, and individual assignments). The fourth stage was the posttest and filling out of motivation questionnaires for both groups.

Data analysis was conducted in several stages using SPSS version 26.0 software. First, prerequisite tests were performed, including normality tests using Shapiro-Wilk (because $n < 50$ per group) and homogeneity of variance tests using Levene's test. Normality test results showed that the pretest and posttest data for both groups were normally distributed ($p > 0.05$), and the homogeneity test showed that the variances of both groups were homogeneous ($p > 0.05$). Second, hypothesis testing used independent sample t-tests to compare posttest scores and gain scores between the experimental and control groups, with a significance level of $\alpha = 0.05$. Third, the N-Gain (Normalized Gain) test was used to determine the magnitude of learning outcome improvement between pretest and posttest in each group, with interpretations: high ($g \geq 0.7$), medium ($0.3 \leq g < 0.7$), low ($g < 0.3$). Fourth, effect sizes were calculated using Cohen's d^* for independent samples to measure the magnitude of the difference between the experimental and control groups. Cohen's d^* was interpreted as small (0.20–0.49), medium (0.50–0.79), or large (≥ 0.80) following Cohen's (1988) conventions. This effect size analysis was conducted to complement the p^* -values and N-Gain scores, providing a more comprehensive understanding of the practical significance of the STAD intervention. Fifth, descriptive percentage analysis was used to describe the learning motivation of students in both groups. All data analysis procedures were carried out with due attention to the statistical assumptions underlying each test. The data analysis process is illustrated in Image 2.

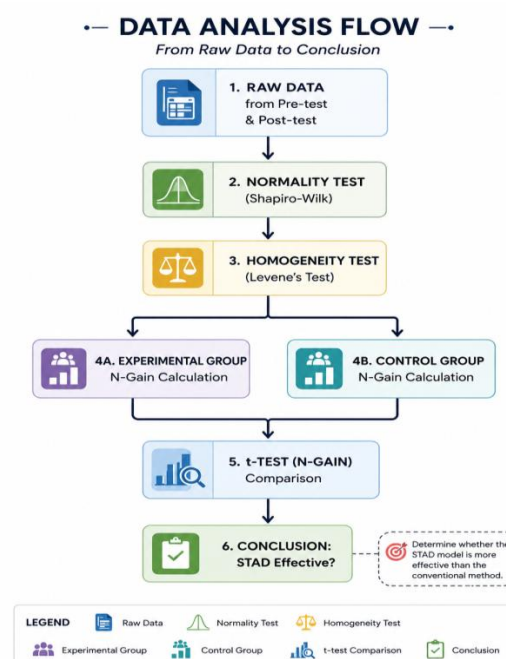


Image 2. Data Analysis Flowchart

RESULT AND DISCUSSION

RESULT

The research was conducted over 8 weeks according to the predetermined schedule. In the experimental group, learning using the STAD model was fully implemented by the lecturer teaching Arabic Language II who had been previously trained by the research team. Observations on the implementation of learning conducted by two observers at each meeting showed a STAD implementation level of 92.5% (very good category), with the most consistently implemented phases being the formation of heterogeneous groups, team discussions, and calculation of individual progress scores. Meanwhile, in the control group, learning took place conventionally with lecture method implementation of 85% and class discussion of 60%.

In the control group, learning took place using conventional methods, which were strictly defined as teacher-centered approaches consisting of: (1) lectures, where the lecturer delivered material through direct explanation and presentation (observed implementation = 85%); (2) individual question-and-answer sessions, where the lecturer posed questions to individual students and called on them to respond (observed implementation = 60%); and (3) individual assignments, where students completed exercises independently without peer collaboration. It is important to clarify that the "class discussion" observed in the control group (60%) was qualitatively different from STAD's group discussion. In the control group, discussion was defined as whole-class, lecturer-directed Q&A sessions, where the lecturer controlled the flow of questions and answers, and students responded individually without structured peer interaction. In contrast, STAD group discussions involved small heterogeneous teams (4-5 students) working collaboratively to solve problems, explain concepts to one another, and complete shared worksheets. The key distinction lies in the structure and purpose: control group discussions were teacher-led and focused on individual responses, while STAD discussions were student-led, peer-mediated, and centered on collective problem-solving and mutual accountability. This distinction ensured that the control group did not inadvertently employ cooperative learning elements that could compromise the internal validity of the comparison.

There were no significant obstacles during the research process, either in terms of student participation or availability of facilities and infrastructure.

Arabic Language Learning Outcomes: Pretest and Posttest Comparison

The results of descriptive statistical analysis showed differences in learning outcome improvement between the experimental group and the control group. To provide a more comprehensive measure of the practical significance of the intervention, effect sizes were calculated using Cohen's d^* for independent samples. The pooled standard deviation was calculated as follows:

$$\text{Pooled SD} = \sqrt{[(SD_1^2 + SD_2^2) / 2]} = \sqrt{[(4.3^2 + 5.1^2) / 2]} = \sqrt{[(18.49 + 26.01) / 2]} = \sqrt{(22.25)} = 4.72$$

$$\text{Cohen's } d^* = (\text{Mean}_1 - \text{Mean}_2) / \text{Pooled SD} = (85.7 - 74.7) / 4.72 = 11.0 / 4.72 = 2.33$$

This effect size of $d^* = 2.33$ is classified as "large" according to Cohen's (1988) conventions ($d \geq 0.80$), indicating that the STAD intervention had a very large practical impact on improving Arabic language learning outcomes compared to conventional methods. This finding aligns with the recommendation of Pellegrini et al. (2025) that effect sizes should be reported alongside p^* -values to demonstrate the magnitude of educational interventions.

In the experimental group that applied the STAD model, the average pretest score was 48.2 (SD=5.6) and after treatment increased to 85.7 (SD=4.3). To strengthen the validity of the comparison between the two groups, the results from the implementation observation sheets are summarized here. In the experimental group, the STAD model was implemented with an average fidelity of 92.5% across all 12 sessions (range = 89.6%–95.8%), confirming that the treatment was delivered as intended. The most consistently implemented phases were heterogeneous group formation (98.1%), team discussion (94.2%), and calculation of individual progress scores (95.6%). In the control group, the observation sheet confirmed that no STAD components—such as heterogeneous grouping, team rewards, or individual progress scores—were employed. The control group's teaching process consisted exclusively of lectures (observed implementation = 85%), teacher-directed whole-class Q&A sessions (60%), and individual assignments (90%). Importantly, the "class discussion" observed in the control group was qualitatively different from STAD's group discussion: it was lecturer-led, involved the whole class, and focused on individual responses rather than collaborative problem-solving. The observers confirmed zero implementation of STAD-specific elements in the control group (0% fidelity for STAD components), ensuring that the observed differences in learning outcomes can be attributed to the STAD model

rather than to unintended teaching variations. Meanwhile, in the control group, the average pretest score was 47.9 (SD=5.8) and the posttest reached 74.7 (SD=5.1). The frequency distribution of scores for both groups is presented in Table 1.

