Prosiding
Konferensi Nasional Inovasi Pembelajaran Bahasa dan Sastra Arab
Volume 5 Desember 2025
ISSN (p) 2828-5603
ISSN (e) 3026-1511

Peer Tutoring Strategy for Improving Learning Achievement in the *Qirā'ah* Course: A Study of Arabic Language Education (*PBA*) Students at UIN Sunan Kudus and UIN Sultan Thaha Saifuddin Jambi

Habib Maulana Maslahul Adi[1*], Zakiyatunnisa Al Mubarokah[2]
UIN Sunan Kudus[1], UIN Sulthan Thaha Saifuddin Jambi[2]
e-mail : [1]maslahulhabib@uinsuku.ac.id, [2]zakiyatunnisa@uinjambi.ac.id
*Corresponding Author

Abstrak: Reading (*qirā'ah*) courses in Arabic teacher education often face limited student engagement, uneven participation, and insufficient formative feedback during text processing. This study investigates whether a structured peer tutoring strategy can improve qirā'ah achievement among Arabic Language Education students at two Indonesian Islamic universities (UIN Sunan Kudus and UIN Sulthan Thaha Saifuddin Jambi). Using a pre-experimental one-group pretest–posttest design, students participated in guided peer tutoring sessions in which trained peers supported classmates through staged reading activities (previewing, vocabulary support, questioning, and summarising) followed by brief reflection. Achievement data were collected through parallel reading comprehension tests administered before and after the intervention and analysed using paired-difference procedures with a nonparametric robustness check. The findings indicate that peer tutoring leads to a statistically meaningful improvement in *qirā'ah* performance across both sites, with generally consistent gains across student groups, although contextual classroom dynamics may shape the magnitude of benefits for specific subgroups.

Keywords: peer tutoring; *qirā'ah*; reading comprehension; Arabic learning

INTRODUCTION

Reading proficiency (*qirā'ah*), understood as the ability to comprehend Arabic discourse accurately, fluently, and analytically, constitutes a main pillar of academic literacy in Arabic Language Education (*PBA*) programmes at the university level. The demands of *qirā'ah* in undergraduate courses do not stop at recitation and literal translation, but presuppose competence in drawing inferences, assessing rhetorical structures, and connecting authentic information across texts.(Hartati & Hanafi, 2022). In the global context of twenty-first-century education, this competence is closely related to higher-order thinking skills (HOTS), collaboration, and digital literacy, all of which are now mainstreamed in the design of second/foreign language learning (Susanto et al., 2022).

Qirā'ah instruction in Indonesian higher education still faces several challenges, including materials that are insufficiently contextual (insufficiently grounded in local culture), dependence on Middle Eastern textbooks that do not always align with the profile of Indonesian learners, and the use of media and assessment that is not yet systematically oriented towards fostering critical understanding (Nuruddin et al., 2017). These findings reinforce the need for active, collaborative, and student-centred pedagogical interventions so that *qirā'ah* achievement can improve in a meaningful and sustainable manner (Arifin et al., 2024).

One relevant approach is peer tutoring. Theoretically, this strategy is rooted in Vygotsky's social constructivism, particularly the notion of the zone of proximal development (ZPD), which emphasizes that developmental leaps occur when learners are assisted by a "more capable other" (in this context, a competent peer) (Vygotsky, 1978). The scaffolding that takes place within peer interaction enables a transition from actual competence to a higher potential level, in line with the principle of internalization through socio-cultural practice. At the same time, social interdependence theory asserts that positive interdependence, individual accountability, and promotive interaction within cooperative groups are correlated with increases in academic achievement and social skills (Johnson & Johnson, 2009). In a similar vein, the cognitive elaboration approach shows that rich explanations and high-quality feedback among peers are a key lever for deeper understanding, not only for those being helped but also for the "tutors" themselves (Roscoe & Chi, 2007; Webb, 1989).

Empirical evidence from various studies reinforces this foundation. A broad synthesis of hundreds of meta-analyses in education reports a moderate effect of peer tutoring strategies on learning outcomes, placing them above the average impact of conventional teaching practices (Hattie, 2009). Classical narrative reviews from Topping (2005) also confirm the positive trend of peer learning (peer tutoring, cooperative learning, and peer assessment), provided that the integrity of implementation is maintained. More recent findings, particularly meta-analyses in higher education, report a moderately positive effect of peer tutoring on university students' academic achievement, with stronger effects observed in more individualized interventions and Asian contexts (Zhang et al., 2025). Meanwhile, among emergent bilinguals, peer tutoring has consistently been shown to exert a positive impact on academic outcomes across educational levels (Romero et al., 2025).

