

Optimizing ChatGPT as a Scaffolding Tool for Arabic Text Comprehension: Evidence from Islamic Higher Education

Apri Wardana Ritonga¹⁾, Ayu Desrani²⁾, Agus Supriatna³⁾

^{1,3)} Institut Agama Islam As-Syifa, Subang, Indonesia

²⁾ Universitas Negeri Yogyakarta, Indonesia

E-mail: ¹⁾apriwardanaritonga@stiq.assyifa.ac.id, ²⁾ayudesrani@gmail.com,

³⁾agussupriatna@stiq.assyifa.ac.id

 <https://doi.org/10.54214/alfawaid.Vol16.Iss2.1061>

ABSTRACT

Arabic text comprehension is essential in Islamic higher education, yet many students struggle to understand Arabic texts. This study examines how ChatGPT functions as a scaffolding tool for Arabic reading comprehension, focusing on students' engagement patterns, scaffolding roles, and perceived impact on comprehension. Using an explanatory sequential mixed-methods design, data were collected from 37 undergraduate students through a structured questionnaire and supplemented by phenomenological interviews, participatory observation, and documentation. Quantitative data were analysed using frequencies, percentages, mean scores, and descriptive comparisons across prior ChatGPT-use groups, while qualitative data were analysed using Colaizzi's seven-step procedure. Findings indicate that ChatGPT functions as a moderate and selective scaffolding tool. Its perceived support was strongest for adaptive explanations and contextual meaning (mean 2.51) and weaker for nahwu and sharf (means 2.05–2.08). Students also reported faster comprehension and easier vocabulary access, but noted inaccurate responses requiring verification. Overall, the findings suggest that ChatGPT is perceived as more useful for contextual and lexical support than for technical grammatical assistance and should be used with critical verification.

Keywords : ChatGPT, Scaffolding, Arabic Text Comprehension, Islamic Higher Education

INTRODUCTION

Arabic language instruction in Islamic higher education carries a dual mandate that distinguishes it from all other foreign language contexts. Students must develop communicative competence while simultaneously mastering Arabic as the primary access key to the Qur'an, hadith, classical jurisprudence, and centuries of Islamic scholarship (Susanto et al., 2022). This dual obligation creates a reading demand that far exceeds standard second-language acquisition. Students are expected to navigate morphologically dense classical texts for which neither everyday exposure nor general linguistic knowledge adequately prepares them. Research consistently confirms that this structural pressure produces a persistent gap between students' formal grammar knowledge and their

actual reading performance (Zheng et al., 2023). Understanding this pressure is essential context for evaluating how AI tools are reshaping Arabic learning in Islamic higher education today.

Arabic text comprehension presents challenges that are both numerous and interconnected. Vaid (2022) identifies Arabic as a typologically extreme case for second-language readers, combining a non-Latin script, a root-and-pattern morphological system, and a case-marking system (*i'rab*) with no equivalent in Indonesian. The diglossic gap between Modern Standard Arabic and the classical register used in Islamic texts adds a further layer that learners of other languages simply do not face (Kamal, 2025). In practice, students who perform adequately on *nahwu* and *sharf* examinations frequently cannot read an unvocalised classical text independently (Adawiyah et al., 2025). This gap between declarative grammatical knowledge and fluent contextual reading is the core instructional problem that technology-based interventions in Arabic education must address.

ChatGPT has become one of the most rapidly adopted educational technologies in history. Released by OpenAI in late 2022, it reached 100 million users within eight weeks faster than any application previously recorded (Rudolph et al., 2023). In language learning settings, its appeal is intuitive: it provides contextual vocabulary explanations, sentence-level translations, and real-time grammatical clarification without the social anxiety of asking a human instructor. Mundofi (2025) found that university students across multiple disciplines reported ChatGPT significantly reduced the time they spent on language-related tasks, particularly reading comprehension activities. These affordances have made spontaneous ChatGPT adoption especially pronounced among students navigating the linguistic density of Arabic academic texts (Zubaidi et al., 2025).

The theoretical framework guiding this study is Vygotsky's (1978) sociocultural theory, specifically the Zone of Proximal Development (ZPD). Vygotsky defined the ZPD as the distance between what a learner achieves independently and what they achieve with appropriate guidance. To operationalise scaffolding in the present study, ChatGPT's scaffolding function was examined through four observable dimensions derived from the principles of contingent and graduated support. First, adaptive support refers to the extent to which ChatGPT adjusts its explanations to students' specific questions and perceived learning needs. Second, comprehension support refers to the extent to which ChatGPT assists students in constructing the meaning of an Arabic text, including vocabulary and overall contextual meaning. Third, linguistic support refers to assistance with specific linguistic features, particularly *nahwu* and *sharf*. Fourth, transfer toward independent comprehension refers to whether students perceive themselves as increasingly able to understand an Arabic text without continued assistance after interacting with ChatGPT. These dimensions

operationalise scaffolding as more than simple access to information: the tool is considered to provide scaffolding when its assistance responds to a learner's immediate difficulty, supports the construction of understanding, and potentially enables the learner to continue the task with reduced assistance.