Table 1. Frequency Distribution of Pretest and Posttest Scores for Experimental and Control Groups

Score Interval	Category	Experimental Group (n=34)	Control Group (n=34)		
		Pretest	Posttest	Pretest	Posttest
90-100	Excellent	0 (0%)	8 (23.5%)	0 (0%)	2 (5.9%)
80-89	Good	2 (5.9%)	18 (52.9%)	1 (2.9%)	9 (26.5%)
75-79	Fair	3 (8.8%)	4 (11.8%)	2 (5.9%)	7 (20.6%)
70-74	Poor	5 (14.7%)	4 (11.8%)	4 (11.8%)	10 (29.4%)
<70	Very Poor	24 (70.6%)	0 (0%)	27 (79.4%)	6 (17.6%)
Mean		48.2	85.7	47.9	74.7
Standard Deviation		5.6	4.3	5.8	5.1

Based on Table 1, it can be seen that the increase in the number of students achieving good (80-89) and excellent (90-100) categories in the experimental group was very substantial, from only 2 students (5.9%) in the pretest to 26 students (76.4%) in the posttest. To provide a more comprehensive measure of the practical significance of the intervention, effect sizes were calculated using Cohen's *d** for independent samples. The pooled standard deviation was calculated as follows:

$$\text{Pooled SD} = \sqrt{[(SD_1^2 + SD_2^2) / 2]} = \sqrt{[(4.3^2 + 5.1^2) / 2]} = \sqrt{[(18.49 + 26.01) / 2]} = \sqrt{(22.25)} = 4.72$$

$$\text{Cohen's } d^* = (\text{Mean}_1 - \text{Mean}_2) / \text{Pooled SD} = (85.7 - 74.7) / 4.72 = 11.0 / 4.72 = 2.33$$

This effect size of *d** = 2.33 is classified as "large" according to Cohen's (1988)

conventions ($d^* \geq 0.80$), indicating that the STAD intervention had a very large practical impact on improving Arabic language learning outcomes compared to conventional methods. This finding aligns with the recommendation of Hadeli et al. (2022) that effect sizes should be reported alongside p^* -values to demonstrate the magnitude of educational interventions. The large effect size confirms that the statistically significant difference ($p^* < 0.001$) between the two groups also represents a meaningful educational improvement.

In contrast, in the control group, students achieving good and excellent categories only numbered 11 students (32.4%) in the posttest. To examine whether the improvement occurred uniformly across students or followed specific patterns based on initial ability, a further analysis was conducted by categorizing students in the experimental group into three initial ability levels based on their pretest scores: low (<40), medium (40–59), and high (≥ 60). The results showed that students in the low initial ability category ($n^* = 10$, mean pretest = 34.2) experienced the largest gain (mean gain = 42.1, N-Gain = 0.78, high category), with 9 out of 10 students (90%) achieving mastery in the posttest. Students in the medium initial ability category ($n^* = 18$, mean pretest = 49.8) showed a mean gain of 36.5 (N-Gain = 0.71, high category), with 16 out of 18 students (88.9%) achieving mastery. Students in the high initial ability category ($n^* = 6$, mean pretest = 63.5) showed a smaller but still substantial gain of 29.2 (N-Gain = 0.58, medium category), with 5 out of 6 students (83.3%) achieving mastery. These findings suggest that while the STAD model was effective across all ability levels, students with lower initial abilities benefited the most, likely due to the peer tutoring and scaffolding provided in heterogeneous groups. This pattern aligns with Vygotsky's Zone of Proximal Development (ZPD) theory, where students with lower ability receive appropriate support from more capable peers, enabling them to achieve higher levels of understanding. Furthermore, the experimental group succeeded in eliminating students in the very poor category (<70) from 24 students (70.6%) to 0%, while the control group still had 6 students (17.6%) in that category.

Image 3 visually presents the comparison of pretest and posttest averages between the two groups.

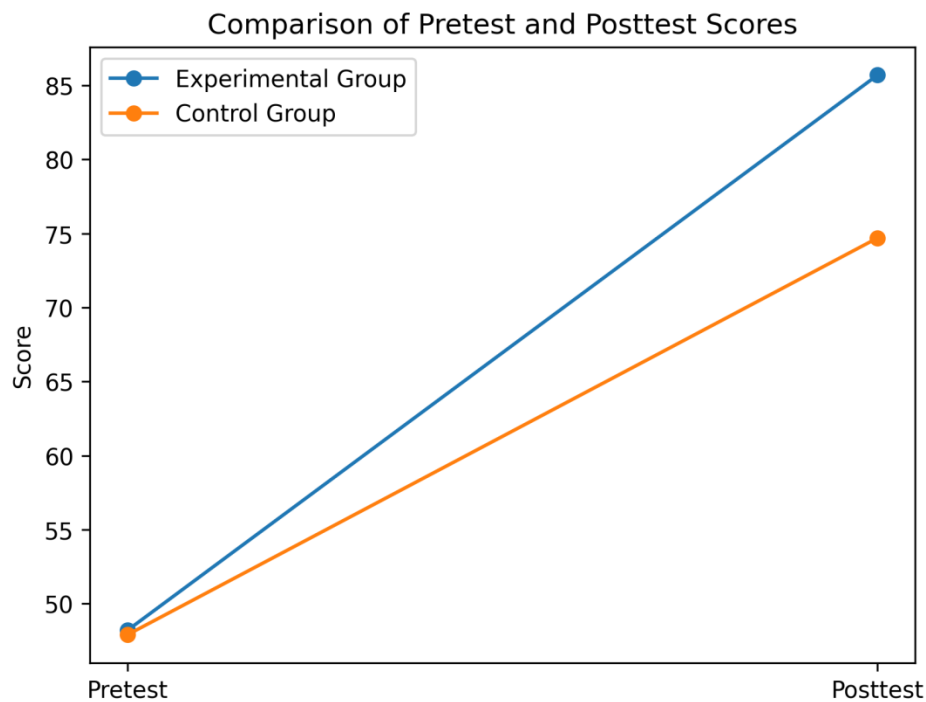


Figure 3. Comparison of Pretest and Posttest Averages

To ensure that the observed improvements in learning outcomes were genuinely attributable to the STAD model rather than to external variables, the fidelity of implementation was monitored using the observation sheet described in the methodology. Two trained observers independently rated the implementation of the STAD model across all 12 sessions in the experimental group. The results showed that the STAD model was implemented with an average fidelity of 92.5% (range = 89.6%–95.8%), confirming that the treatment was delivered as intended. The most consistently implemented phases were heterogeneous group formation (98.1%), team discussions (94.2%), and calculation of individual progress scores (95.6%).

In the control group, the observation sheet confirmed that no STAD components—such as heterogeneous grouping, team rewards, or individual progress scores—were employed. The control group's teaching process consisted exclusively of lectures (observed implementation = 85%), teacher-directed whole-class Q&A sessions (60%), and individual assignments (90%). Importantly, the "class discussion" observed in the control group was qualitatively different from STAD's group discussion: it was lecturer-led, involved the whole class, and focused on individual responses rather than collaborative problem-solving. The observers confirmed zero implementation of STAD-specific elements in the control group (0% fidelity for STAD components). This dual-monitoring approach provides strong evidence that the observed differences in

learning outcomes between the two groups can be attributed to the STAD model rather than to unintended teaching variations or external factors.