In the domain of Arabic language instruction, especially *qirā'ah*, various studies in Indonesia indicate the effectiveness of peer tutoring or related peer-based practices. Research on peer tutoring in *qirā'ah* instruction has documented increases in students' scores and in their learning preferences for more structured *qirā'ah* activities (Astifa et al., 2023). Other studies integrating peer tutoring with short-card media in madrasah settings have shown improvements in participation and learning outcomes in Arabic (Rizal et al., 2023; Taufiq et al., 2022). Beyond pure *qirā'ah* contexts, peer tutoring has also been effective in enhancing the ability to read and memorize Qur'anic verses, further underscoring the leverage of peer-based strategies for receptive skills (Amri, 2024). At the same time, the literature on HOTS-based *qirā'ah* assessment and interactive testing grounded in the CEFR highlights the need for valid and reliable instruments to capture progress in comprehension rather than merely phonetic accuracy (Azzahra & Baroroh, 2024; Mustakim et al., 2024).

Nevertheless, a research gap remains evident, particularly in the context of Islamic higher education institutions. *First*, there is still a limited number of quantitative studies that explicitly compare *qirā'ah* achievement before and after the implementation of peer tutoring at the university level; many studies focus on describing implementation or developing media without isolating and measuring the impact (Khuluq & Imamah, 2024). Second, the diversity of instruments (rubrics, benchmarks, and item types) makes cross-study comparison difficult–even though mastery of *qirā'ah* discourse requires indicators that encompass fluency, accuracy, vocabulary, coherence, and inference (Azzahra & Baroroh, 2024). Third, there are still relatively few reports that sharpen the parameters

ISSN (p) 2828-5603
ISSN (e) 3026-1511

of peer tutoring implementation (e.g., frequency of sessions, reciprocal vs. non-reciprocal role models, or feedback mechanisms) that have been shown to moderate learning outcomes in higher education (Topping, 2005).

On the basis of the foregoing discussion, the novelty of this study lies in: (1) its specific focus on *qirā'ah* achievement in Islamic higher education, with a clear comparison of learning outcomes before and after the implementation of peer tutoring; (2) the refinement of *qirā'ah* achievement dimensions (fluency, accuracy, text comprehension, and inference) so that they are aligned with academic literacy; and (3) its contextual contribution from Indonesian State Islamic Higher Education Institutions (*PTKIN*), which remain relatively underexplored in the international literature, thereby adding a Indonesian perspective to global discussions on Arabic language pedagogy. Accordingly, the research objective is formulated explicitly: to compare students' learning achievement in the *qirā'ah* course before and after the implementation of peer tutoring. The expected outcome is not only to enrich the quantitative evidence on the effectiveness of peer tutoring in *qirā'ah*, but also to offer practical recommendations for Arabic Language Education lecturers in designing activities and assessments that more accurately reflect progress in reading comprehension.

METHOD

This study employed a quantitative approach using a pre-experimental multisite one-group pretest–posttest design to examine changes in *qirā'ah* learning achievement before and after the implementation of a peer tutoring strategy. This design choice is consistent with the aim of assessing changes in scores following the intervention in an intact class (Creswell & Creswell, 2018). The research was conducted in the Arabic Language Education (PBA) Study Programs at UIN Sunan Kudus and UIN Sultan Thaha Saifuddin Jambi, involving all third-semester students enrolled in the *qirā'ah* course (N = 151), comprising 84 students from UIN Sunan Kudus and 67 students from UIN Sultan Thaha Saifuddin Jambi; by sex, the sample included 96 female and 55 male students. *qirā'ah* achievement was measured using an achievement test and/or performance-based reading assessment scored with a structured rubric. The research procedures consisted of (1) a pretest, (2) a peer tutoring intervention implemented for approximately 6 weeks (1 session per week), including brief tutor training and monitoring of implementation fidelity using a checklist, and (3) a posttest administered using instruments of equivalent difficulty. Data were analyzed using descriptive statistics, assumption testing, pretest–posttest comparisons (paired-samples t-test or Wilcoxon signed-rank test, depending on assumptions), and gain-score comparisons to examine differences in improvement by institution and sex, with effect sizes reported, calculated using the SPSS version 25 application..

RESULTS AND DISCUSSION

Data Screening and Participant Profile

The dataset comprised 151 paired observations (pretest and posttest) from third-semester Arabic Language Education (PBA) students enrolled in the *qirā'ah* course at two public Islamic universities. Data screening indicated no missing values in the

achievement scores and a plausible score range (pretest 60–100; posttest 70–100). The gain score (post − pre) ranged from −5 to 25, with 135 students showing positive gains, 15 showing no change, and only one student showing a negative gain. This profile suggests that the intervention's improvement was broadly distributed across the cohort, although the magnitude of improvement varied across individuals.

Table 1 *Participant distribution by institution and sex*

Institution	Female (n)	Male (n)	Total (n)
UIN Sunan Kudus	55	29	84
UIN Sultan Thaha Saifuddin Jambi	41	26	67
Total	**96**	**55**	**151**

Because this study used an intact-class, one-group pretest–posttest design, the analyses focus on: (a) whether mean achievement changed after the peer tutoring intervention; and (b) whether the observed gains were robust across institutions and sexs. Interpretations are offered with explicit attention to design-based validity constraints, including common threats in single-group pretest–posttest studies (e.g., testing, maturation, history) (Shadish et al., 2002).