Self-regulated learning theory provides a complementary lens for understanding how students engage with ChatGPT. Zimmerman (2000) identifies three phases of effective self-regulated learning and argues that cycling actively through all three produces more durable competencies than reactive processing alone. Bernacki et al. (2021) synthesised 129 studies and found that AI tools support SRL when students engage reflectively but undermine it when students engage passively, substituting AI output for their own cognitive work. Abdulaal et al. (2022) cognitive load theory reinforces this: the productive struggle with difficult Arabic texts generates germane cognitive load that drives schema formation. Eliminating this struggle without replacing it with equivalent engagement leaves comprehension capacity underdeveloped.

Lo et al. (2024) reviewed 70 studies on ChatGPT in ESL/EFL education and found scaffolding effects strongest for vocabulary and well-bounded tasks, weaker for complex grammatical processing. Morell-Mengual et al. (2025) surveyed 1,200 university students across Europe and reported that AI chatbot adoption is widespread but typically occasional, with perceived usefulness moderated strongly by prior domain knowledge. Ridho et al. (2025) documented a consistent gap between student enthusiasm for AI assistance and measurable improvement in independent performance, attributing this to over-reliance on AI output. Ismail et al. (2026) demonstrated through expert evaluation that ChatGPT loses morphosyntactic coherence in Arabic case marking and agreement precisely the features most critical for comprehension of classical Islamic texts.

Despite this growing body of research, three important gaps remain. First, Arabic language learning in Islamic educational settings is almost absent from reputable publication AI-and-language scholarship; existing Arabic-focused studies are methodologically limited to single-strand designs. Second, scaffolding is frequently invoked but rarely operationalised precisely enough to distinguish scaffolded performance from genuinely developed comprehension capacity the distinction that matters most theoretically. Third, no study has disaggregated ChatGPT's effects across the multiple dimensions of Arabic reading to map where AI alignment with learner needs breaks down. This study addresses all three gaps through an integrated mixed-methods design combining quantitative survey data with phenomenological interview evidence, producing a more complete and differentiated account than any single-method approach could yield.

This study examines three research questions corresponding to three phases of the scaffolding interaction. RQ1 asks how students at Islamic universities engage with ChatGPT in Arabic reading contexts. RQ2 asks what scaffolding functions ChatGPT actually performs, and where its effects are strongest and weakest. RQ3 asks what impact ChatGPT use has on comprehension, and how that impact varies by prior experience with the tool. The study contributes three novel elements: an integrated mixed-methods account capturing both distributional patterns and experiential meanings; a granular map of AI scaffolding alignment across vocabulary, grammar, and independent comprehension dimensions; and a theoretically grounded analysis of ChatGPT in Islamic higher education, a context whose stakes for Arabic reading competence extend well beyond academic performance into religious and scholarly practice.

METHOD

Research Design

This study employed an explanatory sequential mixed-methods design. Quantitative data were analyzed first to identify patterns in students' use of ChatGPT and their perceived comprehension outcomes. Qualitative data were then used to explore these patterns in greater depth, focusing on students' perceptions of ChatGPT as a scaffolding tool. Quantitative findings also informed qualitative participant selection based on semester, ChatGPT-use frequency, and variation in questionnaire responses. The two strands were integrated through a joint display organized around the three research questions. Quantitative results were compared with relevant qualitative themes to determine whether the qualitative findings explained, extended, or qualified the quantitative patterns.

Setting and Participants

The study was conducted at Institut Agama Islam As-Syifa, a private Islamic higher education institution in Subang Regency, West Java, Indonesia, where Arabic is a compulsory course for undergraduate students across Semesters II, IV, VI, and VIII, with a total population of 100 students. For the quantitative strand, 37 students completed a structured questionnaire. They were selected through a purposive criterion based on prior exposure to Arabic reading tasks in an academic context, followed by random selection within the eligible pool. The sample comprised 28 female (75.7%) and 9 male (24.3%) students, distributed across Semester II ($n = 9$, 24.3%), Semester IV ($n = 13$, 35.1%), Semester VI ($n = 13$, 35.1%), and Semester VIII ($n = 2$, 5.4%). For the qualitative strand, a subset of 9 to 12 students was selected from the survey respondents through maximum

variation purposive sampling, prioritising variation in semester, self-reported ChatGPT experience level, and survey response patterns.

Data Collection

Quantitative data were collected through a structured questionnaire administered in March 2026. The instrument comprised three thematic constructs: (1) general use of ChatGPT in Arabic reading (four items), (2) ChatGPT's role as a scaffolding tool for Arabic reading comprehension (six items), and (3) the impact of ChatGPT use on Arabic reading comprehension (six items). All items used a four-point ordinal scale (1 = strongly disagree / never, 4 = strongly agree / always), with a forced-choice format (no midpoint) to avoid central tendency bias. An open-ended item asked students to describe, in their own words, the positive and negative effects of using ChatGPT in their Arabic studies. The questionnaire was pilot-tested with a subsample of ten students not included in the main study; internal consistency was assessed using Cronbach's alpha, with a threshold of $\alpha \geq 0.70$ (Creswell & Poth, 2018).

Qualitative data were collected from February to May 2026 through three complementary techniques. In-depth phenomenological interviews with student and lecturer participants used open-ended questions to elicit narratives about the lived experience of using ChatGPT in Arabic reading contexts, following the iterative interview structure recommended by Moustakas (1994): an initial session to build rapport and gather broad narratives, a second session to probe themes and clarify meanings, and a third session where needed for member-checking. Each interview session lasted between 60 and 90 minutes and was audio-recorded with participants' written consent, then transcribed verbatim within 48 hours. Participatory observation of students engaging with ChatGPT during Arabic reading tasks was conducted on separate occasions to capture behavioural and interactional dimensions not readily accessible through interview alone. Documentation was collected as a secondary triangulation source.