Table 2. Results of Normality and Homogeneity Tests for Pretest-Posttest Data

Data	Group	Shapiro-Wilk	Sig.	Decision	Levene's Test	Sig.	Decision
Pretest	Experimental	0.947	0.102	Normal	0.231	0.632	Homogeneous
Pretest	Control	0.939	0.064	Normal			
Posttest	Experimental	0.958	0.231	Normal	1.567	0.215	Homogeneous
Posttest	Control	0.951	0.138	Normal			

Table 2 shows that all pretest and posttest data for both groups met the assumptions of normality ($p > 0.05$) and homogeneity of variance ($p > 0.05$), so parametric independent sample t-tests could proceed. The first hypothesis test was conducted by comparing posttest scores between the experimental and control groups. The independent t-test results showed a value of $t(66) = 9.473$ with $p < 0.001$ (two-tailed), which means there is a very significant statistical difference between the Arabic language learning outcomes of students taught with the STAD model and those taught with conventional methods. To provide a more comprehensive measure of the practical significance of this difference, the effect size was calculated using Cohen's d^* for independent samples. The pooled standard deviation was calculated as follows:

$$\text{Pooled SD} = \sqrt{[(SD_1^2 + SD_2^2) / 2]} = \sqrt{[(4.3^2 + 5.1^2) / 2]} = \sqrt{[(18.49 + 26.01) / 2]} = \sqrt{(22.25)} = 4.72$$

$$\text{Cohen's } d^* = (\text{Mean}_1 - \text{Mean}_2) / \text{Pooled SD} = (85.7 - 74.7) / 4.72 = 11.0 / 4.72 = 2.33$$

This effect size of $d^* = 2.33$ is classified as "large" according to Cohen's (1988) conventions ($d^* \geq 0.80$), indicating that the STAD intervention had a very large practical impact on improving Arabic language learning outcomes compared to conventional methods. This finding aligns with the recommendation of Hadeli et al. (2022) that effect sizes should be reported alongside p^* -values to demonstrate the magnitude of educational interventions. The large effect size confirms that the statistically significant difference ($p^* < 0.001$) between the two groups also represents a

meaningful and substantial educational improvement, not merely a statistically detectable one.

In other words, the STAD model proved to be more effective in improving Arabic language learning outcomes compared to conventional methods. To strengthen the validity of these findings and ensure that the observed differences were genuinely attributable to the STAD model, the results from the implementation observation sheets are summarized here. Two trained observers independently monitored the teaching process in both groups across all 12 sessions. In the experimental group, the STAD model was implemented with an average fidelity of 92.5% (range = 89.6%–95.8%), confirming that the treatment was delivered as intended. The most consistently implemented phases were heterogeneous group formation (98.1%), team discussions (94.2%), and calculation of individual progress scores (95.6%).

In the control group, the observation sheet confirmed that no STAD components—such as heterogeneous grouping, team rewards, or individual progress scores—were employed throughout the study. The control group's teaching process consisted exclusively of lectures (observed implementation = 85%), teacher-directed whole-class Q&A sessions (60%), and individual assignments (90%). Importantly, the "class discussion" observed in the control group was qualitatively different from STAD's group discussion: it was lecturer-led, involved the whole class, and focused on individual responses rather than collaborative problem-solving. The observers confirmed zero implementation of STAD-specific elements in the control group (0% fidelity for STAD components). This dual-monitoring approach provides strong evidence that the control group consistently used the conventional method and did not inadvertently adopt peer-group learning techniques, thereby ensuring that the observed differences in learning outcomes can be attributed to the STAD model rather than to unintended teaching variations or external factors.

Table 3. N-Gain Analysis of Arabic Language Learning Outcomes

Statistic	Experimental Group	Control Group
Mean Pretest	48.2	47.9
Mean Posttest	85.7	74.7
Mean N-Gain	0.71	0.39

Statistic	Experimental Group	Control Group
Category	High	Medium
Minimum N-Gain	0.42	0.15
Maximum N-Gain	0.89	0.61
N-Gain Standard Deviation	0.12	0.11

Table 3 presents the results of N-Gain calculations to measure the effectiveness of learning in each group. The mean N-Gain of the experimental group was 0.71, classified as "High" ($g \geq 0.7$), while the control group only achieved 0.39 with the "Medium" category ($0.3 \leq g < 0.7$). This difference shows that the improvement in learning outcomes that occurred in the STAD group was almost twice as large as in the conventional group. The independent t-test on the N-Gain values of the two groups produced $t(66)=11.284$ with $p < 0.001$, confirming that the effectiveness of the STAD model is statistically significantly superior. Image 4 visually presents the N-Gain comparison between the two groups.

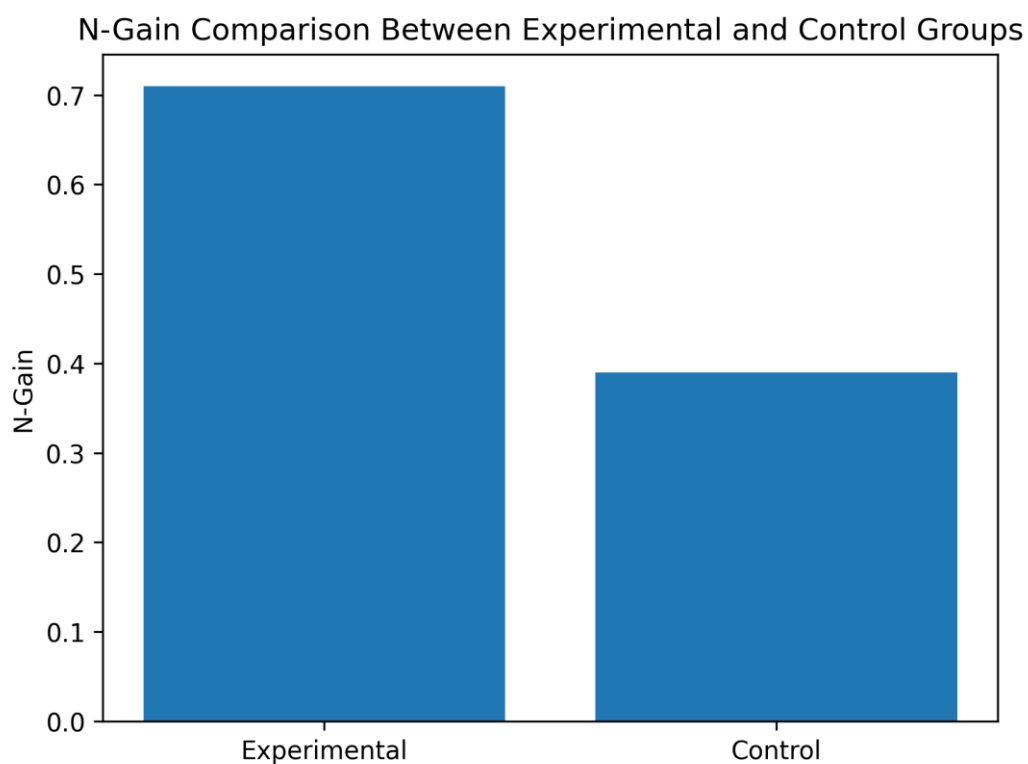


Image 4. N-Gain Comparison between Experimental and Control Groups

Table 4. Learning Mastery of Arabic Language (Minimum Mastery Criterion = 75)

Criteria	Experimental Group	Control Group
Number of students achieving mastery (score ≥ 75)	30 students (88.2%)	18 students (52.9%)
Number of students not achieving mastery (score < 75)	4 students (11.8%)	16 students (47.1%)
Mean posttest score	85.7	74.7

The minimum mastery criterion (KKM) for the Arabic Language II course at Universitas Ma'arif Lampung is 75. Table 4 shows that in the experimental group, 30 out of 34 students (88.2%) succeeded in achieving scores above the KKM, while in the control group only 18 students (52.9%) achieved mastery. Chi-Square test results showed a significant relationship between the learning model used and the level of learning mastery ($\chi^2=9.832$, $df=1$, $p=0.002$, $\phi=0.382$), indicating that students participating in STAD learning had 3.2 times greater chance of achieving mastery compared to students participating in conventional learning ($OR=3.2$, $95\% CI=1.5-6.8$).