Table 2 summarises the overall descriptive statistics. Mean *qirā'ah* achievement increased from pretest (M = 74.77, SD = 9.63) to posttest (M = 84.11, SD = 8.83), yielding a mean gain of 9.34 points (SD = 6.02). In substantive terms, the mean shifted by approximately 9–10 points on a 0–100 scale, which is typically large enough to be pedagogically visible in course-level assessment contexts.

Table 2 *Descriptive statistics for pretest, posttest, and gain (N = 151)*

Measure	M	Std. Dev.	Min	Max
Pre-test	74.77	9.63	60	100
Post-test	84.11	8.83	70	100
Gain (post−pre)	9.34	6.02	-5	25

Note. Scores are presented on a 0–100 scale.

Inferential testing confirmed that the improvement of *qirā'ah* achevement was statistically reliable.

The Effectiveness of Peer Tutoring Strategy in *Qirā'ah* Learning

a. Overall Effectiveness of Peer Tutoring on Qirā'ah Achievement

A paired-samples t-test indicated a significant increase in posttest scores relative to pretest, $t(150) = 19.06$, $p < .001$, with a large within-subject effect size (Cohen's $d_z = 1.55$). The 95% confidence interval for the mean gain was [8.37, 10.31], suggesting that the average improvement is unlikely to be a trivial fluctuation. Because the gain distribution departed from normality (Shapiro–Wilk $W = 0.927$, $p < .001$), a nonparametric robustness check was also conducted. The Wilcoxon signed-rank test similarly indicated a significant positive shift ($W = 25$, $p < .001$), with a large rank-based effect ($r \approx 0.86$). The convergence of parametric and nonparametric results strengthens confidence that the observed change reflects a consistent pattern across the cohort rather than sensitivity to distributional assumptions.

Table 3 *Summary of key statistical analyses*

Analysis	Test statistic	p	Effect size	Notes
Paired t-test (post vs. pre)	t(150) = 19.06	< .001	Cohen's $d_z = 1.55$	Mean gain = 9.34; 95% CI [8.37, 10.31]

ISSN (p) 2828-5603
ISSN (e) 3026-1511

Wilcoxon signed-rank (nonparametric)	W = 25	< .001	r = 0.86	n (nonzero) = 136
Shapiro–Wilk normality test (gain)	W = 0.927	< .001	–	Non-normal gain distribution
Two-way ANOVA on gain: Institution main effect	F(1,147) = 0.99	= 0.321	ηp^2 = 0.007	–
Two-way ANOVA on gain: Sex main effect	F(1,147) = 0.05	= 0.820	ηp^2 = 0.000	–
Two-way ANOVA on gain: Institution × Sex	F(1,147) = 7.43	= 0.007	ηp^2 = 0.048	Crossover interaction observed

Note. Effect sizes are reported as Cohen's d_z for the paired t-test, r for the Wilcoxon signed-rank test, and partial eta squared (ηp^2) for the ANOVA.

Figure 1 visualises the pretest and posttest means with 95% confidence intervals. The separation between the intervals offers a clear visual complement to the inferential tests and underscores the magnitude of improvement.

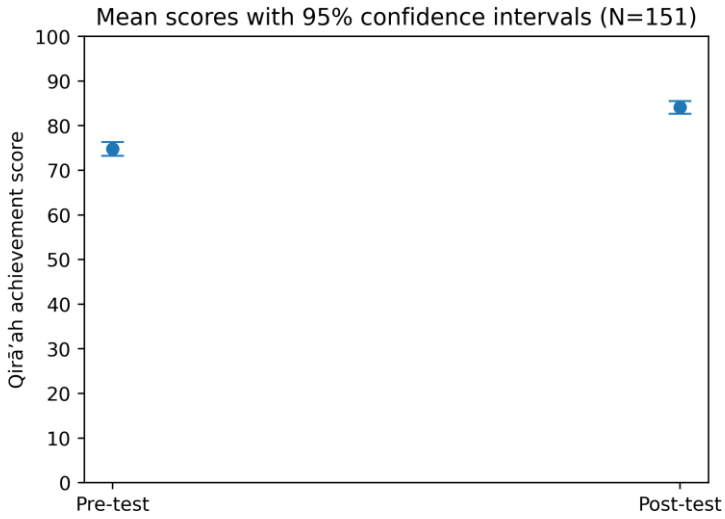


Figure 1 *Mean qirā'ah achievement scores with 95% confidence intervals*

Benchmarking the observed improvement against prior evidence is informative. Meta-analytic reviews of peer-assisted learning and tutoring generally report positive, educationally meaningful effects on achievement, often in the small-to-moderate range depending on design and outcome domain (Ginsburg-Block et al., 2006; Rohrbeck et al., 2003). In the present study, the within-subject effect size is very large (d_z = 1.55), which may reflect both genuine instructional impact and design-related inflation. Single-group pretest–posttest designs can yield larger apparent effects because they do not subtract gains that might occur under business-as-usual instruction, and because practice with similar task formats can increase posttest performance even without a targeted intervention (Shadish et al., 2002). Accordingly, the most defensible conclusion is that students' *qirā'ah* achievement improved substantially during the peer tutoring period, while a stronger causal claim would require a control or comparison condition.

