



## Bridging the Nahwu–Insyā' Gap: The Effectiveness of Syntactic Integration in Arabic-as-a-Foreign-Language Writing Instruction

Aidillah Suja<sup>1\*</sup>, Syarifaturrahmatullah<sup>2</sup>, Moh. Syaifudin<sup>3</sup>, Renti Yasmar<sup>4</sup>, Ade D. Deviana<sup>5</sup>,  
Achmad T. Rusady<sup>6</sup>

<sup>1</sup> Sekolah Tinggi Agama Islam Negeri Sultan Abdurrahman Kepulauan Riau, Indonesia

<sup>2</sup> Universitas Islam Negeri Sultan Aji Muhammad Idris, Indonesia

<sup>3</sup> Sekolah Tinggi Ilmu Bahasa Arab dan Dakwah Masjid Agung Sunan Ampel, Indonesia

<sup>4</sup> Institut Agama Islam Negeri Curup, Indonesia

<sup>5</sup> Sekolah Tinggi Ilmu Tarbiyah Darul Hijrah Martapura, Indonesia

<sup>6</sup> Universitas Muhammadiyah Malang, Indonesia

Correspondence: E-mail: [aidillahsuja87@gmail.com](mailto:aidillahsuja87@gmail.com)

### ARTICLE INFO

#### Article History:

Submitted/Received 10 Aug 2026

First Revised 31 Aug 2026

Accepted 12 Oct 2026

Publication Date 17 Oct 2026

#### Keywords:

Arabic as a foreign language

Arabic writing instruction

Quasi-experimental study

Syntactic integration

Writing proficiency

### ABSTRACT

Although syntax is recognized as a fundamental component of foreign-language writing ability, syntactic knowledge acquired separately from writing practice rarely transfers into actual text-production performance. This study examined the effectiveness of a syntax-integration model in Arabic writing instruction within the Indonesian Islamic higher-education context, using a quasi-experimental pretest–posttest control-group design involving 108 Arabic Language Education students from three universities. The experimental group ( $n = 53$ ) received integrated instruction—authentic-text modeling, sentence-and-paragraph reconstruction, collaborative editing, and explicit syntactic feedback—while the control group ( $n = 55$ ) received conventional writing instruction. Because the data were not normally distributed, analysis used the Wilcoxon signed-rank test for within-group comparisons and the Mann-Whitney U test for between-group gain scores, complemented by the rank-biserial correlation effect size. Both groups showed significant improvement in writing ability. However, the experimental group showed a significantly higher gain score ( $Z = -2.788$ ,  $p = .005$ ,  $r = .27$ ). These findings indicate a measurable added value of syntactic integration, albeit with a small-to-moderate effect size, and offer contextual empirical evidence for pedagogy in Arabic as a foreign language.

Copyright (c) [2026] Aidillah Suja, Syarifaturrahmatullah, Moh. Syaifudin, Renti Yasmar, Ade Destri Deviana, Achmad Tito Rusady.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## 1. INTRODUCTION

The relationship between syntactic knowledge and writing ability has long been a central concern in foreign-language pedagogy. Although syntax is widely recognized as a fundamental component of writing ability, a persistent challenge in instructional contexts is that grammatical knowledge acquired in isolation rarely transfers into effective writing performance (DeKeyser & Suzuki, 2025). Learners may be able to identify or analyze syntactic structures, yet struggle to apply that knowledge functionally when constructing extended text. This gap between declarative syntactic knowledge and procedural writing performance is a recurring problem across foreign-language learning contexts, with direct implications for curriculum design, instructional practice, and language-teacher education.

Syntax is one of the linguistic components consistently identified as having a significant influence on writing ability, particularly in foreign-language contexts. In the context of Arabic as a foreign language, syntax is operationalized in this study through *Nahwu*—the Arabic grammatical discipline governing the structural organization of words into well-formed sentences, including case-ending rules (*i'rāb*), subject–predicate agreement (*muthābaqah*), and the formation of verbal sentences (*jumlah fi'liyah*) and nominal sentences (*jumlah ismiyah*) through constituents such as verb–subject (*fi'l-fā'il*), object (*maf'ul bih*), adverbials of place or time (*ẓarf al-makān/al-zamān or jār-majrūr*), and topic–comment structures (*mubtada'–khabar*).

This set of categories corresponds directly to the *Nahwu* content targeted in the instructional intervention and the writing-ability rubric used in this study (see Methods), ensuring that the construct of syntax discussed throughout this literature review is conceptually and operationally anchored to the specific grammatical content under investigation (Ryding, 2005). A number of studies have shown that syntactic awareness contributes to sentence-structure comprehension and overall writing quality (Ishak, 2018; Aziza & Sutanto, 2024). These findings are consistent with recent research confirming that syntactic features—such as phrase complexity and the use of subordinate clauses—correlate directly with foreign-language writing quality (Zheng & Barrot, 2024; Wang et al., 2025).