Data Analysis

Quantitative data were analyzed using descriptive statistics, including item-level means, construct-level means, response frequencies, and descriptive comparisons across ChatGPT-use groups. The analysis aimed to identify patterns within the sample rather than test hypotheses or make statistical generalizations. These findings informed the focus of the qualitative analysis. Qualitative data were analyzed using Colaizzi's (1978) seven-step phenomenological procedure: reading transcripts, identifying significant statements, formulating meanings, organizing themes, developing an exhaustive description, identifying the essential structure, and validating the findings

through participant feedback. This procedure was appropriate for interpreting participants' lived experiences, provided that the qualitative data consisted of interview transcripts suitable for phenomenological analysis.

RESULT AND DISCUSSION

Percentage of Students' ChatGPT Use in Arabic Text Learning

This section presents the distribution of students' prior ChatGPT use to establish their level of familiarity with the tool before examining its scaffolding functions and perceived impact on Arabic text comprehension. Figure 1 shows the respondents' usage frequency, while their gender, semester, and duration of Arabic study provide additional context for interpreting the findings.

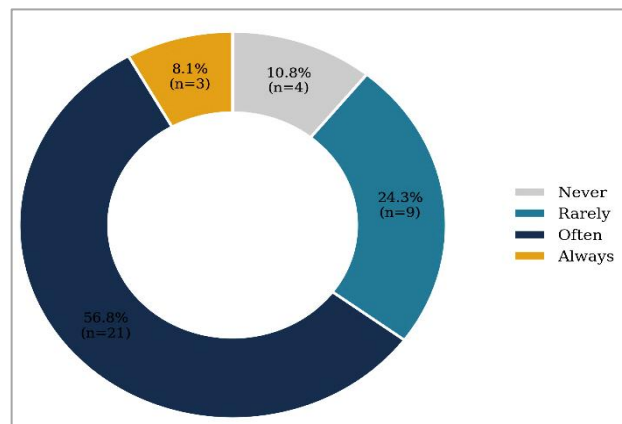


Figure 1. Distribution of prior ChatGPT experience levels (n = 37)

Figure 1 shows that 56.8% of respondents used ChatGPT often, while 24.3% used it rarely, 10.8% had never used it, and 8.1% used it constantly. Thus, 65% of respondents reported using ChatGPT often or constantly, indicating that most participants had prior experience with the tool. Of the 37 respondents, 28 (75.7%) were female and 9 (24.3%) were male. Semester IV and VI each accounted for 13 students (35.1%), followed by Semester II with 9 (24.3%) and Semester VIII with 2 (5.4%). Regarding Arabic learning experience, 22 students (59.5%) had studied Arabic for one to three years, 13 (35.1%) for four to seven years, and 2 (5.4%) for eight to fifteen years. These distributions describe the sample's background and prior familiarity with ChatGPT, providing context for the subsequent analysis of its perceived scaffolding functions and impact.

This finding is consistent with the research of Morell-Mengual et al. (2025), who similarly report that roughly two-thirds of university students use chatbots in their studies, though typically with occasional rather than constant frequency. Not only that, the study by Amoah et al. (2025) analyzing more than 20,000 student responses across 109 countries found that over 70% of students

already use ChatGPT, with usage somewhat higher in developed countries than in developing ones, a proportion that further confirms how widespread prior familiarity with the tool has become. This convergence suggests that functional, need-driven engagement rather than constant reliance has become the common pattern of student-AI interaction. However, Polyportis (2024) tracked Dutch higher-education students over an eight-month period and found a significant decline in ChatGPT usage behavior over time, driven by shifts in trust and perceived ease of use, indicating that the relatively high familiarity captured in this cross-sectional snapshot should not be read as a stable or permanently increasing trend.

These adoption patterns take on particular significance when situated within language-learning research specifically. Lo et al. (2024), in a systematic review of 70 empirical studies on ChatGPT in ESL/EFL education, found that the tool most often scaffolded learners by acting as a partner in practising language, providing feedback, and recommending further practice activities, a function that aligns closely with how respondents in this study described using ChatGPT for stepwise text explanation and vocabulary clarification. Similarly, Aljohani (2026) reported in a 2025 systematic review that students increasingly used ChatGPT for interactive, scenario-based language practice rather than passive translation alone. At the institutional level, Murgia et al. (2023) found considerable variation in ChatGPT adoption patterns depending on student awareness, academic use, and perceived instructional support, which helps explain why, despite widespread familiarity, comfort with using ChatGPT for Arabic learning remained the lowest-scoring item in this study's usage construct. This body of work reinforces the interpretation that scaffolding through ChatGPT is real but conditional, shaped by how students are guided to use it.