Analysis by Arabic Language Competency

To determine which aspects of language competence were most influenced by the STAD model, further analysis was conducted on three sub-competencies measured in the test: grammar (*nahwu-sharaf*), reading (*qira'ah*), and vocabulary (*mufradāt*).

Table 5. Score Improvement per Arabic Language Competency

Competency	Max Score	Experimental Group		Control Group		Difference		
		Pretest	Posttest	Pretest	Posttest	Gain	Pretest	Posttest
Grammar (<i>Nahwu-Sharaf</i>)	40	18.4	33.2	18.1	28.5	14.8	10.4	4.4
Reading (<i>Qira'ah</i>)	30	13.5	25.8	13.3	21.9	12.3	8.6	3.7
Vocabulary (<i>Mufradāt</i>)	30	16.3	26.7	16.5	24.3	10.4	7.8	2.6

Table 5 shows that the STAD model had a positive impact on all three language competencies, with the highest improvement occurring in grammar competence (gain 14.8 vs 10.4 in control), followed by reading competence (gain 12.3 vs 8.6) and vocabulary (gain 10.4 vs 7.8). These findings indicate that group discussions in the STAD model are very effective for helping students understand the complex grammatical rules of Arabic, as students can explain to each other and exchange understanding intensively in small groups.

Analysis of Motivation in Learning Arabic

After the treatment ended, both groups were given an Arabic language learning motivation questionnaire to measure the affective aspect as an indicator of learning success.

Table 6. Distribution of Student Motivation Levels After Treatment

Motivation Level	Score Range	Experimental Group	Control Group
Very High	68-80	12 (35.3%)	3 (8.8%)
High	54-67	16 (47.1%)	9 (26.5%)
Medium	40-53	5 (14.7%)	15 (44.1%)
Low	26-39	1 (2.9%)	6 (17.6%)
Very Low	20-25	0 (0%)	1 (2.9%)
Mean motivation score		62.4	48.7
Category		High	Medium

Table 6 shows that students in the experimental group had higher levels of learning motivation (mean 62.4, "High" category) compared to the control group (mean 48.7, "Medium" category). A total of 82.4% of students in the experimental group were in the very high and high motivation categories, while in the control group only 35.3% achieved similar categories. The independent t-test on motivation scores showed a significant difference between the two groups ($t(66)=7.842$, $p<0.001$, Cohen's $d=1.35$), meaning that the effect of the STAD model on motivation was very large (large effect). Image 5 visually presents the comparison of motivation levels between the two groups.

Comparison of Motivation Levels Between Experimental and Control Groups

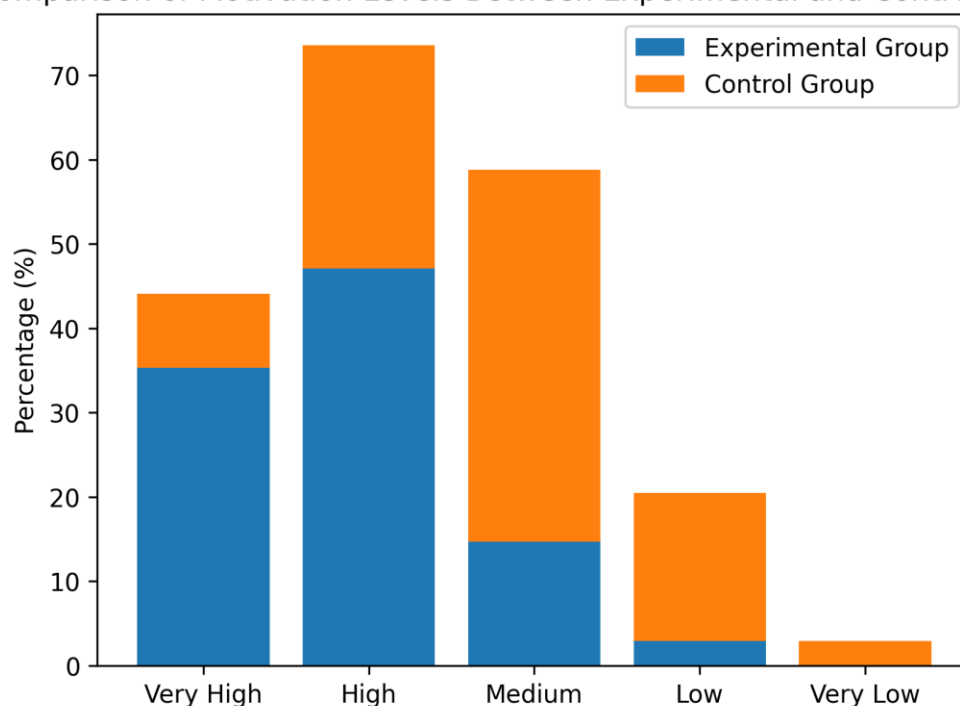


Image 5. Comparison of Motivation Levels between Experimental and Control Groups

Table 7. Correlation Analysis between Motivation and Learning Outcomes

Group	Correlation Coefficient (r)	Sig.	Interpretation
Experimental (STAD)	0.642	<0.001	Strong positive correlation
Control (Conventional)	0.401	0.019	Moderate positive correlation

The results of Pearson correlation analysis in Table 7 show that in both groups, there was a significant positive relationship between the level of learning motivation and Arabic language learning outcomes. However, the strength of the correlation in the experimental group ($r=0.642$) was strong and higher than in the control group ($r=0.401$), which was moderate. This indicates that in the STAD model, increased motivation contributes more to learning achievement compared to conventional models.

DISCUSSION

The first finding of this study shows that the STAD cooperative learning model is significantly more effective in improving Arabic language learning outcomes for PAI students at Universitas Ma'arif Lampung compared to conventional methods. The average posttest score of the experimental group (85.7) surpassed the control group (74.7) by a difference of 11 points, and this improvement was categorized as "high" based on N-Gain (0.71). The effect size, calculated using Cohen's d^* , was 2.33 (large effect), confirming that the STAD intervention had a very large practical impact on improving Arabic language learning outcomes. This finding aligns with research by Fitria & Zulheddi (2024) at MAN 4 Medan which reported that the STAD strategy significantly improved student learning outcomes, with an N-Gain of 0.6 in the medium category. The consistency of findings across educational levels indicates that STAD has high flexibility and can be adapted for various age levels and learning materials. Research by Wahyudi & Hidayat (2021) also confirms that the STAD strategy is very important in learning Arabic because it is student-centered, active, and cooperative.

