

AI-Mediated Discourse in Arabic Grammar Acquisition: Linguistic Interaction Patterns in ChatGPT-Assisted *Nahwu* Learning

Tamim Mulloh ^[1], M. Abdul Hamid ^[2], Ahmad Mubaligh ^[3]

^[1] tamim@bsa.uin-malang.ac.id

^[1], ^[2], ^[3] Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia

First Received: 09/04/2026; Final Proof Received: 26/06/2026; Accepted: 27/06/2026

Abstract

This paper analyzes the learning of *nahwu* with the assistance of ChatGPT at an Indonesian Islamic university, where there is a perception that classical Arabic grammar is not merely a linguistic subject but also a doorway to the scholarship of Islamic texts. The urgency of this research is that *nahwu* learning is often based on memorization and teacher-centered explanations, whereas engaging in conversations with AI may give learners more opportunities to question, engage in interaction focused on form, and conduct metalinguistic reasoning. This study involved thirty-four bachelor's degree students from Maulana Malik Ibrahim State Islamic University and employed an explanatory sequential mixed-methods design. 17 students were placed in a conventional *nahwu* class, while the rest were placed in a class that integrates ChatGPT. This study lasts six weeks, and its intervention encompasses structured prompt procedures, guided classroom tasks, pre- and post-tests, weekly interaction logs, ChatGPT dialogue records, and classroom observations. Apart from that, the authors selected eight participants through purposive sampling and conducted semi-structured interviews. The authors analyzed quantitative data through independent-sample tests, paired-samples tests, a mixed-design sensitivity analysis, and effect sizes. Meanwhile, the qualitative data were analyzed thematically using codes derived from the metalinguistic output of the Interaction Hypothesis, Sociocultural Theory, and the Community of Inquiry. Results show that both groups improve, with the ChatGPT group achieving greater gains. The post-test difference was significant: $t(32) = 6.72$, $p < 0.001$, with a large between-group effect size, $d = 2.31$. Three patterns can be observed in the qualitative findings: more precise metalinguistic discourse, the repositioning of the instructor as a verifier and facilitator, and dialogic inquiry. Results show that ChatGPT usage is effective for Arabic grammar pedagogy as it functions beyond an answer generator: it creates an interactional environment which can aid learners in becoming more initiative. Even so, the AI output still requires human monitoring to guarantee accuracy.

Keywords: AI-mediated discourse; Arabic grammar; ChatGPT; community of Inquiry; *nahwu*; second language acquisition

INTRODUCTION

Nahwu (the study of Arabic syntax and grammar) occupies a unique and challenging position among the persistent problems that Arabic language education faces in Islamic universities in Indonesia. *Nahwu* is not a marginal parlor game but rather an integral intellectual instrument that helps students learn to analyze, understand, and decode classical texts in Arabic. Students' understanding of verbal morphology, agreement, nominal predication, *I'rab*, and the relationships between clauses enables them to understand how syntactic roles create meanings, and why certain words have certain grammatical endings. These matters have additional weight in Islamic universities as the interpretation of grammar often relates to students' ability to access classical Islamic texts. A student who fails to understand the relationship between *Khabar* and *mubtada'*, or

between *ma'mul* and *'amil*, will be limited in their participation in the longer tradition of Arabic textual analysis. Thus, it cannot be deemed that the student is merely making an isolated grammatical error (Gass & Mackey, 2015).

Despite its importance, learning *nahwu* is often described as abstract-heavy and procedure-intensive. Previous work on Arabic pedagogy in Indonesia has highlighted issues of weak involvement, uneven prior exposure to *pesantren* (Islamic boarding school) or *madrasah* (Islamic school) learning, limited time for instruction, and reliance on teacher-centered explanations or grammar-translation (Aziz, 2020; Kamal, 2025). In such classrooms, a learner may have memorized what they are told is necessary to obtain the right answer, but fail to apply it when the sentence order changes. They may also become confused when more than one grammatical relation occurs. The challenge thus is both cognitive and discursive: learners have to be offered opportunities not just to articulate hypotheses, test examples, ask why a particular case ending is justified, and to receive feedback that moves them beyond rule-naming toward grammatical reasoning.

Given the strong research base behind interactional approaches to second language acquisition, it is no surprise that many researchers and practitioners believe grammar develops as part of an interpersonal process (Gass & Mackey, 2015; Long, 1996). The Interaction Hypothesis (Long, 1996) states that learners will benefit from language development if they negotiate meaning and form (Godwin-Jones, 2022). They should also notice gaps in their knowledge regarding their learning process during communicative exchange. What this means is that, in grammar learning, language acquisition becomes more effective when learners are not simply asked to listen to rules, but also ask questions for clarification, compare examples, and reformulate erroneous explanations by reformulating grammatical areas with corrective feedback. This view is complemented by sociocultural theory, which considers mediation, scaffolding, and internalization (Lantolf & Thorne, 2006; Vygotsky, 1978). Then, from this point of view, learners' grammatical knowledge is developed through mediated practice with teachers, peers, texts, and tools. The relevant mediation in *nahwu* classrooms is highly metalinguistic because students learn how to speak and think with technical categories (like *i'rab*, *'amil*, *mansub*, *marfu'*, *jar--mubtada' khabar fi'l*).