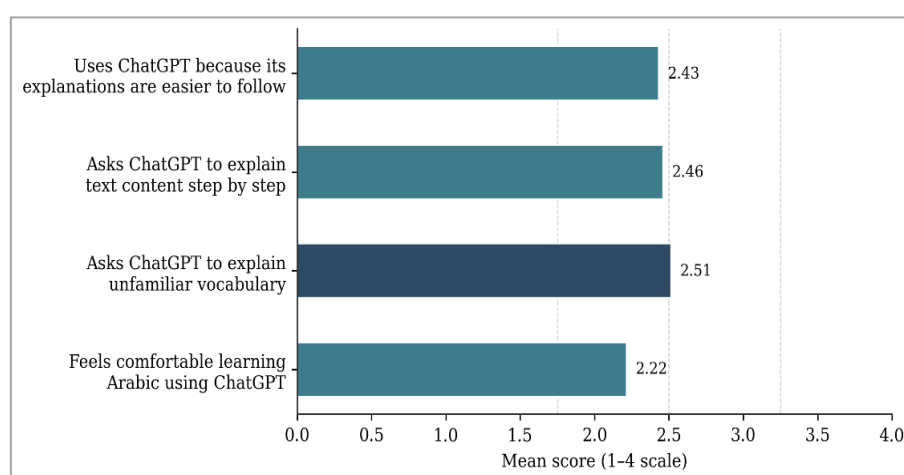


Figure 2. General use of ChatGPT in reading Arabic texts

The general-use items point to a moderate reliance on ChatGPT for Arabic reading support, with a construct mean of 2.41, categorized as moderate. As Figure 2 shows, students turned to ChatGPT most often to clarify unfamiliar vocabulary (mean 2.51), a pattern consistent with Ismail et al. (2026), who grounded their work in Vygotsky's sociocultural theory and found that classroom integration of ChatGPT measurably improved vocabulary retention compared to conventional teacher-led instruction. The advantage over older reference habits is not merely perceptual either: in a controlled comparison, Zubaidi et al. (2025) found that students using ChatGPT for lexical tasks completed them faster and more accurately than those relying on a conventional dictionary, which helps explain why vocabulary lookup edged out every other use case in this sample. Close behind, students also asked ChatGPT to walk through text content step by step (mean 2.46), echoing Lo et al. (2024) review of ESL/EFL research, where stepwise, conversational explanation was identified as one of ChatGPT's most consistently reported scaffolding functions. Together, these two leading items position the tool primarily as a lexical and comprehension aid rather than a substitute for independent reading.

Comfort with using ChatGPT for Arabic learning told a different story, scoring lowest among the four items. Klimova et al. (2024) observed a similar split in their own student sample: participants described themselves as fascinated and satisfied with ChatGPT, yet that enthusiasm coexisted with real reservations about relying on it for language learning. The present data mirror this exactly, students draw on the tool functionally without having fully made it part of their comfort zone. This kind of gap has less to do with how often a tool gets used than with the trust and familiarity built around it; work summarizing findings from Ismail et al. (2026), Pham & Nguyen (2025) suggests that any anxiety-reducing benefit of ChatGPT is inconsistent and depends on how much prior exposure and classroom guidance a student has had.

Open-ended responses indicate varied patterns of ChatGPT use. Of the 37 respondents, 16 described their experiences positively, often using terms such as *mudah* (easy) and *terbantu* (helped), while nine specifically mentioned translation and vocabulary lookup. One student explained,

"I use ChatGPT when I encounter unfamiliar Arabic vocabulary because it helps me understand the meaning of a text more quickly. However, I still ask my lecturer or friends when the explanation is unclear" (PM, Interview, 2026).

In contrast, 11 students reported rarely or never using ChatGPT, citing personal preferences and limited access to technology. As one student stated,

“ChatGPT is helpful for translating Arabic texts, but I do not use it regularly because I sometimes have limited access to a mobile phone. When I cannot access it, I prefer to ask my friends” (MA, Interview, 2026).

These accounts illustrate how students’ engagement varies according to perceived usefulness, learning habits, and access to technology.

Mean Scores for ChatGPT’s Role as a Scaffolding Tool in Arabic Learning

This section examines whether students’ use of ChatGPT functions as meaningful scaffolding for learning, rather than simply reflecting frequent use. The six items focus on how ChatGPT supports students in understanding Arabic texts, where it is most helpful, and where its support remains limited.

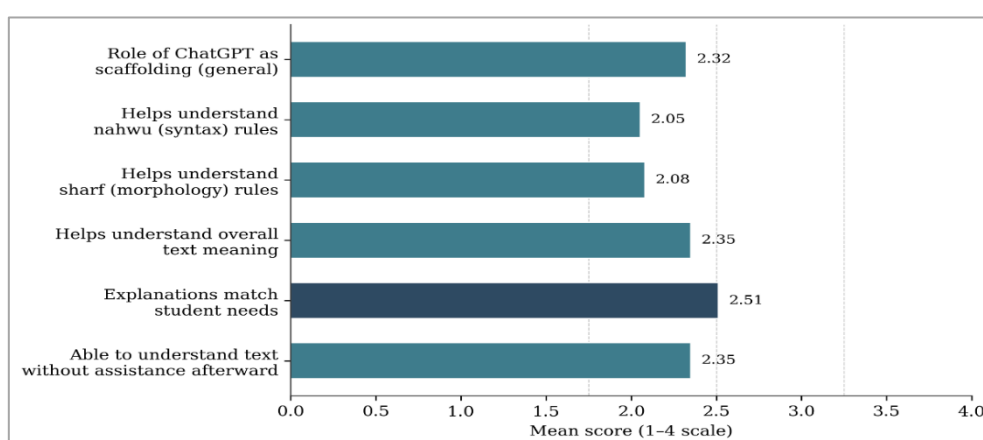


Figure 3. ChatGPT’s role as scaffolding for Arabic reading comprehension

Based on Figure 3 above, the construct as a whole scored 2.28, “moderate”, but the item-level spread tells a more specific story than the average alone. The single highest-scoring item was *“explanations match student needs”*, edging out even the more general items about helping with meaning or building independent comprehension. This is a meaningful detail: students did not rate ChatGPT highest for what it explains, but for how well it adapts that explanation to them. Two items tied for the next tier suggesting that once ChatGPT’s explanation lands, students do walk away somewhat more capable of reading independently, at least in this sample’s own estimation.