From a theoretical perspective, the size and consistency of the improvement is compatible with major explanatory accounts of peer tutoring and peer learning: (1) Social interdependence theory predicts that structured collaborative arrangements increase achievement when positive interdependence, individual accountability, and promotive interaction are present (Johnson & Johnson, 2009). A peer tutoring format makes these conditions more likely, particularly when tutors receive role guidance and

implementation fidelity is monitored, because students' success becomes linked to cooperative task completion rather than isolated performance; (2) Research on tutor learning indicates that tutors consolidate and reorganise knowledge when they provide elaborated explanations and diagnostic questions, moving beyond simple answer-giving (Roscoe & Chi, 2007); and (3) the ICAP framework predicts that interactive learning activities, dialogic exchange with co-construction, produce greater learning than merely active rehearsal (Chi, 2009). In *qirā'ah* tutoring sessions, students can shift from individual decoding to interactive comprehension-building by negotiating meaning, justifying inferences, and jointly repairing misunderstanding. These scaffolding dynamics align with the zone of proximal development perspective, in which learning is supported through guided participation with a more capable peer (Vygotsky, 1978).

This theoretical interpretation is consistent with applied peer tutoring research in reading. Large-scale evidence indicates that peer tutoring supports reading comprehension when pairs engage in explicit strategy use (e.g., predicting, clarifying, questioning, summarising), repeated text engagement, and joint construction of answers (Flores et al., 2024). In Indonesian Arabic language education contexts, peer tutoring has also been reported to increase students' willingness to ask questions, reduce anxiety, and improve engagement during reading activities–factors that can plausibly strengthen comprehension outcomes (Hasanah, 2020).

b. Practical Significance and Distribution of Learning Gains

Beyond mean differences, the distribution of gains provides important information for curriculum decisions. Figure 2 indicates that gains were predominantly positive and clustered between 5 and 15 points, with a long right tail suggesting that a subset of students achieved gains above 15 points. Only one student experienced a decline, and 15 students showed no net change. Such a distribution is typical for interventions that combine structured practice with individual differences in engagement and prior proficiency (Rohrbeck et al., 2003; Topping, 2005).

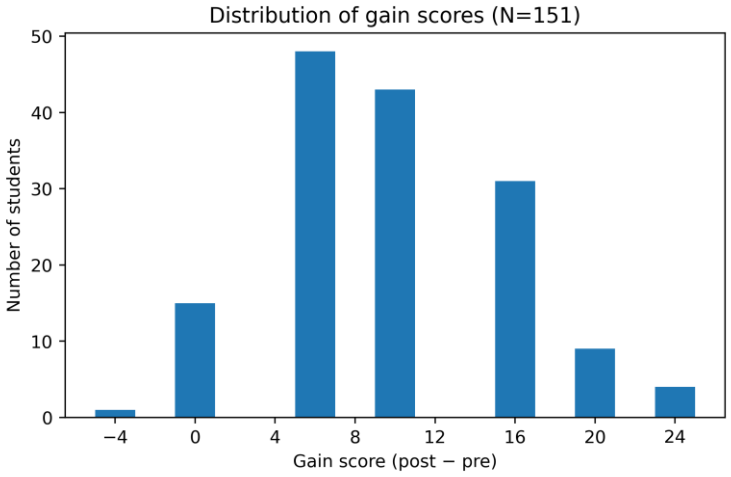


Figure 2 *Histogram of gain scores (post – pre)*

To express improvement in a way that is less sensitive to baseline score differences, a normalised gain (N-gain) index was computed for students who were not at ceiling on the pretest (i.e., pretest < 100). Using the widely cited normalised gain formulation (Hake, 1998), the mean N-gain was approximately 0.40 (40.2%), commonly interpreted as a 'moderate' learning gain in pretest–posttest contexts. As shown in Table 4, 45.6% of students fell in the moderate N-gain band (0.30–0.70), 13.6% reached the high band

ISSN (p) 2828-5603
ISSN (e) 3026-1511

(>0.70), and 40.8% were in the low band (<0.30). The practical implication is twofold: (1) Peer tutoring generated meaningful progress for many students; (2) but a substantial minority may require additional scaffolding (e.g., more explicit comprehension strategy prompts, closer monitoring of tutor questioning quality) to realise stronger gains (Flores et al., 2024; Roscoe & Chi, 2007).