Within this context, conversational AI, particularly ChatGPT, represents a fresh type of mediational tool (Cope & Kalantzis, 2022). ChatGPT can answer students' questions over multiple turns, provide alternative examples of the grammar in action, explain differences between learners' expressed concepts, and even continue a back-and-forth conversation if students are confused (Huang et al., 2023). Recent comprehensive surveys of ChatGPT-motivated work in foreign language education have similarly demonstrated exponential growth since 2022. However, the continued predominance of writing about English writing or vocabulary (or perceptions) over fine-grained grammatical discourse is also consistent across these reviews (Fang & Han, 2025; Law, 2024). Systematic reviews of chatbot-based language education also cite largely favorable effects on skills, attitudes, autonomy, and engagement, while noting caveats regarding accuracy, overdependence, and pedagogical design (Cisłowska & Pena-Acuna, 2024; Lyu et al., 2024). These findings are not fully transferable to classical Arabic grammar, as *nahwu* features an idiosyncratic vocabulary, a heritage of authority, and a high sensitivity to even small grammatical distinctions (Cohen, 1988).

There is a growing but uneven body of research directly related to ChatGPT and grammar learning. For example, Kucuk (2024) showed that a ChatGPT-integrated grammar instruction in the EFL context led to better gains than teacher- and textbook-centered teaching methods, while Azwar and Sujarwati (2025) reported positive effects of ChatGPT usage on Indonesian EFL learners' grammar achievement (Hubbard, 2013). These research articles strengthen the case for examining ChatGPT in grammar pedagogy. However, they are primarily focused on English-language learning and often emphasize achievement rather than the interactional processes through which achievement develops. Nasaruddin's research (2024), for example, explained how teachers of Arabic as a foreign language may use ChatGPT to prepare teaching materials, questions, and texts. Another study compared the use of Gemini and ChatGPT for analyzing *i'rab* (Maulidiya et al., 2025), stating that although AI for language learning is useful, its output still requires human correction and relevance checks. There are many benefits of studies that focus on Arabic roots, but it leaves the empirical question of how students can use ChatGPT to acquire *nahwu*.

A further limitation of the existing literature is methodological. A considerable number of research studies describe learners' perceptions towards ChatGPT or report post-intervention score differences. However, only a few have combined score data with interaction logs, classroom observations, and interviews. For grammar teaching, this combination is essential because the object of learning is not only the correct answer but also the process of grammatical justification. A learner, an exemplar-based one, can produce a correct case label by guessing, copying, or even memorizing it. However, this learner would not be able to explain the syntactic relation that makes the label valid (Norris & Ortega, 2000). On the other hand, a learner who struggles with the final label, as they continue to ask increasingly precise questions, demonstrates that they are developing mastery. The design of the present study centers on that distinction, as discourse data were conceived as essential evidence rather than merely exemplary illustration (Anderson et al., 2025).

Therefore, the gap that this research addresses is quite specific. However, existing scholarship has failed to provide adequate empirical evidence on how conversational AI shapes learner discourse, metalinguistic reasoning processes, and class participation in *nahwu* learning in Islamic universities. This gap is not just whether ChatGPT raises scores (Kessler, 2018). In addition, it examines how students encounter grammatical problems with the assistance of an AI interlocutor, how Arabic grammatical terminology is used in dialogue and classroom discourse when students bring AI-mediated questions to the instructor (Nassaji & Fotos, 2011). This is an important distinction, as a tool can increase test scores through remedial practice, while classroom conversation remains essentially unchanged. This distinction matters because a tool may raise test scores through extra practice while leaving classroom discourse essentially unchanged. On the contrary, the tool may transform learner initiative and metalinguistic talk even when the effect on outcomes is modest. A rigorous account must therefore examine both competence and discourse (Al-Khatib, 2016).

In this study, ChatGPT is addressed not as an autonomous teacher or merely a source to be consulted, but as a discourse partner within the web of pedagogical interactions. This study investigated two research questions: (1) What is the effect of ChatGPT-assisted interaction on learners' grammatical competence in classical Arabic *nahwu*? (2) What discursive patterns and interactional shifts define learner engagement in *nahwu* acquisition contexts with ChatGPT? The study aims to explain how AI-assisted interaction operates in a grammar classroom in which linguistic accuracy, scholarly tradition, and learner participation are tightly interwoven, combining test-based evidence with observations, interviews, and ChatGPT interaction logs (Schegloff, 2007).

Theoretical framework and analytical operationalization

Three theories were used in this paper. In this study, Long's Interaction Hypothesis Framework was used. Interaction was defined as learner participation in the negotiation of form and meaning. The code was based on the following: it lists all possible clarification responses generated by ChatGPT (requests, contrastive examples, uptake after feedback, and confirmation checks) [4]. Classroom observations were also analyzed to determine whether learners asked analytical questions (initiated) rather than simply answering the teacher's questions. To guide this interpretation, the authors used the model to link additional form-focused interaction with competence gains for RQ1 and identify the discourse moves through which learners made explicit attention to specific grammatical forms for discussion as per RQ2.

The second framework is the Vygotskian Sociocultural Theory. Its purpose was to ground ChatGPT in a system of mediating activity rather than to treat it as a mere transmission device. Thus, in the analysis, the authors identified scaffolding, assisted performance, the shift from other-regulation to self-regulation, and the transfer of AI-supported reasoning into classroom discourse. In *nahwu* learning, scaffolding traces can be observed when learners break down analytical steps of a grammatical task through ChatGPT, request examples with varying levels of difficulty using ChatGPT, or compare potential analyses in light of the explanation before asking the lecturer for verification. The important analytical question was not whether ChatGPT provided an answer, but whether and to what extent the interaction led learners to generate more explicit and regulated grammatical reasoning (Sinclair & Coulthard, 1975).

The last framework is the Community of Inquiry (CoI) model. CoI originally explains educational experience in terms of cognitive, social, and teaching presence (Garrison et al., 2000). For this study, cognitive presence was defined as the progression of learners from identifying a grammatical issue to finding instances of target-language rules and ultimately arriving at a solution through analysis. Social presence is the extent to which learners feel they can ask questions without embarrassment and advance ideas that were in some way incomplete by having them discussed or commented upon. Teaching presence included the design, facilitation, and verification functions performed by the instructor. All recent work on AI and CoI suggests that generative AI has a role in inquiry when it supports instructional design, guides learners, participates in discussion, and sustains dialogical partnerships while keeping human agency and critical reflection in the foreground (Stenbom & Garrison, 2026).