The bottom of the ranking is where the pattern sharpens. Both grammar-related items sit clearly below every other item in the construct, a full quarter-point or more under the construct mean. This is not a marginal dip; it is a consistent gap across two related items, which suggests it reflects something structural about the kind of help ChatGPT provides rather than random variation in individual responses. Put plainly, students rate ChatGPT as a stronger partner for arriving at what a text means than for understanding why it is grammatically constructed the way it is. The data

echo this split almost exactly: of the thirty-seven open-ended responses describing ChatGPT's role, eight emphasized its general importance, six pointed to translation, and six to vocabulary but only two mentioned grammar rules at all.

This split lines up closely with how scaffolding is understood in Vygotsky's sociocultural theory of learning. Central to this theory is the Zone of Proximal Development (ZPD) the space between what a learner can do unaided and what they can do with the right kind of support. Vygotsky himself never used the word "scaffolding"; the term was coined later by Wood et al. (1976) to describe how a more capable other, whether teacher, peer, or tool, provides support that is responsive and temporary, withdrawn as competence grows. Cong-Lem & Daneshfar (2024) apply this framework directly to generative AI in language learning, arguing that tools like ChatGPT can occupy the role of that "more capable other" when their support is genuinely calibrated to a learner's current gap in understanding, rather than simply supplying finished answers. Seen through this lens, it is not surprising that "explanations match student needs" scored highest in this study: it is the one item that describes ZPD-consistent support directly, rather than describing ChatGPT's output in the abstract.

The qualitative pattern reinforces this reading in a way the numbers alone cannot. Students who described ChatGPT's role in terms of translation or vocabulary were, in effect, describing task-specific, bounded interactions the kind of narrow, well-defined request that a system built on pattern completion handles comfortably. Cong-Lem & Daneshfar (2024) account of AI-mediated ZPD support anticipates exactly this limitation: scaffolding is only as strong as the AI's capacity to accurately model the learner's specific gap, and that capacity is uneven across different types of linguistic knowledge. Lexical meaning is generally easier to model and verify than grammatical structures, particularly in morphologically complex languages such as Arabic (Yuliani et al., 2025). This may help explain why only two students identified grammar as a specific area in which they used ChatGPT. Their limited references to grammar do not necessarily indicate that they consider grammar unimportant; rather, they may reflect uncertainty about the tool's reliability in this area.

That sense is also consistent with recent linguistic evaluations of ChatGPT-generated Arabic. Ismail et al. (2026), for example, identified recurring problems in grammatical agreement, inflection and case marking, sentence structure, preposition use, and transitivity. For illustration, an agreement problem in Arabic may occur when a plural subject is paired with a verb that does not appropriately reflect the required agreement pattern, such as using "الطلاب يدرس" instead of the grammatically

appropriate “الطلاب يدرسون” in a context requiring full subject-verb agreement. Similarly, case-marking errors may involve an inappropriate nominative, accusative, or genitive ending in a syntactic position where another case is required. These examples are illustrative of the grammatical domains identified in the literature and are not presented as errors extracted from the participants’ ChatGPT interactions in the present study. The distinction is important because the present study measured students’ perceptions of ChatGPT’s grammatical assistance rather than conducting a systematic linguistic error analysis of its outputs.

The Impact of ChatGPT Use on Comprehension of Arabic Texts

This section examines students’ perceived comprehension outcomes across different levels of prior ChatGPT use. Figure 4 compares the mean scaffolding and impact scores across the four usage groups, showing how students’ perceptions vary according to their frequency of using ChatGPT.

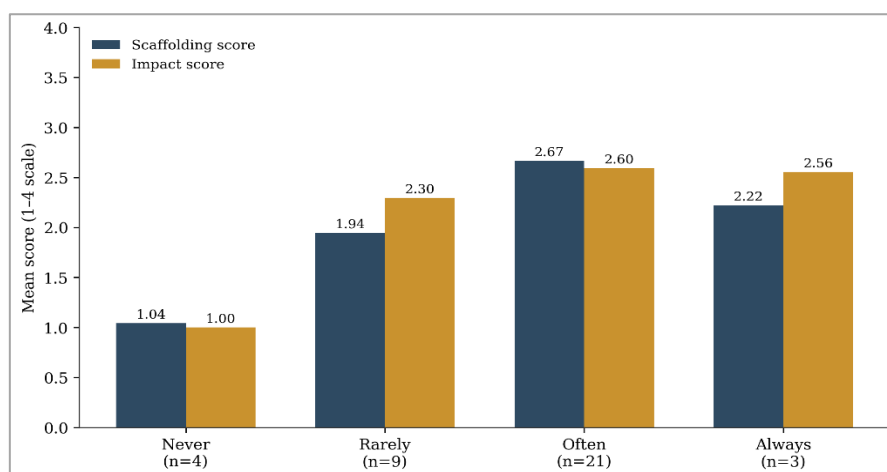


Figure 4. Mean scaffolding and impact scores by level of prior ChatGPT experience

Based on Figure 4 above, the pattern across experience levels is, on its own terms, orderly. Students with no prior ChatGPT experience report the lowest scores on both constructs (scaffolding 1.04, impact 1.00), functioning almost as a validity check: those who have not used the tool cannot logically rate its help highly, and these near-floor scores confirm the scale is behaving as expected. Scores then rise together through “Rarely” and peak at “Often” (scaffolding 2.67, impact 2.60), the same group that made up the majority of the sample ($n = 21$). Read against Figure 4’s distribution, this means the moderate construct means reported earlier are being pulled upward mainly by this one large, moderately-frequent-use group, rather than reflecting uniformly moderate experiences across the whole sample.