Table 4 *Normalised gain (N-gain) categories among students with pretest < 100 (n = 147)*

Category	n	Percent
Low (<0.30)	60	40.8%
Moderate (0.30–0.70)	67	45.6%
High (>0.70)	20	13.6%

Ceiling effects and regression patterns were also evident. There was a moderate negative correlation between pretest scores and gain scores ($r = -0.44$, $p < .001$), indicating that "students who began at higher proficiency tended to gain fewer raw points". This is expected when the test has a fixed maximum and when instruction primarily helps learners stabilise comprehension routines that are not yet automatised. In practical terms, "peer tutoring may function as an equalising mechanism by accelerating growth among mid- and lower-performing students", which is a pattern frequently reported in peer-assisted learning research (Ginsburg-Block et al., 2006; Rohrbeck et al., 2003).

c. Robustness Across Institutions and Sex

Because the study was conducted in two institutional contexts, it is important to examine whether the pre–post improvement generalised across institutions. At the descriptive level, both institutions showed substantial gains. Students at UIN Sunan Kudus improved from M = 73.99 (SD = 9.11) to M = 83.75 (SD = 9.15), with mean gain 9.76. Students at UIN Sultan Thaha Saifuddin Jambi improved from M = 75.75 (SD = 10.23) to M = 84.55 (SD = 8.47), with mean gain 8.81.

	n	Pre-T Mean (SD)	Post-T Mean (SD)	Gain Mean (SD)
UIN Sunan Kudus	84	73.99 (9.11)	83.75 (9.15)	9.76 (6.20)
UIN STS Jambi	67	75.75 (10.23)	84.55 (8.47)	8.81 (5.78)

An independent-samples comparison of gain scores between institutions (Welch's t-test) did not indicate a statistically significant difference, $t(145.41) = 0.98$, $p = .330$. A Mann–Whitney U test yielded the same conclusion ($U = 3061.5$, $p = .339$). The practical implication is that "the peer tutoring model appears to function similarly across the two sites, despite potential differences in local teaching culture, cohort composition, or instructional routines".

When gains were compared by sex alone, there was no meaningful difference at the overall sample level. Female students had a mean gain of 9.27 (SD = 5.76), while male students had a mean gain of 9.45 (SD = 6.50). Welch's t-test indicated no significant difference, $t(101.73) = -0.17$, $p = .862$, and the Mann–Whitney U test was likewise nonsignificant ($p = .946$). This suggests that "peer tutoring, as implemented, was broadly equitable with respect to sex at the aggregate level".

Sex	n	Pre-T Mean (SD)	Post-T Mean (SD)	Gain Mean (SD)
Male	55	73.45 (8.04)	82.91 (8.32)	9.45 (6.50)
Female	96	75.52 (10.40)	84.79 (9.09)	9.27 (5.76)

In the peer learning literature, such cross-context robustness is taken as evidence that an intervention is not merely an artefact of a single classroom's idiosyncrasies (Topping, 2005) and is consistent with theoretical expectations that cooperative

structures operate through general learning mechanisms rather than location-specific factors (Johnson & Johnson, 2009).

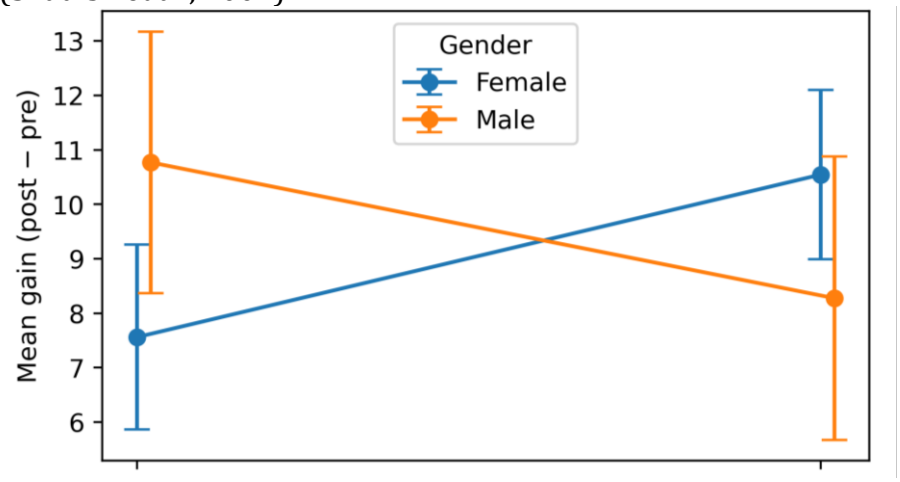
d. Institutional Interaction Patterns × Sex

A More nuanced pattern emerged when institution × sex context was considered simultaneously. A two-way ANOVA on gain scores revealed a statistically significant institution × sex interaction, $F(1,147) = 7.43$, $p = .007$, $\eta p^2 = .048$, while the main effects of institution and sex were nonsignificant. As displayed in Table 5 and Figure 3, this interaction took a crossover form: "at UIN Sunan Kudus, female students tended to gain more than male students" (female M gain = 10.55; male M gain = 8.28), whereas "at UIN Sultan Thaha Saifuddin Jambi, male students tended to gain more than female students" (male M gain = 10.77; female M gain = 7.56). An ANCOVA predicting posttest scores while controlling for pretest yielded a similar interaction ($p = .034$), suggesting that the crossover is not solely a reflection of baseline differences.