Instead, in this context, AI-mediated presence can be thought of as an analytic extension of CoI whilst remaining cautious of replacing its three canonical presences. It refers to the perceived and observable contribution of a conversational AI system to the persistence, responsiveness, and direction of inquiry. It differs from teaching presence in that it does not design the course, make moral judgments, or take institutional responsibility for right vs. wrong. Unlike cognitive presence, however, it is not the construction of meaning by the learner itself that it achieves; rather, a new concept may stimulate such construction with prompts, feedback, and examples. It differs from social presence in that it does not indicate the identity and relational commitment of a human participant, but its non-judgmental reactivity may lower learner anxiety. In practical coding, this refers to instances of AI-mediated presence when students turn to ChatGPT, a constant, responsive, and iterative interlocutor, to design activities that maintain the galaxy of grammatical inquiry alive for them outside of teacher-fronted classroom time (Warschauer & Healey, 1998).

Swain's Output Hypothesis was introduced as a new construct within the theoretical framework. It is not treated as a fourth overarching framework that is equal to the Interaction Hypothesis, Sociocultural Theory, and CoI. Instead, it accounts for one mechanism: language production can prompt learners to notice an absence of knowledge, test hypotheses, and process structure more deeply (Swain, 1985, 1995). In this analysis, what was relevant was not conversational fluency but metalinguistic output: learner-generated explanations, questions, counterexamples, and justifications using *nahwu* terminology (see Table 1).

Table 1. Operationalization of theoretical frameworks in the analysis

Framework/ construct	Analytical focus	Observed indicators	Role in interpretation
Interaction Hypothesis	Negotiation of form and meaning during exchange.	Clarification request; contrastive examples; feedback uptake; self-repair; multi-turn grammatical inquiry.	Coding of ChatGPT logs and classroom discourse; interpretation of competence gains as interaction-supported rather than tool-determined
Sociocultural Theory	Mediation, scaffolding, ZPD, internalization.	Learner requests a step-by-step explanation, assisted analysis, and movement from AI-supported reasoning to classroom questions.	Interpretation of ChatGPT as a mediational artifact and of the lecturer as verifier/facilitator
Community of Inquiry	Cognitive, social, and teaching presence in educational inquiry.	Triggering grammatical problems; exploration of examples; integration of rules; resolution; safe participation; instructor design.	Coding of inquiry phases and classroom roles
AI-mediated presence	Distinct AI contribution to continuity and responsiveness of inquiry.	Always-available dialogue; adaptive response; non-judgmental repair; iterative prompts outside of class.	Explanation of how ChatGPT extends inquiry without replacing teaching presence
Output Hypothesis	Metalinguistic production as pushed output.	Learner explanation; hypothesis testing; explicit use of terms such	Interpretation of metalinguistic register and learner-generated analysis

as i'rab, 'amil, mansub,
marfu'

Source: Author's analytical framework based on the study design

METHOD

This study used an explanatory sequential mixed-methods design (Creswell & Plano Clark, 2018). Quantitative data were initially obtained using pre- and post-tests to compare the *gains in nahwu* competence of students in the ChatGPT-integrated class with those in the conventional class. The quantitative data pattern was further detailed by the collection and analysis of qualitative data to explain the discourse processes behind these findings. This design suited the empirical objective well, for it was not merely a ranking of the two ways that learners can interact with grammar. It sought to associate competence outcomes with the interactional mechanisms through which learners engaged with grammar (Yin, 2018).

The research site was an Arabic Language and Literature program in an Islamic University in Indonesia. Such an institutional setting was analytically crucial, as *nahwu* is connected to both the study of the Arabic language and the interpretation of Islamic texts in the larger sense. The subjects in this study were 34 undergraduate students enrolled in a *nahwu* course. Two intact classes were used: an experimental group receiving ChatGPT-integrated instruction (n=17) and a control group receiving conventional instruction (n=17). Because intact classes were used, the study was considered quasi-experimental rather than fully randomized. Pre-test equivalence was assessed before the intervention, and the independent-samples results indicated no statistically significant initial difference, $t(32) = 0.22$, $p = 0.830$.

The intervention was experimental and continued for six weeks. Both groups shared the same *nahwu* content: case and *i'rab* assignment, selected verb morphology, *mubtada'* and khabar structures, and basic subordinate clause patterns. To minimize variations due to the teacher effect, both groups were taught by the same instructor. The control group was taught via lecture as a form of traditional instruction, analysis of examples, question-answer sequences, and homework problems from textbooks. The lesson content is identical to that of the control group but also incorporates ChatGPT as a structured, dialogic tool for grammar exploration. ChatGPT was integrated in two ways: short classroom demonstrations by teachers and out-of-class use by students with weekly logs.

For transparency and reproducibility, the intervention utilized the public-facing ChatGPT web interface. At the time of data collection (i.e., November–December 2025), the interface showed that students used GPT-4o/default ChatGPT as the model. Because AI systems change over time, this research archived interaction dates, prompt templates, teacher verification notes, and sampled outputs rather than only depending on a model identity as the basis for replication.

Experimental group participants were guided through a prompt procedure comprising four steps: state the grammatical problem; provide the Arabic sentence or phrase; ask for step-by-step reasoning using *nahwu* terminology; and request one or more contrasting examples. A common query format was: 'examine the *i'rab* of the underlined word in this sentence; describe the '*amil*, sign of case, and why the other case is not possible; and then generate one similar and one contrasting example.' Students were guided to converse in Indonesian for explanations, and Arabic examples were added in the prompt. They may even use ChatGPT to provide Arabic terms and transliteration. The dominant use of Indonesian during this time enabled teachers to provide high levels of pedagogical explanation, which is crucial when they are required to communicate complex ideas. This bilingual process mirrored the realities of the classroom, where Indonesian emerges as the dominant language of explanation, whilst Arabic assumed dual roles: first as a target (object) language and secondly as an important source for technical grammatical terminology (Zawacki-Richter, 2019).