A cross-linguistic and developmental meta-analysis by Tong et al. (2025) further confirms that syntactic skill is a consistent predictor of reading ability across age groups and language backgrounds. Recent developments in the literature, however, mark an important shift in how syntax is understood. Syntax is no longer viewed solely as an indicator of linguistic ability, but as a literacy practice actively involved in meaning construction. Pölcz (2022) positions syntax as a rhetorical device for organizing relationships between ideas, while Akbari (2014) argues that sentence-structure choices affect text cohesion and coherence.

Within this framework, syntax is not merely a tool for producing grammatical sentences, but a cognitive and rhetorical resource that enables writers to build arguments and develop discourse effectively. These two research traditions—viewing syntax as a predictor and viewing it as a literacy practice—are not mutually exclusive. Rather, they form complementary conceptual foundations for this study: if syntax is both a predictor of writing ability and an active literacy practice in text production, then instruction that explicitly integrates syntax into the writing process has the

potential to produce more functional and sustainable improvement.

The theoretical foundation for integrating syntax instruction into writing pedagogy can be traced to the tradition of form-focused instruction (FFI)—a pedagogical approach that directs attention to linguistic form within the context of meaningful language use (Long, 1991). Unlike explicit grammar teaching detached from communicative practice, FFI holds that form awareness develops most effectively when learners are engaged in tasks that demand authentic language use (Ellis, 2024).

A number of recent empirical studies illustrate this position, though each also leaves an identifiable gap that motivates the present study. Xie and Yeung (2022), studying 184 Hong Kong primary-school L2 learners over one academic year, showed that syntactic awareness and reading comprehension mutually predict each other's development over time; however, their design was correlational and developmental rather than instructional, leaving open whether a deliberately designed intervention would produce comparable gains. Marjokorpi (2023) found, in a study of Finnish L1 secondary students, that grammatical understanding developed within the context of text production correlated more strongly with writing skill than grammatical knowledge acquired declaratively—a finding that supports integration in principle but was again observational rather than experimental.

More directly relevant, Jaya et al. (2025) conducted a quasi-experimental study in which Indonesian undergraduate EFL students received explicit sentence-combining instruction; the experimental group showed significantly greater gains in syntactic complexity and accuracy than a control group receiving conventional instruction. Similarly, Khalaf (2025) tested a syntax-focused instructional technique with EFL intermediate-school students using a quasi-experimental design and found significant writing-achievement gains for the treatment group. Both studies provide encouraging quasi-experimental evidence that syntax-focused intervention can improve writing outcomes. However, both were conducted in a general EFL context with a morphosyntactic structure that is typologically simpler than Arabic. Relatively few studies have examined how such interventions interact with diglossic language situations or with curricula that formally separate the teaching of grammar and writing, as is the case with *Nahwu* and *Insyā'* in Arabic language education programs in Indonesia.

Furthermore, a meta-analysis by Wang et al. (2025) found that although the relationship between syntactic features and writing performance is generally weak in the aggregate, it varies meaningfully depending on student characteristics and how syntax is measured—suggesting that the effectiveness of syntax-focused instruction is likely to be context-dependent rather than universal. In addition, Zhou & Lü (2022) showed that FFI significantly increased syntactic complexity in L2 Chinese writing, indicating that similar effects may hold across languages with complex morphosyntactic systems.

However, their study, like the others reviewed here, did not extend to Arabic. This principle underlies the argument that integrating syntax into writing instruction is more productive than teaching *Nahwu* as a stand-alone course. Nevertheless, research on grammar–writing integration remains dominated by English-as-a-foreign or second-language contexts, and experimental evidence

from languages with more complex morphosyntactic systems, including Arabic, remains very limited (Awang et al., 2025).

This challenge is felt even more acutely in the teaching of Arabic as a foreign language in non-Arab contexts. In Indonesia—home to the world's largest Muslim population and a large number of Arabic Language Education programs in higher education—students are required to master complex syntactic structures (*Nahwu*) while simultaneously developing academic writing competence (*Insyā'*) (Suja, Muassomah, et al., 2025). Arabic exhibits a close interdependence between morphology and syntax: an error at the morphological level, such as in case marking (*i'rāb*) or gender agreement (*muthābaqah*), directly affects the meaning and grammatical acceptability of a sentence. This means that syntactic mastery in Arabic cannot be separated from morphological mastery, and both must be developed in an integrated manner through text-production practice (Awang et al., 2025).