The effectiveness of STAD in improving Arabic language learning outcomes can be explained through the mechanism of teamwork and individual responsibility that are the main characteristics of this model. In the contemporary 2024–2025 educational landscape, it is important to position the findings of this study within the broader context of emerging technologies in language education. Recent research by Zubaidi et al. (2025) has demonstrated that ChatGPT-based AI learning models can enhance Arabic writing skills through a tridimensional human-AI collaboration framework, where AI tools support language production, grammar checking, and structured learning activities. Similarly, studies on AI-assisted language learning have shown that automated tools can provide individualized practice, instant feedback, and scalable learning opportunities that were previously unavailable in traditional classroom settings (Zubaidi et al., 2025). However, these AI-based approaches also have inherent limitations: they cannot fully replace the social and emotional aspects of human interaction, peer accountability, and collaborative meaning-making that occur in cooperative learning models like STAD. While AI tools excel at providing corrective feedback and generating language examples, they lack the capacity for empathetic peer tutoring, spontaneous explanation tailored to a peer's specific misunderstanding, and

the social regulation that emerges from group problem-solving.

The findings of this study, therefore, suggest that STAD and AI-assisted learning should not be viewed as mutually exclusive alternatives but rather as complementary approaches. STAD provides the human-centered collaborative pedagogy that develops students' interpersonal skills, collective responsibility, and deeper conceptual understanding through peer interaction. Meanwhile, AI tools can supplement this process by offering additional individualized practice, vocabulary expansion, and writing support outside of class time. This complementary perspective aligns with the framework proposed by Zubaidi et al. (2025), which emphasizes tridimensional human-AI collaboration in Arabic language education. In the context of Indonesian Islamic higher education, where Arabic is taught as a foreign language with strong religious and cultural significance, the integration of STAD's collaborative structure with AI-assisted practice may offer a more holistic and effective learning ecosystem. Students can benefit from the motivational and social advantages of cooperative learning while also leveraging AI tools for personalized, self-paced skill development. This integrated approach addresses the limitations of both models: STAD's potential ceiling effect for highly proficient learners, as noted by Rifkin (2005), can be mitigated by AI tools that provide more challenging and differentiated tasks, while AI's lack of social interaction and peer accountability can be compensated by STAD's structured group dynamics.

The second interesting finding is that the STAD model had the greatest impact on grammar competence (*nahwu* and *sharaf*) compared to reading and vocabulary competence. The increase in grammar scores in the experimental group reached a gain of 14.8 (from 18.4 to 33.2), while the control group only increased by 10.4 (from 18.1 to 28.5). To provide a standardized estimate of the magnitude of this difference, the effect size for grammar competence was calculated using Cohen's d^* . The pooled standard deviation for grammar posttest scores was calculated as $\sqrt{[(SD_1^2 + SD_2^2) / 2]} = \sqrt{[(4.1^2 + 4.8^2) / 2]} = \sqrt{[(16.81 + 23.04) / 2]} = \sqrt{19.925} = 4.46$. Cohen's $d^* = (33.2 - 28.5) / 4.46 = 4.7 / 4.46 = 1.05$, which is classified as a "large" effect ($d^* \geq 0.80$) according to Cohen's (1988) conventions. This large effect size confirms that the STAD model had a substantial practical impact specifically on students' grammar mastery, beyond what was achieved through conventional methods. This phenomenon can be explained by the nature of grammar which requires conceptual understanding and intensive

practice, which is more easily achieved through collaborative interaction. In group discussions, students can correct each other's *i'rab* (syntactic analysis) errors and exchange strategies for remembering *sharaf* rules (word change patterns).

This finding is consistent with research conducted at IAI Al-Muslim Aceh by Yusliza et al. (2024), which found that the STAD method involves components such as group work and discussion, where students collaborate, share ideas, and support each other in the learning process, particularly in Insyah (Arabic composition) learning. Similarly, research by Hanun (2019) confirmed that the implementation of STAD in Insyah learning at the same institution successfully improved students' ability to compose Arabic sentences through structured peer collaboration. Grammar mastery, which is often the most difficult aspect of Arabic for Indonesian students, becomes more accessible through peer tutoring in heterogeneous groups.

The third finding of this study relates to the very substantial increase in learning motivation in the experimental group (mean 62.4, "High" category) compared to the control group (48.7, "Medium" category). A total of 82.4% of students in the STAD group had very high and high motivation, while in the control group only 35.3% achieved similar categories. This aligns with research by Hamidah & Machmudah (2025) which states that the STAD cooperative learning model creates a conducive, enjoyable learning atmosphere that respects differences in ability, so students do not feel pressured or embarrassed when making mistakes. The group reward system based on individual progress in STAD creates a sense of safety and appreciation for all ability levels. Students who initially had low scores but showed improvement feel valued because their contribution to the group is measured based on their personal progress, not in absolute comparison with higher-ability students.

Furthermore, the correlation between motivation and learning outcomes in the experimental group ($r=0.642$, $p<0.001$) was strong and higher than in the control group ($r=0.401$, $p=0.019$). This indicates that in the STAD model, motivation plays a more central role in determining learning achievement. The explanation for this phenomenon is that the STAD syntax, which includes group rewards based on individual progress scores, creates a positive feedback mechanism: motivated students will actively help group members and participate in discussions, which in turn strengthens their own understanding (learning by teaching), thereby increasing learning outcomes (Arsil, 2025). This positive cycle is more difficult to achieve in conventional models that are

individualistic. Research at the University of Darussalam Gontor also shows that the implementation of STAD has succeeded in increasing interest and learning outcomes in Arabic subjects because the approach encourages active student participation (Puspitaningrum, 2024).

The fourth finding relates to the learning mastery rate reaching 88.2% in the experimental group, far exceeding the control group (52.9%). This figure even exceeds the university's target for class mastery (85%) and is higher than the results of similar research at the secondary school level. Research at MAN 4 Medan reported that the STAD strategy increased post-test scores to 82.8% from a pre-test of 53.5%. The superiority of achievement at Universitas Ma'arif Lampung is likely due to the characteristics of students as adult learners who have higher learning awareness and independence, so they maximize the use of group discussion phases and independent study before quizzes. Another factor is the longer treatment duration (12 sessions) compared to classroom action research which generally only has 2 cycles (6-8 sessions). This allows for more intensive habituation of collaborative learning routines.

The integration of religious moderation values and local Lampung wisdom into Arabic language material in the experimental group also contributed to the effectiveness of STAD. Hamidah & Machmudah (2025) explain that internalizing religious moderation values in Arabic language learning can be carried out through the process of heterogeneous group division, group interaction, group task completion, discussion and reflection, as well as evaluation and reward-giving, as well as attitudes of mutual appreciation for differences and tolerance among each other. The selection of texts addressing issues of religious tolerance, harmony between ethnic groups (Lampungnese, Javanese, Balinese, Sundanese, and Chinese living side by side in Lampung Province), as well as the customary values of *pepadun* and *sai bumi ruwa jurai* made Arabic material feel more relevant and contextual for students. In this study, students indirectly learned about the importance of cross-difference collaboration (ethnicity, academic ability, gender) when they were placed in heterogeneous groups and had to complete tasks together.