To minimize interference with natural inquiry, usage frequency was standardized as much as possible. Experimental students performed two ChatGPT-guided tasks weekly and at least one independent inquiry log per week. For each log, the authors recorded the initial prompt, at least one follow-up question, the AI answer, whether students found it clear (or unclear), and a brief reflection on what to check with the lecturer or in the textbooks. Students were also told not to

assume ChatGPT output is correct by default. They had to compare the AI explanations with those in the course materials and be prepared to raise any answers that were unclear or raised red flags in class. This verification process requires students to mitigate the risk of erroneous AI grammar explanations being internalized.

A purpose-constructed 40-item test was used to measure grammatical competence across four domains, namely case assignment (*i'rab*), verb morphology and conjugation, nominal predication structures, and subordinate clause formation. The combination consisted of recognition, short-answer analysis, and very brief production items. Content validity was examined by three Arabic linguistics specialists, who assessed coverage, item difficulty, and relevance to the six-week syllabus. Reliability was acceptable, with Cronbach's alpha of 0.87. The same test blueprint was used for pre- and post-tests, although the item order and surface examples were varied to avoid memory effects.

The qualitative data were collected from three bases: classroom observations, ChatGPT interaction logs, and semi-structured interviews. The authors collected observations over six sessions per group, using a structured protocol to document student-initiated turns (i.e., turns where students initiated a question), types of questions, turn length, teacher evaluation moves, instances of metalinguistic reasoning, as well as instances when students made use of or pushed back on AI-generated explanations. Only the experimental group provided ChatGPT logs, which were anonymized prior to the analysis. The logs were used to explore prompt structure, follow-up behavior, repair sequences, and the emergence of metalinguistic discourse. After the intervention, the authors conducted interviews to understand how students interpreted their own learning process.

Maximum variation sampling was used to select the eight interviewees purposively. Four students were chosen from the experimental and the control group, respectively. In each group, students were selected to represent relatively high- and moderate-gain students, as well as more- and less-active participants in classroom tasks. Selection in the experimental group also included the richness of ChatGPT log submissions to link interview data to observable interactional behavior. The rationale for this sampling strategy was that the interviews were not intended to be representative in a numerical sense but rather to obtain explanatory depth. This was because the study required opinions explaining why some students used interactional opportunities more analytically than others.

Quantitative data were evaluated using descriptive statistics; paired-samples t-tests for within-group gains; independent-samples t-tests for comparisons between groups; associated effect sizes (Cohen's *d*); and a mixed-design sensitivity analysis of the overall pattern of mean changes across groups by time. Independent-sample statistics were computed from the reported means, standard deviations, and sample sizes. The t-test comparison of the mean score on the low-calorie version between groups at pre-test was not statistically significant, $t(32) = 0.22$, $p = .830$. The post-comparison test was significant, $t(32) = 6.72$, $p < .001$. Paired effect sizes were reported as *d_z*, while the between-group post-test effect size was reported as Cohen's *d*.

Qualitative analysis was carried out following Braun and Clarke's (2006) thematic approach, but was operationalized through a theory-informed codebook. In the first cycle, open coding was employed to identify discourse moves from logs, observations, and interviews. The second cycle grouped these moves into categories based on the Interaction Hypothesis, Sociocultural Theory, CoI, and metalinguistic output. The third cycle formed themes that explored the answer to RQ2. To strengthen the audit trail, each theme was supported by at least two data sources. After two weeks, the codes were reviewed through code-recode and peer discussion with an Arabic linguistics colleague. The study did not claim formal inter-rater reliability; rather, credibility was grounded on clear coding categories, source triangulation, and representative excerpts.

Ethical procedures included voluntary participation, anonymization of student names, and removal of personally identifiable information from ChatGPT logs before the analysis. It was explained to students that the use of ChatGPT was part of the research intervention, but that their grades would not be affected by their decision to give positive or negative evaluations of the AI tool. The instructor manually checked AI outputs via weekly sampling of submitted logs and classroom verification of problematic explanations. In cases where ChatGPT analysis was

questionable or incomplete, the instructor used it as a teaching object, asking students to identify the relevant rules and compare the AI explanation with course references. This process embodies a core pedagogical position of the study, that conversational AI can assist inquiry, but grammatical authority remains distributed between learner verification, disciplinary knowledge, and teacher expertise.

ANALYSIS

ChatGPT-mediated interaction and nahwu competence

Prior to the intervention, the control and experimental groups had similar levels of *nahwu* competence. The mean pre-test score for the experimental group was 62.4 (SD = 8.3), while the mean for the control group was 61.8 (SD = 7.9). Descriptive statistics showed that this difference was not statistically significant, $t(32) = 0.22$, $p = .830$, 95% CI [-5.06, 6.26]. This result confirms baseline comparability and demonstrates why statistical values must be aligned with sample sizes, means, and standard deviations.

After six weeks, both groups showed improvement, but the gain was greater in the ChatGPT-integrated class. The experimental group increased from 62.4 to 87.6, i.e., reaching a mean gain of 25.2 points. The control group increased from 61.8 to 71.3, with a mean gain of 9.5 points. Within-group paired comparisons remained significant for both groups: experimental, $t(16) = 12.47$, $p < .001$, $d_z = 3.02$; control, $t(16) = 4.83$, $p < .001$, $d_z = 1.17$. The post-test between-group comparison showed a significant advantage for the experimental group, $t(32) = 6.72$, $p < .001$, 95% CI [11.36, 21.24], with a large between-group effect, $d = 2.31$. A change-score sensitivity check also indicated a significant group-by-time pattern, $t(32) = 5.57$, $p < .001$, equivalent to $F(1, 32) = 31.00$. These results support the interpretation that interaction with an integrated use of ChatGPT was associated with greater short-term gains. Meanwhile, the quasi-experimental design prevents strong causal claims.