What the pattern obscures at a glance, however, is that scaffolding and impact do not move in lockstep within each group and the direction of that gap flips in a way that is worth reading closely.

Published by STAI Ali bin Abi Thalib Surabaya

Website Jurnal : <https://jurnal.stai-ali.ac.id/index.php/Alfawaaid>

DOI Jurnal : <https://doi.org/10.54214/alfawaaid>

Among “Rarely” users, impact (2.30) clearly outpaces scaffolding (1.94), a gap of 0.36; the same is true, even more sharply, for “Always” users (impact 2.56 vs. scaffolding 2.22, a gap of 0.34). Only in the “Often” group does scaffolding edge slightly ahead of impact (2.67 vs. 2.60). In other words, students at both the low and the very high ends of experience report feeling more benefit than they attribute to ChatGPT’s role as a structured learning aid, while the large middle group is the only one where perceived scaffolding matches or slightly exceeds perceived impact. One reading is that infrequent users experience ChatGPT as a convenient shortcut whose benefit is felt but not really processed as “scaffolding,” while the heaviest users may have folded the tool so fully into their routine that its structural, teaching-like role becomes less visible to them even as its practical payoff remains high. The “Always” group’s small size ($n = 3$) means this last point should be treated as suggestive rather than conclusive, but the consistency of the gap’s direction across two independent groups on either end of the experience spectrum makes it a pattern worth flagging rather than dismissing as noise.

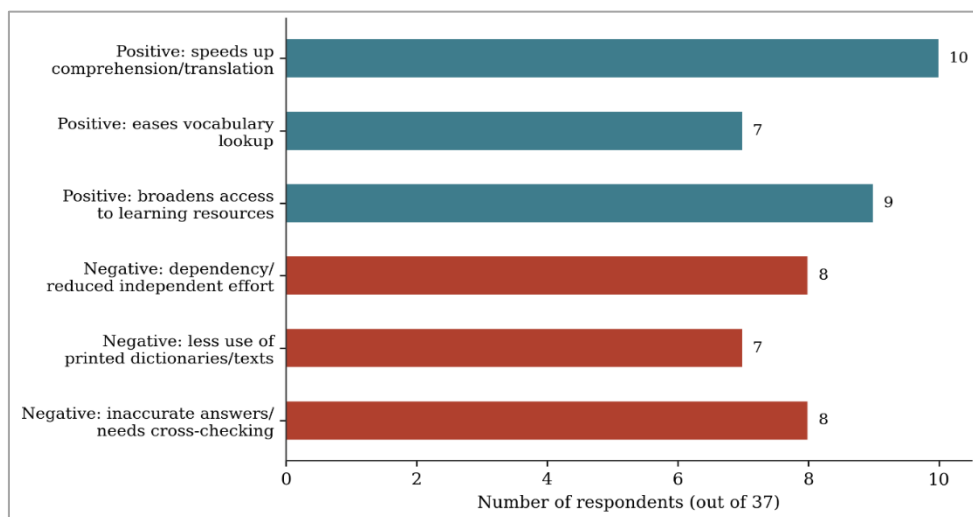


Figure 5. Themes of positive and negative impact from open-ended responses

Figure 5 above shows that the six themes split almost exactly down the middle in volume which is itself a notable finding: this is not a student body enthusiastically endorsing ChatGPT with occasional grumbling on the side, nor one raising serious alarm with a few positive footnotes. The near-even split mirrors the “moderate” categorization of the impact construct almost exactly, suggesting the quantitative and qualitative data are not just broadly consistent but converging on the same underlying ambivalence from two independent angles. Because open-ended responses were not mutually exclusive this balance likely reflects genuine dual awareness within individual students rather than two separate camps of critics and enthusiasts talking past each other.

The content of the themes also lines up tightly with the item-level data from Figure 4, which sharpens what the balance actually means. The leading positive theme, faster comprehension and translation ($n = 10$), tracks directly onto the item “understands text faster” (mean 2.35); and the leading negative theme, inaccurate answers requiring cross-checking ($n = 8$), tracks onto both “received inaccurate answers” (mean 2.30) and the highest-scoring item in the entire impact construct, “always cross-checks answers with other sources” (mean 2.65). That the single most-endorsed behavior in the whole construct is a corrective one reframes the negative theme cluster: this is not students discovering a flaw despite themselves, but students actively building a workaround into how they use the tool. Speed is being gained, but only at the cost of an extra verification step students have already learned to take, which is a more specific and more actionable finding than “impact is moderate” on its own would suggest.