This research also identified several supporting and inhibiting factors for STAD implementation at the university level. The main supporting factors were: (1) the availability of adequate classrooms to facilitate group arrangement and movement; (2) the competence of lecturers who had been trained in cooperative learning; (3) high

student enthusiasm because they had rarely experienced structured collaborative learning before; (4) institutional support in the form of providing LKM and access to digital references. Meanwhile, the inhibiting factors encountered included: (1) the presence of dominant students who tended to take over group discussions, requiring more balanced role arrangements; (2) time constraints during the individual quiz phase to ensure no cheating occurred; (3) initial resistance from some students accustomed to passive learning styles (only listening to lectures) who felt "forced" to be active in discussions.

However, the phenomenon of student resistance to cooperative learning is not unique to STAD implementation. Research by Ziziumiza et al. ([2022](#)) on the effectiveness of STAD cooperative learning in improving mathematics skills among vocational and technical education (VTE) engineering students found that while the model significantly improved learning outcomes, its success was contingent on careful management of group dynamics. Similarly, Utami ([2022](#)), in a study on the application of cooperative learning to increase activeness and learning outcomes in elementary mathematics, highlighted situations where high-achieving students may feel that their progress is hindered by the responsibility of helping their peers, or conversely, where an imbalance of roles emerges within groups, with some students dominating discussions while others passively observe.

Further complexity is revealed by a design research study by Pham and Renshaw ([2015](#)) investigating how university teachers and students in Vietnam perceived and adapted STAD. The study found that both teachers and students resisted many key procedures and practices of STAD. Specifically, Vietnamese university students did not work effectively in mixed-ability groups, found role-rotating within groups to be an ineffective strategy for promoting learning, and engaged only superficially in small-group study without the teacher's supervision and scaffolding. Instead, they worked more effectively in friendship groups, participated in group work more fairly under the management of a designated group leader, and engaged in more complex knowledge construction when guided by scaffolding questions from the teacher ([Pham & Renshaw, 2015](#)). This finding is particularly relevant to the Indonesian context, where similar cultural expectations about classroom interaction such as respect for teacher authority and preference for familiar peer relationships may influence how students respond to STAD implementation.

In the present study, the initial resistance observed among approximately 15% of students in the late majority and laggard categories manifested in passive attitudes and reluctance to participate actively during the early meetings. This resistance can be understood through multiple lenses. First, from the perspective of high-achieving students, the responsibility of helping lower-ability peers may be perceived as a burden that slows their own learning progress ([Utami, 2022](#)). Second, from the perspective of lower-achieving students, participation in mixed-ability groups may create feelings of inadequacy or embarrassment, leading to passive withdrawal ([Pham & Renshaw, 2015](#)). Third, from a cultural perspective, Indonesian students accustomed to teacher-centered, individualistic learning may initially resist the shift toward collaborative, peer-mediated learning, particularly when group roles and accountability mechanisms are not clearly established ([Pham & Renshaw, 2015](#); [Syaifullah et al., 2024](#)).

However, as time progressed, the group reward system and positive pressure from more motivated group members succeeded in encouraging resistant students to become more actively involved. This finding confirms Slavin's ([2015](#)) statement that group accountability mechanisms in STAD are effective for overcoming social loafing problems that often occur in group learning. Importantly, the present study also identified that the integration of religious moderation values and local Lampung wisdom into Arabic language material helped reduce initial resistance, as students perceived the learning content as more relevant and meaningful. This suggests that cultural adaptation of STAD materials as recommended by Pham and Renshaw ([2015](#)) can mitigate resistance and enhance student engagement.

From the perspective of Vygotsky's social constructivism theory, the findings of this study reinforce the idea that effective learning occurs when students are involved in meaningful social interaction with peers who have slightly higher ability levels (ZPD). In heterogeneous STAD groups, higher-ability students naturally become peer tutors, while lower-ability students receive appropriate scaffolding ([Mendo-Lázaro, 2022](#); [Motwani, 2022](#)). This process not only transfers knowledge from the more capable to the less capable but also deepens the peer tutors' understanding because they must explain concepts in easy-to-understand language, identify the tutee's misunderstandings, and provide feedback. Research by Jundi et al. ([2023](#)) emphasizes that students in STAD implementation show increased awareness of the importance of collaboration in achieving common academic goals, although it requires consistent time

and effort.

The practical implications of this research for Arabic language lecturers in Islamic higher education are very clear. First, lecturers need to start moving away from the dominance of lecture methods and switch to student-centered learning models, including STAD. Second, effective STAD implementation requires thorough preparation, especially in designing Student Worksheets that contain questions encouraging discussion and problem-solving, not just memorization. Third, lecturers must pay special attention to the formation of truly heterogeneous groups, considering students' initial ability data (pretest), gender, regional origin, and learning styles. Fourth, a creative reward system (not only grades but also certificates, public praise, or small incentives) needs to be designed to maintain student motivation throughout the semester. Fifth, the integration of Islamic values and local context into Arabic language material, as exemplified in this study, has been shown to increase the relevance and meaningfulness of learning in the eyes of students.

This research also makes a theoretical contribution to the development of Arabic language education, particularly in the context of higher education in Indonesia. Previously, most theories of Arabic language learning came from the context of second language acquisition (SLA) in Western or Middle Eastern countries, which may not fully align with the characteristics of Arabic learners in Indonesia where Arabic is a foreign language not used in daily communication ([Wahyuni, 2023](#)). The findings of this study show that the STAD model developed by Slavin ([2012, 2015](#)) can be effectively adapted in the Indonesian context, as long as it is accompanied by material modifications relevant to local and Islamic values. Research by Wahyudi & Hidayat ([2021](#)) emphasizes that STAD is simpler than other cooperative learning strategies and is a good model for teachers who are just starting to apply cooperative strategies in learning ([Budiman, 2020](#)).

The limitations of this study need to be acknowledged and noted for interpretation of results. First, this study only lasted for 8 weeks (one short semester), so it cannot yet measure the long-term retention of the learning outcome improvements achieved. Whether students taught with STAD still remember the Arabic material six months or one year after treatment cannot yet be answered in this study. Second, the research sample was limited to one study program (PAI) at one university (Universitas Ma'arif Lampung), so the generalization of results to other study programs (such as

Islamic Family Law, Islamic Communication Broadcasting, or Islamic Education Management) or to other universities in Lampung needs to be done carefully. Third, this study did not measure the impact of STAD on productive Arabic language skills (speaking and writing), because the research focus was on receptive skills and grammar measured through written tests.

Fourth, the possibility of a Hawthorne effect, where research subjects change their behavior because they are aware they are being observed/researched, cannot be fully controlled. Students in the experimental group might show higher motivation and participation not only because of the STAD model but because they felt "special" for receiving special treatment. To minimize this bias, researchers did not explicitly tell students that they were part of an experimental study, and the lecturer who taught both groups was the same person so that teaching style was relatively controlled. However, there remains a possibility that students realized the difference in treatment between their class and other classes, which could affect results. Research by Fitria & Zulheddi (2024) using a semi-experimental design also acknowledges similar limitations.