Table 2. Descriptive statistics and within-group gains

Group	n	Pre-test M (SD)	Post-test M (SD)	Gain	Paired t(16)	p	d _z
Experimental (ChatGPT)	17	62.4 (8.3)	87.6 (5.1)	+25.2	12.47	< .001	3.02
Control (Conventional)	17	61.8 (7.9)	71.3 (8.6)	+9.5	4.83	< .001	1.17

Source: The researchers' raw data.

Table 3. The comparison between-group and change-score tests

Comparison	Mean difference	t	df	p	95% CI	Interpretation
Pre-test between groups	0.6	0.22	32	.830	[-5.06, 6.26]	No significant baseline difference
Post-test between groups	16.3	6.72	32	< .001	[11.36, 21.24]	The ChatGPT group shows a large advantage; $d = 2.31$
Gain difference/ group x time	15.7	5.57	32	< .001	[9.96, 21.44]	Sensitivity check supports larger experimental gain; $F(1,32) = 31.00$

Source: The researchers' raw data.

The domain-level results provide a more specific explanation of the concentration of differences. The greatest advantages of the experimental group were evident in subordinate clause

structures and case assignment. These two domains require relational analysis rather than simple recall: learners must identify clause boundaries, governors, the grammatical consequences of particles or verbs, and syntactic roles. Such a pattern is consistent with the intervention's interactional logic. It was shown that ChatGPT was most useful when students used it to test why a particular relation produced a particular case ending, not when they used it only to memorize definitions.

Table 4. Post-test subcategory scores by grammatical domain

Grammatical category	Experimental M	Control M	p-value	Interpretive note
Case Assignment (<i>i'rab</i>)	89.2	69.4	< .001	High demand for relational analysis of 'amil, case sign, and syntactic role
Verb Morphology	85.7	72.1	< .001	Improved through repeated examples, but less dependent on multi-turn reasoning
Nominal Agreement (<i>mubtada'-khabar</i>)	86.3	71.8	< .001	Supported by contrastive examples and an explanation of agreement
Subordinate Clause Structures	88.9	70.6	< .001	Strong advantage where learners needed to identify embedded relations

Source: The researchers' raw data.

Discourse patterns in AI-mediated nahwu interaction

The findings reveal how students' discourse changed when ChatGPT became part of the learning ecology. Three themes were developed: a movement from monologic reception to dialogic inquiry, the emergence of a more explicit metalinguistic register, and the repositioning of the instructor from sole authority to facilitator and verifier. The themes were not based on interviews alone. They drew on classroom observation notes, anonymized ChatGPT logs, and interview excerpts, with the coding structure summarized below (Table 5).

Table 5. Qualitative coding categories and evidence sources

Theme/category	Operational definition	Sample indicators	Theoretical link	Data source
Dialogic inquiry	Learner initiates multi-turn exploration of a grammatical problem rather than requesting a single answer	Clarification requests, counterexamples, follow-up prompts, and self-repair	Interaction Hypothesis; CoI cognitive presence	ChatGPT logs; classroom observations
Metalinguistic register	Learner uses nahwu terminology to justify, compare, or question the	The use of <i>i'rab</i> , 'amil, mansub, marfu', mubtada', khabar in Analysis, Reasoning Frames,	Output Hypothesis; Sociocultural Theory	Logs; interviews; observations
Scaffolded reasoning	AI or the instructor breaks a grammar problem into steps that the learner later reuses	Step-by-step analysis; request for simpler example; movement from assisted to independent explanation	Sociocultural Theory	Logs; interviews
AI-mediated presence	Learner treats ChatGPT as an available interlocutor that keeps the inquiry going outside of class	Repeated prompts; non-judgmental repair; learner reports reduced embarrassment	Community of Inquiry extension	Logs; interviews

Instructor as verifier	instructor evaluates, corrects or extends AI-mediated analysis rather than initiating all discourse	Student brings AI explanation to class; teacher asks for rule justification; class compares alternatives	CoI teaching presence	Classroom observations
------------------------	---	--	-----------------------	------------------------

Source: Interview data, classroom observations, and ChatGPT logs.

The initial focus is the transition from monologic reception to dialogic exploration. Discourse in the control group classroom primarily revolved around initiation-response-feedback: the instructor posed a question, the student answered briefly, and the instructor provided feedback on that response before continuing the class. Observation notes showed that student-initiated turns accounted for only 11.4% of total classroom discourse in the conventional class. These turns were typically brief, transactional exchanges — whether an answer should be memorized or whether a particular form would appear on the test. While such participation is not without meaning, it provides limited evidence of how learners negotiate grammatical form.

However, the same pattern did not emerge from the experimental group. The logs from ChatGPT indicated that learners would manage a grammatical inquiry for multiple turns (e.g., an average of 6.3 turns before a given problem was designated as solved in the log reflection). For example, one exchange involved a student asking why the noun in (1) *inna al-taliba mujtahidun mansub*, whereas the predicate remains *marfu'*. The student's follow-up was as follows: If it is *kana*, does the effect not shift to *khabar*? It is analytically important because the learner did not stop after receiving a label. Using ChatGPT's explanation, the learner generated a contrast between two grammatical operators in Arabic — *inna* and *kana* — and tested how changing the governing relation influences the case pattern.