Table 5 *Descriptive statistics by institution and sex*

	Sex	n	Pre M (SD)	Post M (SD)	Gain M (SD)
UIN STS Jambi	Female	41	77.80 (10.96)	85.37 (8.76)	7.56 (5.38)
UIN STS Jambi	Male	26	72.50 (8.15)	83.27 (7.99)	10.77 (5.95)
UIN Sunan Kudus	Female	55	73.82 (9.72)	84.36 (9.38)	10.55 (5.75)
UIN Sunan Kudus	Male	29	74.31 (7.99)	82.59 (8.72)	8.28 (6.85)

Figure 3 provides a visual summary of the crossover. Importantly, all four institution–sex subgroups improved significantly from pretest to posttest, so the interaction should not be interpreted as "effective for one subgroup and ineffective for another". Rather, it indicates that the magnitude of gain was different across contexts. Because the design lacks a control group and does not randomise students into different tutoring structures, the interaction should be treated as exploratory and hypothesis-generating (Shadish et al., 2002).



Note. Left: UIN STS Jambi & Right: UIN Sunan Kudus

Figure 3 *Mean gain by institution and sex with 95% confidence intervals*

A plausible explanatory pathway for such an interaction can be formulated using established peer-learning mechanisms. Peer tutoring effectiveness depends on the quality of interaction–particularly the extent to which pairs engage in elaborated explanations, strategic questioning, and joint monitoring of understanding (Chi, 2009; Roscoe & Chi, 2007). Differences in how students participate in peer dialogue–potentially influenced by local classroom norms, mixed-sex interaction patterns, and the degree to which tutors are trained to elicit contributions from quieter partners–may affect who benefits most. Evidence from peer tutoring for reading indicates that explicit strategy talk

ISSN (p) 2828-5603
ISSN (e) 3026-1511

and joint construction of answers are core features linked to gains (Flores et al., 2024). Thus, the crossover pattern observed here may reflect institution-level differences in how consistently these interactional features were achieved within tutoring sessions.

e. Implications for *Qirā'ah* Pedagogy in Arabic Teacher Education

Taken together, the findings provide convergent evidence that a structured "peer tutoring intervention can improve *qirā'ah* achievement in Arabic teacher education contexts". The magnitude of the observed gain aligns with broader peer-assisted learning evidence that collaborative, structured peer support can raise achievement (Ginsburg-Block et al., 2006; Rohrbeck et al., 2003) and with Indonesian Arabic language education research reporting that peer tutoring can increase engagement and reduce anxiety in reading activities (Hasanah, 2020). At the same time, the distribution of gains suggests that implementation refinements could improve benefits for students who remained in the low N-gain band.

For curriculum implementation, four design principles are particularly relevant: **1)** Structure sessions around comprehension processes rather than answer sharing. Reading research suggests that improvement is linked to explicit strategy use (predicting, clarifying, questioning, summarising), repeated text engagement, and joint construction of responses (Flores et al., 2024); **2)** Provide concise tutor training focused on explanation and questioning. A research about Learning by teaching, indicates that tutoring is most productive when tutors are prompted to ask diagnostic questions, demand justification, and guide peers through reasoning rather than simply provide correct answers (Fiorella & Mayer, 2013; Roscoe & Chi, 2007); **3)** Ensure individual accountability and (where feasible) rotate roles. Social interdependence theory emphasises accountability as a safeguard against passive participation and as a driver of promotive interaction (Johnson & Johnson, 2009); and **4)** Monitor implementation fidelity with lightweight checklists and periodic observation. Peer tutoring's impact depends on consistent enactment of its core features (Topping, 2005).

These implications are directly relevant for Arabic teacher education programmes because *qirā'ah* competence is foundational for subsequent coursework in *tafsīr*, *ḥadīth*, Arabic linguistics, and pedagogical practicum. Embedding peer tutoring as a regular feature of *qirā'ah* courses can also model collaborative pedagogies that future teachers may use in their own classrooms. Moreover, if the course uses an assessment framework aligned with recognised proficiency descriptors (e.g. ACTFL reading proficiency descriptors and CEFR) based performance rubrics–peer tutoring activities can be aligned with target 'can-do' descriptors and used as structured formative assessment opportunities (ACTFL, 2012; COE, 2020).