This unease is well documented elsewhere. Adawiyah et al. (2025) traces a comparable trajectory in a qualitative case study of university students, where a tool initially adopted for efficiency gradually became something closer to a cognitive crutch, undermining the very critical thinking and independent effort it was meant to support. Saif et al. (2024), applying the Technology Acceptance Model (TAM) to ChatGPT specifically in education, similarly found that perceived usefulness reliably drove adoption, but did not by itself guarantee deeper engagement with the material usefulness and genuine learning gain are not the same thing, and this study’s data reflect that same gap between efficiency and depth. Taken together with the item-level finding that “always checking answers against other sources” was one of the highest-scoring items in the impact construct, this literature suggests the negative themes identified here are not a sign of naivety on students’ part, but rather an early, appropriate wariness that the technology acceptance literature would predict.

This brings the three research questions to a single, coherent answer. Students’ experience with ChatGPT (RQ1) is functional and selective, not habitual; its role as scaffolding (RQ2) is real but concentrated where the tool’s own linguistic competence is strongest, namely vocabulary and meaning rather than grammar; and its impact on comprehension (RQ3) follows the same logic, scaling with use but tempered by an emerging awareness that speed and ease come with a cost to independent skill-building. Optimizing ChatGPT as a scaffold for Arabic reading comprehension, then, is less a matter of increasing how often students use it, and more a matter of directing that use toward the tasks where it is genuinely reliable.

CONCLUSIONS

This study set out to examine how ChatGPT functions as a scaffolding tool for Arabic text comprehension among undergraduate students at an Islamic higher education institution, addressing three research questions concerning student engagement patterns, scaffolding roles, and perceived comprehension impact. The findings indicate that students use ChatGPT selectively, particularly when encountering unfamiliar vocabulary or difficulties in understanding Arabic texts. Its perceived scaffolding role is uneven across linguistic domains: adaptive explanations and contextual meaning receive higher ratings, whereas assistance with *nahwu* and *sharf* receives lower ratings. The descriptive comparison across prior-use groups shows that students reporting “Often” use have the highest perceived scaffolding and impact scores in this sample, while the small “Always” group shows a different pattern. Qualitative responses further indicate that students perceive practical benefits such as faster comprehension and easier vocabulary access, while also reporting inaccurate answers and the need for cross-checking. Taken together, these findings suggest a tendency for ChatGPT to be perceived as more useful when it is used selectively and critically, rather than as an unconditional substitute for independent reading. However, this study does not establish a causal relationship between ChatGPT use and cognitive engagement, and the observed group differences should therefore be interpreted as descriptive patterns rather than experimental effects.

These findings carry significance that extends beyond the immediate context of Arabic language instruction. Theoretically, they contribute empirical evidence to the emerging construct of AI-mediated scaffolding, demonstrating that Vygotsky’s ZPD framework remains productive for analysing AI-assisted learning, but only when the analysis is sensitive to the alignment between AI competence and target linguistic demands. This study is not without limitations. The sample of 37 students from a single institution restricts the transferability of quantitative findings, and the cross-sectional design cannot establish whether observed comprehension benefits persist over time. Future research should employ longitudinal designs tracking comprehension development across multiple semesters, expand to multiple Islamic universities to test the generalisability of these patterns, and investigate the differential effects of structured versus unstructured ChatGPT use through experimental or quasi-experimental designs.

ACKNOWLEDGMENT

The authors would like to thank the Research and Community Service Institute of Institut Agama Islam As-Syifa Subang for funding this research under grant number: 11/SKP/LPPM/Institut/As-syifa/VI/2026. The authors declare no conflict of interest.

REFERENCES

- Abdulaal, M. A. A. D., Alenazi, M. H., Tajuddin, A. J. A., & Hamidi, B. (2022). Dynamic vs. Diagnostic Assessment: Impacts on EFL Learners' Speaking Fluency and Accuracy, Learning Anxiety, and Cognitive Load. *Language Testing in Asia*, 12(1). <https://doi.org/10.1186/s40468-022-00179-0>
- Adawiyah, S. R., Wasilah, W., Prasetyo, B., & Mardha Tilla. (2025). Optimalisasi Pembelajaran Nahwu-Shorof dengan Buku Al-Miftāḥ li Al-'Ulūm Studi di MTs Ponpes Sabiilillah Kayuagung. *Jurnal Al-Fawa'id : Jurnal Agama Dan Bahasa*, 15(1), 121–140. <https://doi.org/10.54214/alfawaid.Vol15.Iss1.813>
- Aljohani, N. J. (2026). ChatGPT in Language Learning: A Systematic Review of Applications and Challenges. *Social Sciences & Humanities Open*, 13, 102357. <https://doi.org/https://doi.org/10.1016/j.ssaho.2025.102357>
- Amoah, A., Asiama, R. K., & Kwablah, E. (2025). ChatGPT Early Usage Among Students: A Global Evidence of Determinants. *Development and Sustainability in Economics and Finance*, 7, 100065. <https://doi.org/10.1016/j.dsef.2025.100065>
- Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A Systematic Review of Research on Personalized Learning: Personalized by Whom, to What, How, and for What Purpose(s)? *Educational Psychology Review*, 33(4), 1675–1715. <https://doi.org/10.1007/s10648-021-09615-8>
- Colaizzi, P. F. (1978). Psychological Research as the Phenomenologist Views it. In *Existential-phenomenological alternatives for psychology* (pp. 48–71). Oxford University Press.
- Cong-Lem, N., & Daneshfar, S. (2024). *Generative AI and Second/Foreign Language Education from Vygotsky's Cultural-Historical Perspective* (pp. 175–188). https://doi.org/10.1007/978-3-031-63447-5_10
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications.
- Ismail, A. F., Alqudah, R. A., Al-Saliti, R. A. M. N., & Hamid, A. A. (2026). Grammatical Error Patterns in ChatGPT-Generated Modern Standard Arabic Texts: A Linguistic Analysis of Recurrent Patterns. *Languages*, 11(5), 86. <https://doi.org/10.3390/languages11050086>
- 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>
- Klimova, B., Pikhart, M., & Al-Obaydi, L. H. (2024). Exploring the Potential of ChatGPT for Foreign