A similar transfer appeared in classroom observations. Once, during a session while the researcher was experimenting with this material, a student entered and said, "ChatGPT says that this word is *mansub* because it is *maf'ul bihi* and it's not, but rather follows the previous noun as *na't*." Which relation should we test first?' The instructor did not just say that the answer was incorrect; rather, the instructor asked the student to find a possible head noun, the verb, and finally, a semantic role. This moment highlights a significant shift in the interaction. The student no longer only asks for the right answer. Although it was a suggestion to test grammar, the instructor's response became facilitative rather than solely evaluative.

The second topic is the creation of a higher metalinguistic discourse register. For the control group, metalinguistic terms were largely used as recalled labels. When asked, students could almost always name terms like *fa'il*, *maf'ul*, *mubtada'*, or *khabar*; but they often did very little with those terms to construct an explanation. In contrast, a higher proportion of the terms appeared in analytical frames in the experimental group: students asked about the relation between an *'amil* and a case ending, compared the effects of particles, or requested examples that differentiated grammatical categories, compared effects of particles, or requested examples that would differentiate grammatical categories. This pattern corroborates with the view that ChatGPT-mediated interaction prompted learners to produce metalinguistic output rather than merely recognize terminology.

This shift in approach was reflected by one interviewee who explained, 'When I asked ChatGPT what *i'rab* was, it forced me to think about why the case changes and not merely what the case is. So I asked: Is it the verb which governs the noun, or is there another *amil* responsible?' It is a significant comment, as it reflects a shift from declarative recall to relational reasoning. In this case, the use of *'amil* is not as a rote definition, but is an explanation of how grammatical authority plays out in a sentence. Swain's framework explains this kind of metalinguistic output: the learner formulated a specific question to draw attention to gaps in understanding.

The third theme is the repositioning of the instructor as both a verifier and facilitator. As a result, in *nahwu* classrooms, the instructor becomes the center source of grammatical knowledge, the judge of correctness, and the controller of turn-taking. The experiment class did not abolish this authority, nor should it do so. Instead, it redistributed the sequence of inquiry. Students often start a grammatical problem with ChatGPT outside of class, encounter either a partial answer or a

confusing explanation, and bring that half-finished question to the instructor. Thus, the instructor simply verified, corrected, contextualized, and added to the analysis. This pattern might be more appropriately described as the reorganization of teaching presence than as the teacher's replacement.

This differentiation matters within the context of Islamic universities. *Nahwu* is historically linked with the transmission of scholarship, teacher authority, and disciplined interpretation of sacred texts. Any assertion that ChatGPT merely 'reconfigures authority' would be extremely difficult without ethnographic evidence. The data lends itself to a more modest assertion: that ChatGPT generated new entry points into grammatical inquiry, and these entry points occasionally provided students with better opportunities to engage the instructor with well-formed questions. Human expertise is still critical, especially for outputs that are inaccurately classified as incomplete or overly general, and AI still lacks classical grammatical sensitivity. So, AI-mediated presence acts as an extension of inquiry, and the teaching presence continued to provide a framework for curricular direction and accuracy checking.

Also, the qualitative data show that using ChatGPT carries risks. According to some students, the AI's confident tone gave off an impression that it was reliable, even when its explanations were incorrect or incomplete. Some realized that when they posed general prompts, the responses were broad explanations that still failed to address the real *nahwu* problem. These limitations confirm previous warnings that generative AI may reduce barriers to explanation but also introduce new challenges to verification (Kasneci et al., 2023; Maulidiya et al., 2025; Stenbom & Garrison, 2026). In the current study, the most productive interactions occurred when learners used structured prompts that prompted them to ask for justification. Also, they treated AI output as a hypothesis to be checked rather than as an authority to be copied.

The quantitative and qualitative findings indicate that achieving significant learning requires more than mere exposure to additional information. The same explanations and examples were given to the control group. The distinctive feature of the experimental condition was a greater opportunity for iterative, learner-initiated, form-focused dialogue. This interpretation is consistent with the Interaction Hypothesis, as the ChatGPT logs included requests for clarification, counterexamples, and repair sequences. It is consistent with Sociocultural Theory, in which ChatGPT served as a mediational artifact, scaffolding learners' reasoning both before and after classroom instruction. It is consistent with CoI because it extended cognitive presence beyond class time, reduced social risk for some students, and shifted teaching presence from verification to facilitation.

Moreover, the results are generally consistent with recent studies on grammar learning in conjunction with the use of ChatGPT. However, there was an additional context-specific contribution. Similarly, as was the case with Kucuk (2024) and Azwar and Sujarwati (2025), this study reported superior grammar achievement in the group that integrates ChatGPT usage compared to the conventional one. However, this study builds on the literature by targeting classical Arabic *nahwu* and analyzing discourse patterns and perceptions. Apart from that, it also serves as an extension of Nasaruddin's research (2024) and his talk on the role of ChatGPT in Arabic teaching by demonstrating how learners and teachers use conversational AI for grammar-related inquiries. This study also supports the research of Maulidiya et al. (2025), which raises the warning that the AI-generated *i'rab* analysis needs to be manually verified, because the strongest pedagogical value appears when students are trained to question and verify AI output.

Hence, the novelty of the current study is very specific. Provides mixed-methods evidence that, in an Islamic higher education *nahwu* course, the structured use of ChatGPT was linked to greater short-term competence gains and to observable changes in learner discourse. It also operationalizes this new dimension of AI-mediated presence as a coding construct to understand how conversational AI contributes to inquiry without conflating it with human teaching presence.

Several implications follow. ChatGPT could be used more effectively if it were incorporated into Arabic grammar pedagogy as an application of prompt literacy and verification routines, rather than letting it remain as a new-age homework shortcut. These include training in how to seek *i'rab* analysis, how to require grammatical justification, how to compare examples, and how to recognize when an AI response would benefit from teacher review. For instructors, the value of the

tool might not come from what it explains but rather how it shifts students' questions as they arrive at class. For researchers, this indicates that future work should combine achievement measures with discourse data, since test scores alone do not indicate if learners are becoming more active grammatical reasoners.