Finally, the institution-level robustness observed in this dataset indicates that the model is scalable across institutions, but the institution × sex crossover underscores the importance of monitoring participation dynamics. In practical terms, programmes may consider: **1)** explicitly training tutors to invite contributions; **2)** balancing pairings to prevent domination by one partner; and **3)** using scripted prompts to standardise strategy talk. These are low-cost design elements that directly target the interaction mechanisms identified in the peer tutoring literature (Flores et al., 2024; Roscoe & Chi, 2007).

f. Validity Considerations and Directions for Future Research

The findings should be interpreted with appropriate caution due to the study design. In a one-group pretest–posttest design, improvements may reflect threats to internal validity such as maturation, history, testing effects, or regression to the mean

(Campbell & Stanley, 1963; Creswell & Creswell, 2018; Shadish et al., 2002). Although the magnitude and consistency of improvement, together with the short intervention window (6 weeks), make it less likely that maturation alone explains the change, practice effects cannot be ruled out, especially if pretest and posttest tasks are structurally similar. Furthermore, because the intervention and assessments were implemented in intact classes, there may be uncontrolled instructor or classroom factors that covary with the timing of measurement.

Future studies could strengthen causal inference by adding a comparison group (quasi-experimental) or by randomising classes to peer tutoring versus business-as-usual conditions where feasible. Multi-level modelling or mixed-effects models would also allow researchers to account for nested structures (students within classes within institutions) and to examine how tutor training intensity, pairing strategy, or session frequency moderates outcomes. Finally, qualitative analysis of peer interaction–coding the presence of explanation, questioning, and strategy talk–would help connect 'how tutoring was enacted' to 'how much students improved', an approach recommended in recent large-scale peer tutoring research in reading (Flores et al., 2024) and consistent with theoretical accounts emphasising interaction quality as the driver of learning (Chi, 2009; Roscoe & Chi, 2007).

CONCLUSION

This multisite pre-experimental study provides convergent evidence that a structured peer tutoring intervention is associated with a substantial improvement in undergraduate students' *qirā'ah* achievement in two PTKIN contexts. Across 151 paired observations, mean scores increased from 74.77 (SD = 9.63) on the pretest to 84.11 (SD = 8.83) on the posttest, yielding a mean gain of 9.34 points (SD = 6.02) on a 0–100 scale. Inferential tests corroborated this shift, with a significant paired-samples result ($t(150) = 19.06$, $p < .001$) and a large within-subject effect size ($d_z = 1.55$), supported by a consistent nonparametric check (Wilcoxon $p < .001$). Practically, the overall learning gain was moderate by normalised gain (mean N-gain ≈ 0.40), with most students showing positive gains and evidence that peer tutoring may particularly accelerate progress among mid- and lower-performing students.

From an implementation and equity standpoint, the intervention appeared broadly robust across institutions and sexes: gain differences between the two sites and between male–female groups were not statistically meaningful at the aggregate level. However, a significant institution × sex crossover interaction indicates that participation dynamics and local classroom norms may shape "who benefits more" in each context, even though all subgroups improved. Accordingly, Arabic teacher education programmes seeking to scale peer tutoring in *qirā'ah* should emphasise the mechanisms most closely linked to comprehension growth–explicit strategy talk (predicting, clarifying, questioning, summarising), tutor prompting for elaborated explanations, individual accountability, and fidelity monitoring–while actively managing interaction quality to prevent dominance and ensure equitable engagement. Finally, because the one-group pretest–posttest design is vulnerable to threats such as testing and maturation, future research should strengthen causal inference through comparison groups (quasi-experimental or cluster-randomised designs) and complement outcome tests with qualitative/interactional evidence to clarify which tutoring features most strongly drive qirā'ah gains.

ISSN (p) 2828-5603
ISSN (e) 3026-1511

REFERENCES

ACTFL. (2012). *ACTFL Proficiency Guidelines 2012*. American Council on the Teaching of Foreign Languages. https://education.alaska.gov/state_board/october-2023/7.2 ACTFLProficiencyGuidelines2012.pdf

Amri, M. F. (2024). Implementation of Peer Tutoring in Reading and Memorizing Qur'anic Verses for Students of Class IX SMP Negeri 32 Semarang. *JERS: Journal of Education and Religious Studies*, *4*(3). https://doi.org/10.57060/jers-m1jagy60

Arifin, Z., Lubis, T., Ath-Thukhi, A. M., Alsokari, T., Ainin, M., & Taufan, M. (2024). Analyzing the Problems of Arabic Language Learning in Higher Education. *International Journal of Islamic Studies Higher Education*, *3*(3). https://doi.org/10.24036/insight.v3i3.193

Astifa, M. I. R., Syihabuddin, S., & Tantowi, Y. A. (2023). Peer Tutoring Pada Pembelajaran Qiraah di Program Bimbingan Belajar Lisan Arab PBA UPI. *Jurnal Ilmiah Profesi Pendidikan*, *8*(3), 1287–1296. https://doi.org/10.29303/JIPP.V8I3.1468

Azzahra, L. Q., & Baroroh, R. U. (2024). Inovasi Penilaian Maharah al-Qira'ah pada Buku Teks Bahasa Arab Berdasarkan Penilaian HOTS. *Jurnal Alfazuna : Jurnal Pembelajaran Bahasa Arab Dan Kebahasaaraban*, *9*(1), 133–152. https://doi.org/10.15642/alfazuna.v9i1.3962

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

COE. (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Companion). Council of Europe Publishing. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods* (5th ed.). Sage Publications.