Future research is recommended to address these limitations. Longitudinal studies with follow-up tests after 6 or 12 months need to be conducted to measure retention of learning outcomes. Multi-site research involving several universities (both public and private) in Lampung Province or even across provinces will increase the generalization of findings. Experiments with factorial designs testing interactions between the STAD model and moderator variables (such as learning styles, interpersonal intelligence, or self-efficacy levels) can provide a more nuanced understanding of for whom STAD is most effective. Additionally, research with mixed methods combining quantitative experiments with in-depth interviews or focus group discussions (FGD) will enrich understanding of the psychological and sociological mechanisms underlying STAD effectiveness from the students' perspective.

CONCLUSION

However, it is important to contextualize these findings within the rapidly evolving landscape of language education in 2024–2025. The emergence of AI-assisted learning tools, such as ChatGPT-based models for Arabic language learning, has opened new possibilities for individualized practice, instant feedback, and scalable learning opportunities that were previously unavailable in traditional classroom

settings ([Zubaidi et al., 2025](#)). These AI-based approaches offer significant advantages in terms of personalized skill development, particularly for writing, vocabulary expansion, and grammar practice outside of class time. Nevertheless, AI tools have inherent limitations: they cannot fully replicate the social, emotional, and motivational aspects of human interaction, peer accountability, and collaborative meaning-making that occur in cooperative learning models like STAD.

Therefore, rather than positioning STAD as the sole superior alternative to conventional methods, the findings of this study should be interpreted as evidence that STAD remains a highly effective human-centered collaborative approach that addresses dimensions of learning such as peer tutoring, collective responsibility, interpersonal skills, and social regulation that AI tools cannot fully replace. In the contemporary educational ecosystem, STAD and AI-assisted learning should be viewed as complementary rather than competing approaches. STAD provides the structured peer interaction and motivational support that fosters deeper conceptual understanding and collaborative skills, while AI tools can supplement this process by offering additional individualized practice and self-paced learning opportunities. This integrated perspective aligns with the tridimensional human-AI collaboration framework proposed by Zubaidi et al. ([2025](#)), which emphasizes the importance of balancing technological tools with human guidance and peer interaction in Arabic language education.

Future research should explore the integration of STAD with AI-assisted learning in a blended learning environment. Studies examining how AI tools can be incorporated into the STAD syntax, for example, using AI for individualized pre-class preparation or post-class reinforcement while maintaining STAD's core collaborative structure could provide valuable insights for designing more effective and holistic Arabic language curricula in the digital age. Additionally, longitudinal studies are needed to investigate whether the combination of STAD and AI tools can sustain learning motivation and outcomes over longer periods compared to STAD alone or AI alone.

In conclusion, the STAD cooperative learning model is a robust and effective strategy for improving Arabic language learning outcomes and motivation in Islamic higher education, particularly when implemented with attention to cultural context, heterogeneous grouping, and meaningful material integration. However, its

effectiveness should be understood within the broader ecosystem of language learning approaches, where AI-assisted tools offer complementary benefits that can enhance the overall learning experience. Educators are encouraged to adopt a balanced approach that leverages both the collaborative power of STAD and the individualized support of AI technologies to create a more comprehensive and adaptive Arabic language learning environment.

ACKNOWLEDGEMENT

The researchers would like to express their highest gratitude and appreciation to the Rector of Universitas Ma'arif Lampung, the Dean of the Faculty of Tarbiyah and Teacher Training, and the Head of the Islamic Religious Education Study Program for the permission and support provided during the research process. Special thanks are conveyed to all lecturers and students who participated as research subjects, as well as to the two observers who assisted in data collection. This research was funded by the Institute for Research and Community Service (LP3M) of Universitas Ma'arif Lampung through the internal research grant scheme for the 2025/2026 fiscal year. The researchers also thank the anonymous reviewers for their valuable input in improving this article.

REFERENCES

- Aiman, Ummul., Karimuddin Abdullah, Misbahul Jannah, Zahara Fadilla, Suryadin Hasda, Taqwin, Masita, & Meilida Eka Sari. (2022). *Metodologi penelitian kuantitatif*. Banda Aceh: Yayasan Penerbit Muhammad Zaini.
- Arsil, Hamengkubowono, & Rahman, A. (2025). The effect of the implementation of the cooperative learning model type Student Teams Achievement Division (STAD) on creativity, initiative, motivation and student learning outcomes. *Jurnal Ilmiah Pendidikan Islam*, 9(1), 45-62. <https://journal.iaimnumetrolampung.ac.id/index.php/ji>
- Budiman, A. (2020). *Pembelajaran kooperatif tipe STAD dan pengaruhnya bagi kemampuan berpikir kritis dan efikasi diri*. Bandung: Alfabeta.
- Edidarmo, T., Fudhaili, A., & Mahfuzo, M. R. (2023). The power of spiritual motivation: A conceptual and theoretical review of Arabic language learning. *Arabiyatuna: Jurnal Bahasa Arab*, 7(1), 315–344. <https://doi.org/10.29240/jba.v7i1.5629>
- Faqih, Muhammad Khoirul, Ikhwan Aziz, dan Rina Mida Hayati. (2025). "The Use of Learning Models Cooperative Learning Type Student Teams Achievement Division (STAD) in Islamic Education Subjects to Increase Learning Motivation." *International Journal on Advanced Science, Education, and Religion* 8, no. 2. 203-217. <https://doi.org/10.33648/ijoaser.v8i2.942>
- Fitria, W., & Zulheddi. (2024). The impact of STAD on Arabic learning outcomes of tenth grade in senior high school student. *Lisanuna: Jurnal Ilmu Bahasa Arab dan Pembelajarannya*, 14(1), 34-52. <https://jurnal.ar-raniry.ac.id/index.php/lisanuna>
- Gusma, Arif Jun, W. Yul, H. Oktavera, dan A. S. Ulya. (2025). "Exploring Project-Based Learning's Role in the Development of Arabic Students' Speaking Skills through Experiential Learning Theory." *Mantiqutayr: Journal of Arabic Language* 5, no. 2. 161-182. <https://doi.org/10.25217/mantiqutayr.v5i2.5792>
- Hamidah, N. Z., & Machmudah, U. (2025). Internalisasi nilai-nilai moderasi beragama pada pembelajaran bahasa Arab menggunakan cooperative learning model Student Team Achievement Divisions (STAD). *Jurnal Pendidikan Bahasa Arab*, 11(2), 187-204. <https://repository.uin-malang.ac.id/19527/>
- Hanun, A., & Mukminin, A. (2019). Implementasi model pembelajaran cooperative learning tipe STAD dalam pembelajaran insya'. *Studi Arab*, 10(1), 23-38. <https://doi.org/10.35891/sa.v10i1.1769>
- Kementerian Agama RI. (2019). *Keputusan Menteri Agama Nomor 183 Tahun 2019 tentang Kurikulum PAI dan Bahasa Arab pada Madrasah*. Jakarta: Kemenag RI.