However, a critical reading of the quantitative effect is necessary because of the gain in the experimental group. One possible reason is that the ChatGPT group had more opportunities to provide retrieval, explanation, comparison, and feedback than the traditional group. Another explanation is that students use ChatGPT outside of class, which increases the time they spend on grammar. Third, students may have paid more attention since the tool was new and interactive. All these explanations are not mutually exclusive. That is why this statistical outcome should be seen not as evidence that ChatGPT directly caused these improvements, but rather as strong evidence of a correlation between structured use of ChatGPT and short-term *gains in nahwu*.

The lack of an active control group is also vitally important. The control condition represented conventional instruction, rather than an interactive alternative such as peer tutoring, human online feedback, or other non-AI digital exercises on grammar. As such, the study cannot indicate whether any benefit is attributable to ChatGPT per se or to a combination of interaction, autonomy, and additional practice time. This limitation does not invalidate the findings but does narrow the claim. The strongest conclusion that can be drawn is that ChatGPT can be didactically effective when used as a generative, form-focused dialogue and when a qualified instructor moderates its output.

Qualitative evidence explains why the tool may have been useful in this grammar domain. *Nahwu* problems usually take the form of a chain of reasoning: find the word, determine its role in the sentence, identify its governing element, determine the case marker, and demonstrate why other syntactic analyses are less optimal/have more violations. In teacher-dominated classrooms, there may not be enough time for students to practice each step before straining to memorize the larger task. Learners used ChatGPT interaction as a low-stakes context to rehearse this reasoning. The benefits, however, were conditional on prompt quality. On a broader level question, such as 'Explain *i'rab*', students provided an answer that felt rather generic. When students supplied a sentence, identified the problematic word, and asked for a contrastive explanation, their responses were more useful for grammatical reasoning.

This observation connects the current study to an emerging field of scholarship regarding prompt literacy within language education. Studies of ChatGPT-supported writing and grammar learning increasingly show that students do not automatically know how to collaborate productively with AI. Students require models for asking specific questions, analyzing answers, and treating AI output as new material for analysis rather than a finished product (Han, 2024; Kohnke et al., 2023; Law, 2024). For example, the prompt literacy in *nahwu* should be discipline-specific: Students should learn to ask for the '*amil*, grammar position, *harakah*, *sabab* (i.e., why), and *mukhālafah* sentence. Without this disciplinary prompt design, ChatGPT risks becoming a fluent but shallow explainer.

It also refines the concept of AI-mediated presence. The data do not suggest treating AI-mediated presence as an additional human-like participant in the classroom. Rather, AI-mediated presence identifies an affordance pattern: learners perceive the tool's responsiveness (the 24-7, non-judgemental sign that interaction is on demand) as a responsive listener to their inquiry. This facility can promote cognitive presence by assisting learners in traversing disorientation to exploration and initial resolution. It can indirectly boost social presence by reducing the embarrassment of asking basic questions. It also provides support when the instructor designs tasks and checks outputs. AI-mediated presence is both relational and pedagogically dependent; it becomes educationally meaningful only within a designed learning arrangement.

This is a particularly important point in the context of Islamic higher education. In this light, teaching Arabic grammar is not just technical training; it involves becoming part of a scholarly tradition where transmission is rigorous, analysis is careful, and authoritative knowledge commands respect. A poorly framed AI intervention may bypass that tradition. A better-framed intervention is to position ChatGPT as a space for rehearsed questioning before students return to the teacher, the textbook, and the grammar tradition for validation.

For curriculum design, these findings suggest a Hybrid Model of AI in *nahwu* learning. One is assigning topic-based writing prompts to support grammar. Second, learners should post brief interaction logs that include both the AI's response and their own assessments. Third, verification episodes (in which students compare ChatGPT output with authoritative grammar explanations) should be integrated into classroom time. Fourth, assessment should promote reasoning and justification instead of just final labels. These would employ visible, accountable, and transparent use of ChatGPT that is also aligned with the intellectual aims of how learner interaction is structured, monitored, and interpreted throughout the intervention.

The study calls for more rigorous designs for future empirical work. Increasing the sample size would allow ANCOVA with pre-test scores as covariates, resulting in more stable estimates of effect sizes. An active control group can isolate the effect of AI from the effect of additional practice. A usage-tracing protocol would measure time on task more accurately. This systematic accuracy audit would help identify which *nahwu* topics are reliable for ChatGPT and require stricter teacher intervention. Finally, discourse analysis should consider entire interaction sequences (not only selected parts), so that claims about learner initiative and classroom authority are based on a complete interactional record.

CONCLUSION

This study investigated what happens when undergraduate students in Islamic higher education learn Arabic *nahwu* with ChatGPT as a structured interlocutor. The results indicate a substantial relationship between the usage of ChatGPT-based learning and incremental short-term improvements in grammar knowledge. In addition, the most significant pedagogical change might concern discourse: learners in the experimental group asked more analytical questions and used *nahwu* terms in more relational ways, and brought partially developed grammatical inquiries into the classroom. These results indicate that ChatGPT can support Arabic grammar learning in a pedagogical design that values inquiry, verification, and teacher-guided interpretation.

Theoretically, the study contributes to the unique context of Arabic grammar by connecting the Interaction Hypothesis, Sociocultural Theory, CoI, and metalinguistic output. The Interaction Hypothesis explains why multi-turn form-focused exchange matters. Sociocultural Theory explains how ChatGPT can function as a mediational artifact. CoI outlines the interplay among cognitive, social, and teaching presences within a learning ecology and introduces a conversational AI tool. AI-mediated presence is timidly proposed as the visible contribution of AI agents to the continuity and responsiveness of inquiry rather than a replacement for teacher knowledge or learner cognition.