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, *38*(3), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Flores, M., Ribosa, J., & Duran, D. (2024). How does peer tutoring contribute to the development of reading comprehension? Evidence from ten years of practice. *Revista de Psicodidáctica*, *29*(2), 176–184. https://doi.org/10.1016/j.psicoe.2024.05.003

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, *98*(4), 732–749. https://doi.org/10.1037/0022-0663.98.4.732

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. https://doi.org/10.1119/1.18809

Hartati, M., & Hanafi, A. H. (2022). Maharah al-Qira'ah Learning Planning and Strategies in Improving Islamic Understanding. *Ruhama: Islamic Education Journal*, *5*(2), 75–86. https://doi.org/10.31869/ruhama.v5i2.3644

Hasanah, N. (2020). Efektivitas strategi tutor sebaya dalam meningkatkan kemampuan membaca (qirā'ah) mahasiswa. *Shaut Al Arabiyyah*, *8*(2), 101–112. https://doi.org/10.24252/saa.v8i2.15142

Hattie, J. A. C. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge. https://doi.org/10.1007/s11159-011-9198-8

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

Khuluq, M., & Imamah, N. (2024). Enhancing Arabic Reading Skills: A Prezi Program-Based Learning Material Development for Integrated Islamic Elementary Schools in Indonesia. *Journal of Arabic Language Learning and Teaching (JALLT)*, *2*(1), 33–58. https://doi.org/10.23971/jallt.v2i1.157

Mustakim, N., Setyawan, C. E., Hunaenah, H., & Masuwd, M. (2024). CEFR-Based Maharah Qiraah Learning Interactive Test Using the Quizziz App. *Alsuna: Journal of Arabic and English Language*, *7*(1), 145–160. https://doi.org/10.31538/alsuna.v7i1.5086

Nuruddin, Emzir, & Tajuddin, S. (2017). The Need of Culture-Based Arabic Reading Materials by Using Multimedia in Higher Education. *Alsinatuna: Journal of Arabic Linguistics and Education*, *3*(1), 84–96. https://doi.org/10.28918/alsinatuna.v3i1.6762

Rizal, M., Farisi, M. Z. Al, & Sopian, A. (2023). Peer Teaching Method Integrated with Card Media: Its Implementation in Arabic Syntax Learning. *Arabiyatuna: Jurnal Bahasa Arab*, *7*(1 May), 345–362. https://doi.org/10.29240/jba.v7i1.5860

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*(2), 240–257. https://doi.org/10.1037/0022-0663.95.2.240

Romero, M. E., Burns, M. K., Wan, H. T., Abdelnaby, H., & Crawford, T. J. (2025). A Meta-Analysis on the Effects of Peer Tutoring on Emergent Bilinguals' Academic Achievement. *Psychology in the Schools*, *62*(6), 1953–1965. https://doi.org/10.1002/PITS.23443

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding Tutor Learning: Knowledge-building and Knowledge-telling in Peer Tutors' Explanations and Questions. *Review of Educational Research*, *77*(4), 534–574. https://doi.org/10.3102/0034654307309920

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin. https://iaes.cgiar.org/sites/default/files/pdf/147.pdf

Susanto, Ritonga, A. W., Desrani, A., & Rubiyantoro, Y. (2022). Arabic Learning Design Based on 21st Century Skills during the Covid-19 Pandemic in Indonesia. *Jurnal Iqra': Kajian Ilmu Pendidikan*, *7*(2), 1–14. https://doi.org/10.25217/ji.v7i2.2235

ISSN (p) 2828-5603
ISSN (e) 3026-1511

Taufiq, A., Annur, A. F., Rif'iyati, D., & Hasanah, F. N. (2022). Peer Tutor Method Integrated with Short Card Media in Arabic Learning at Madrasah Ibtidaiyah. *Alsinatuna: Journal of Linguistics and Education*, *7*(2), 160–173. https://doi.org/10.28918/alsinatuna.v7i2.5046

Topping, K. J. (2005). Trends in Peer Learning. *Educational Psychology*, *25*(6), 631–645. https://doi.org/10.1080/01443410500345172

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Webb, N. M. (1989). Peer Interaction and Learning in Small Groups. *International Journal of Educational Research*, *13*(1), 21–39. https://doi.org/10.1016/0883-0355(89)90014-1

Zhang, C., Sun, N., Jiang, Y., Liu, H., & Huang, Q. (2025). The Impact of Peer Tutoring Programs on Students' Academic Performance in Higher Education: A Meta-analysis. *Asia-Pacific Education Researcher*, *34*(4), 1495–1506. https://doi.org/10.1007/S40299-024-00960-0