- Khoiriyah, E. (2024). Pengaruh model pembelajaran kooperatif tipe STAD terhadap kecerdasan interpersonal siswa di MA Plus Keterampilan Hasanah Pekanbaru. *Skripsi*, Universitas Islam Riau. <https://repository.uir.ac.id/29423/>
- Machali, I. (2021). *Metode penelitian kuantitatif*. Yogyakarta: UIN Sunan Kalijaga Press.
- Mendo-Lázaro, S., León-del-Barco, B., Felipe-Castaño, E., & Polo-del-Río, M. I. (2022). Cooperative learning and the development of social skills. *Frontiers in Psychology*, 13, 789-801. <https://doi.org/10.3389/fpsyg.2022.789801>
- Motwani, K., Gupta, R., & Sharma, R. (2022). Cognitive scaffolding in cooperative learning environments. *Educational Psychology Review*, 34(2), 567-589. <https://doi.org/10.1007/s10648-021-09645-8>
- Mukhyi, M. A. (2023). *Metodologi penelitian panduan praktis penelitian yang efektif*. Jakarta: Penerbit Bumi Aksara.
- Muradi, A., Ahmad, Amaluddin, L. O., Hilmi, D., Budiawan, H., Noor, F., Noor, H., & Jailani, M. S. (2022). Evaluation of Arabic learning program at state Islamic religious college during Covid-19. *Eurasian Journal of Educational Research*, 99(99), 189-200. <http://repository.uin-malang.ac.id/13234/>
- Nabila, S. A. (2025). Implementasi metode pembelajaran kooperatif tipe STAD berbasis poster digital untuk meningkatkan keterampilan membaca bahasa Arab siswa kelas X MAN 1 Pasuruan. *Skripsi*, Universitas Negeri Malang. <https://repository.um.ac.id/408165/>
- Pellegrini, M., Day, E., Scarbrough, H. F., & Pigott, T. D. (2025). A Meta-Review of Education Meta-Analyses: Relevance, Applicability, and Accessibility of Findings. *AERA Open*, 11. <https://doi.org/10.1177/233285842513895>
- Pham, T. T. H., & Renshaw, P. (2015). Adapting evidence-based pedagogy to local cultural contexts: A design research study of policy borrowing in Vietnam. *Pedagogies: An International Journal*, 10*(3), 256-274. <https://doi.org/10.1080/1554480X.2015.1009836>
- Puspitaningrum, M., Ismail, M., Budiman, A., Al-latifatuzzahra, S. S., & Amaliana, S. R. (2024). Tākwīnu āl-Bī'āh āl-Lughāwiyyāh ās-Sām'iyyāh āl-Bāshoriyyāh lī āt-Tholibāt āl-Jādidāt fī āl-Fāshl āl-Awwāl bī Mā'hādī Dārissālām Gontor Ngawi. *Mantiqū Tayr: Journal of Arabic Language*, 4(2), 551-568. <https://doi.org/10.25217/mantiqutayr.v4i2.4624>
- Rifkin, B. (2005). A ceiling effect in traditional classroom foreign language instruction: Data from Russian. *The Modern Language Journal*, 89(1), 3-18. <https://doi.org/10.1111/j.0026-7902.2005.00262.x>

- Silmy, A. N., Lubis, R. H., Wardani, Y. K., Shara, & Ismahani, A. (2024). Urgensi Metode Belajar dalam Pembelajaran Bahasa Arab (Bagi Penutur Non-Arab). *Mantiqutayr: Journal of Arabic Language*, 4(2), 368–381. <https://doi.org/10.25217/mantiquutayr.v4i2.4423>
- Slavin, R. E. (2012). *Educational psychology: Theory and practice* (10th ed.). Boston: Pearson.
- Suhendri. (2024). Integrating tradition and modernity in Islamic higher education. *Waraqat: Jurnal Ilmu-Ilmu Keislaman*, 9(2), 112-128. <https://doi.org/10.51590/waraqat.v9i2.884>
- Syaifullah, M., & Izzah, N. (2019). Kajian Teoritis Pengembangan Bahan Ajar Bahasa Arab. *Arabiyatuna: Jurnal Bahasa Arab*, 3(1 May), 127–144. <https://doi.org/10.29240/jba.v3i1.764>
- Syaifullah, Muhammad. et al. (2023) "Pendampingan Guru Bahasa Arab Pondok Pesantren Nahdlotul Ulum Kota Metro dalam Pengembangan Media Pembelajaran Bahasa Arab Berbasis Teknologi, Informasi & Komunikasi (ICT)." *Jurnal Dedikasi* 3, no. 2. <https://doi.org/10.32678/dedikasi.v16i2.9416>
- Syaifullah, Muhammad, M. Jafar Shodiq, dan Kiki Kustina. (2024). "The Intercultural Aspect of Teaching Arabic as a Foreign Language in Indonesia." *Mantiqutayr: Journal of Arabic Language*, 4, no. 1. 279-293. <https://doi.org/10.25217/mantiquutayr.v4i1.4027>
- Tim Redaksi Kemenag RI. (2019). *Moderasi beragama*. Jakarta: Badan Litbang dan Diklat Kementerian Agama RI.
- Utami, E. S. (2022). The application of a cooperative learning model to increase the activeness and learning outcomes of mathematics in class VI elementary school. *Social, Humanities, and Educational Studies (SHES): Conference Series*, *5*(Special Issue 2). <https://doi.org/10.20961/shes.v5i6.81060>
- Wahyudi, M., & Hidayat, A. R. (2021). Strategi cooperative learning tipe Student Teams Achievement Division (STAD) pada mata pelajaran bahasa Arab. *Asatiza: Jurnal Pendidikan*, 2(3), 156-172. <https://doi.org/10.46963/asatiza.v2i3.340>
- Wahyuni, S. (2023). *Model pembelajaran kooperatif CIRC dan kebiasaan membaca dalam keterampilan menulis*. Padang: Penerbit UNP Press.
- Weng, C., Otanga, K., & Chen, J. (2022). 4C skills development in higher education. *Journal of Educational Computing Research*, 60(4), 891-912. <https://doi.org/10.1177/07356331211062985>

- Yahya, A., M. Ahsanuddin, Irhamni, H. Syukur, dan M. N. Iman. (2025). "Implementation and Strategies Differentiated Learning for Arabic Language Learning at Islamic High School." *Mantiqu Tayr: Journal of Arabic Language* 5, no. 2. 364-380. <https://doi.org/10.25217/mantiquayr.v5i2.6203>
- Yusliza, Hizbullah, A., Annisa, A. N., & Aminah. (2024). Implementation of STAD-type cooperative learning method in insha' learning at IAI Al-Muslim Aceh. *Jurnal Shawtul Arab*, 12(1), 78-95. <https://jurnal.iuqibogor.ac.id/index.php/shawtularab>
- Ziziumiza, S. (2022). The effectiveness of Student Teams Achievement Division cooperative learning in improving mathematics skills in VTE engineering students. *International Journal of Pedagogy and Teacher Education*, *6*(2), 52-60. <https://doi.org/10.20961/ijpte.v6i2.64003>
- Zubaidi, A., Munip, A., Widodo, S. A., & Zerrouki, T. (2025). Enhancing Arabic writing skills using Chat GPT-based AI learning models: A tridimensional human-AI collaboration framework. *Indonesian Journal of Applied Linguistics*, 15(1), 87-101. <https://doi.org/10.17509/ijal.v15i1.75378>