Published by STAI Ali bin Abi Thalib Surabaya

Website Jurnal : <https://jurnal.stai-ali.ac.id/index.php/Alfawaid>

DOI Jurnal : <https://doi.org/10.54214/alfawaid>

- Language Education at the University Level. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1269319>
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the Application of ChatGPT in ESL/EFL Education and Related Research Issues: A Systematic Review of Empirical Studies. *Smart Learning Environments*, 11(1), 50. <https://doi.org/10.1186/s40561-024-00342-5>
- Morell-Mengual, V., Fernández-García, O., Berenguer, C., Ortega-Barón, J., Gil-Llario, M. D., & Estruch-García, V. (2025). Characteristics, Motivations and Attitudes of Students Using ChatGPT and Other Language Model-based Chatbots in Higher Education. *Education and Information Technologies*, 30(15), 22257–22274. <https://doi.org/10.1007/s10639-025-13650-1>
- Moustakas, C. (1994). *Phenomenological Research Methods*. SAGE Publications. <https://doi.org/10.4135/9781412995658>
- Mundofi, A. A. (2025). Integration of Deep Learning Approach in Transforming Islamic Religious Education Learning in Schools: A Pedagogical and Technological Study. *Journal of Asian Primary Education (JOAPE)*, 2(1), 3062–9268. <https://doi.org/10.59966/joape.v2i1.1787>
- Murgia, E., Abbasiantaeb, Z., Aliannejadi, M., Huibers, T., Landoni, M., & Pera, M. S. (2023). ChatGPT in the Classroom: A Preliminary Exploration on the Feasibility of Adapting ChatGPT to Support Children's Information Discovery. *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, 22–27. <https://doi.org/10.1145/3563359.3597399>
- Pham, T. S., & Nguyen, L. N. Q. (2025). Perceptions of Pre-service Kindergarten Teachers on Using ChatGPT to Support English Learning: A Mixed-Methods Study in Vietnam. *International Journal of AI in Language Education*, 2(3), 40–56. <https://doi.org/10.54855/ijaile.25233>
- Polyportis, A. (2024). A Longitudinal Study on Artificial Intelligence Adoption: Understanding the Drivers of ChatGPT Usage Behavior Change in Higher Education. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.1324398>
- Ridho, R., Perkasa, M. R., & Maujud, F. (2025). Exploring the Application of Artificial Intelligence in Arabic Language Learning: Perspectives of Students and Teachers. *Academia Open*, 10(2). <https://doi.org/10.21070/acopen.10.2025.12886>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education? *Journal of Applied Learning & Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Saif, N., Khan, S. U., Shaheen, I., ALotaibi, F. A., Alnfiai, M. M., & Arif, M. (2024). Chat-GPT; Validating Technology Acceptance Model (TAM) in Education Sector via Ubiquitous Learning Mechanism. *Computers in Human Behavior*, 154, 108097. <https://doi.org/10.1016/j.chb.2023.108097>
- Susanto, S., Desrani, A., Ritonga, A. W., & Rubiyantoro, Y. (2022). Improving Students' Creative Thinking In Learning Arabic Through HOTS Based Project Based Learning Model. *An Nabighoh*, 24(1), 1.

<https://doi.org/10.32332/an-nabighoh.v24i1.3924>

- Vaid, J. (2022). Biscrptality: A Neglected Construct in the Study of Bilingualism. *Journal of Cultural Cognitive Science*, 6(2), 135–149. <https://doi.org/10.1007/s41809-022-00101-3>
- Vygotsky, L. S. (1978). *Mind in Society. The Development of Higher Psychological Processes*. Harvard University Press.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Yuliani, L., Ritonga, A. W., & Machmudah, U. (2025). Efektivitas Media Interaktif Berbasis Digital dalam Meningkatkan Berpikir Kritis Siswa pada Pembelajaran Mahārah Qirā'ah di MTs Plus Al-Azhar Wlingi Blitar. *Jurnal Al-Fawa'id: Jurnal Agama Dan Bahasa*, 15(2), 453–469. <https://doi.org/10.54214/alfawaid.Vol15.Iss2.904>
- Zheng, H., Miao, X., Dong, Y., & Yuan, D.-C. (2023). The Relationship Between Grammatical Knowledge and Reading Comprehension: A Meta-Analysis. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1098568>
- Zimmerman, B. J. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*, 82–91. <https://doi.org/10.1006/ceps.1999.1016>
- Zubaidi, A., Munip, A., Widodo, S. A., & Zerrouki, T. (2025). Enhancing Arabic Writing Skills Using Chat GPT-based AI Learning Models: A Tridimensional Human-AI Collaboration Framework. *Indonesian Journal of Applied Linguistics*, 15(1), 87–101. <https://doi.org/10.17509/ijal.v15i1.75378>