The Output Hypothesis explains the pedagogical significance of learner-generated metalinguistic explanations and questions. Since the current study used a small sample size, intact classes, a brief intervention period, and no active control group, it cannot establish broad causal generalization. Because out-of-class ChatGPT use was recorded rather than assigned experimentally, this substantial gain can only be attributed in part to exposure. The results should not be read as support for self-directed AI grammar tutoring, as AI output can be inaccurate. The most robust conclusion is that using a structured, vetted, and discussion-triggering ChatGPT can create an environment amenable to *nahwu* research and enhance grammar outcomes in the short term.

This interpretation should be tested in future research with larger samples, randomized or active-control designs, longitudinal follow-up, and systematic accuracy audits of AI-generated Arabic grammatical explanations. Researchers may also investigate differences in performance by model version on classical Arabic grammar, how students learn to challenge AI explanations, and how instructors respond when students bring analyses generated by an AI system.

REFERENCES

- Al-Khatib, M. A. (2016). Approaches to the teaching of Arabic grammar: A critical review. *Arab World English Journal*, 7(3), 180–197. <https://doi.org/10.24093/awej/vol7no3.13>
- Anderson, J. E., Nguyen, C. A., & Moreira, G. (2025). Generative AI-driven personalization of the Community of

- Inquiry model: Enhancing individualized learning experiences in digital classrooms. *The International Journal of Information and Learning Technology*, 42(3), 296-310. <https://doi.org/10.1108/IJILT-10-2024-0240>
- Aziz, A. (2020). Pembelajaran nahwu di perguruan tinggi: Tantangan dan inovasi metodologis. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaan*, 7(2), 157-170. <https://doi.org/10.15408/a.v7i2.15579>
- Azwar, T. A., & Sujarwati, I. (2025). Does ChatGPT-integrated learning affect Indonesian EFL learners' grammar achievement? *ELE Reviews: English Language Education Reviews*, 5(1), 1-27. <https://doi.org/10.22515/elereviews.v5i1.10425>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Cisłowska, A. I., & Pena-Acuna, B. (2024). Integration of chatbots in additional language education: A systematic review. *European Journal of Educational Research*, 13(4), 1607-1625. <https://doi.org/10.12973/eu-jer.13.4.1607>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cope, B., & Kalantzis, M. (2022). A little history of new literacies: A history of a big idea. *Education Sciences*, 12(5), 321. <https://doi.org/10.3390/educsci12050321>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Fang, S., & Han, Z.H. (2025). On the nascent stage of ChatGPT in foreign language teaching and learning. *Annual Review of Applied Linguistics*, 45, 148-178. <https://doi.org/10.1017/S026719052510010X>
- Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2-3), 87-105. [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- Gass, S. M., & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (2nd ed., pp. 180-206). Routledge.
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5-24. <https://doi.org/10.64152/10125/73474>
- Han, Z. H. (2024). ChatGPT in and for second language acquisition: A call for systematic research. *Studies in Second Language Acquisition*, 46(2), 301-306. <https://doi.org/10.1017/S0272263124000111>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues, and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112-131. [https://doi.org/10.30191/ETS.202301_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
- Hubbard, P. (2013). Making a case for learner training in technology-enhanced language learning environments. *Calico Journal*, 30(2), 163-178. <https://doi.org/10.11139/cj.30.2.163-178>
- Kamal, H. (2025). Teaching Arabic today: Challenges, strategies, and opportunities in Islamic higher education. *International Journal of Learning, Teaching and Educational Research*, 24(10), 644-659. <https://doi.org/10.26803/ijlter.24.10.31>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, J., Poquet, O., Sailer, M., Schmidt, A., & Seidel, T. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kessler, G. (2018). Technology and the future of language teaching. *Foreign Language Annals*, 51(1), 205-218. <https://doi.org/10.1111/flan.12318>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>

- Kucuk, T. (2024). ChatGPT-integrated grammar teaching and learning in EFL classes: A study of Tishk International University students in Erbil, Iraq. *Arab World English Journal Special Issue on ChatGPT*, 100-111. <https://doi.org/10.24093/awej/ChatGPT.6>
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers & Education Open*, 6, Article 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Lyu, B., Lai, C., & Guo, J. (2024). Effectiveness of chatbots in improving language learning: A meta-analysis of comparative studies. *International Journal of Applied Linguistics*, 34(4), 834–851. <https://doi.org/10.1111/ijal.12668>
- Maulidiya, A. R., Abdurrahman, M., & Saleh, N. (2025). Exploring AI capabilities in Arabic grammar: Comparative analysis of ChatGPT and Gemini. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 11(2), 160–174. <https://doi.org/10.15408/a.v11i2.42671>
- Nasaruddin, N. (2024). Using ChatGPT in teaching Arabic as a foreign language. *Arabiyatuna: Jurnal Bahasa Arab*, 8(1), 1-24. <https://doi.org/10.29240/jba.v8i1.9413>
- Nassaji, H., & Fotos, S. (2011). *Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context*. Routledge.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528. <https://doi.org/10.1111/0023-8333.00136>
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis* (Vol. 1) Cambridge University Press.
- Sinclair, J. M., & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.
- Stenbom, S., & Garrison, D. R. (2026). Artificial intelligence and communities of inquiry: Reimagining educational experiences. *The International Review of Research in Open and Distributed Learning*, 27(2), 114–131. <https://doi.org/10.19173/irrodl.v27i2.9314>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125–144). Oxford University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, 31(2), 57-71. <https://doi.org/10.1017/S0261444800012970>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39). <https://doi.org/10.1186/s41239-019-0171-0>