This complexity is compounded by the phenomenon of diglossia, a defining feature of the Arabic linguistic situation. Ferguson (1959) defined diglossia as a condition in which two functionally distinct language varieties—a high variety (*al-fuṣḥā*) and a low variety (*al-'āmmiyyah*)—are used concurrently within a single speech community. In the context of learning Arabic as a foreign language, diglossia creates a distinct challenge: non-native learners must not only master the grammatical system of formal Arabic, but also navigate the considerable distance between the Arabic they learn in the classroom and the spoken varieties they may encounter outside it. Safrullah et al. (2022) found that this condition makes it more difficult for learners to internalize and apply syntactic rules in extended written text consistently.

Prevailing instructional practice, however, still maintains a sharp separation between these two domains: syntactic rules are taught as decontextualized formal knowledge, while writing is practiced without systematic attention to syntactic integration (Fahrurrozi, 2014; Hidayat, 2020). As a result, students often understand syntactic rules theoretically but are unable to apply them consistently in extended written text (Prihantoro, 2019; Safrullah et al., 2022). This condition is compounded by the heterogeneity of students' educational backgrounds (Batalipu et al., 2026). This instructional gap reflects a broader tendency in foreign-language grammar pedagogy: most existing research treats syntax solely as a predictor of writing and reading ability (Brimo et al., 2017; Ishak, 2018; Marjokorpi & van Rijt, 2025), rather than as a resource that can be actively integrated into the writing process itself—even though international research in other foreign-language contexts consistently shows that integrating grammar into writing instruction is more effective than a separated approach (Xie & Yeung, 2022; Marjokorpi, 2023).

The literature synthesized above points to three interrelated gaps. First, a theoretical gap: syntax continues to be positioned predominantly as a predictor or correlate of writing ability rather than as a resource that can be directly integrated into the writing-instruction process itself. Second, a methodological gap: existing studies on grammar-writing integration are largely correlational, and experimental or quasi-experimental research testing this relationship remains limited—particularly for languages with more complex morphosyntactic systems such as Arabic (Awang et al., 2025). Third, a contextual gap: research on the integration of *Nahwu* and *Insyā'* within Indonesian Islamic higher education remains scarce. Existing research on Arabic as a foreign language in this context

has focused primarily on teachers, teaching methods, or instructional materials (Suja, Muassomah, et al., 2025; Suja, Syarifaturrahmatullah, et al., 2025; Sutaman et al., 2026) rather than on syntax-writing integration itself, despite the context's distinctive features—diglossia, heterogeneous student syntactic backgrounds, and a strong curricular separation between *Nahwu* and *Insyā'*—that remain empirically unexplored.

This study fills these gaps by examining the effectiveness of a syntax-integration model in Arabic-as-a-foreign-language writing instruction at three Indonesian Islamic higher-education institutions. Drawing on a sociocognitive framework that views writing as an interaction between linguistic knowledge, cognitive strategies, and social context (Nishino & Atkinson, 2015), this study positions syntax instruction not as a prerequisite for writing but as an integral component of the writing process itself. The intervention comprised authentic-text modeling, sentence- and paragraph-reconstruction exercises, peer-based collaborative editing, and explicit syntactic feedback—practices designed to help learners transition their syntactic knowledge from the declarative to the procedural domain. This approach not only assesses the relationship between syntax and writing but also intervenes directly in the learning process to examine how syntax functions as a cognitive and rhetorical tool in the production of academic Arabic text, responding to calls for more contextually grounded empirical evidence (Awang et al., 2025; Wang et al., 2025).

Two central research questions guided this investigation, beginning with an examination of whether integrating syntax instruction into Arabic writing instruction produces a significantly greater improvement in students' writing ability than conventional nonintegrated instruction. Additionally, the research analyzed the pattern of writing ability improvement from pretest to posttest within both the experimental and control groups, while evaluating the extent to which the proportion of students who improved differed between the two groups.

Regarding its contributions, the research enriches the existing literature by repositioning syntax as a literacy practice that plays a direct role in meaning construction in Arabic writing rather than merely a predictor of writing ability. It also introduces an integrated instructional model empirically tested in the Indonesian Islamic higher education context, a setting that remains rarely studied experimentally. Finally, the work offers an initial empirical basis for developing a more contextually grounded Arabic curriculum, while emphasizing that generalization to other contexts requires further replication.

## 2. METHODS

This study employed a quasi-experimental pretest–posttest control-group design. This design was chosen because it allows for a systematic comparison between two instructional conditions—an experimental group receiving syntax integration and a control group receiving conventional instruction—before and after the intervention. In higher-education settings, random assignment to groups is generally not feasible because classes are already institutionally formed; a quasi-experimental design is therefore the most appropriate and commonly used choice for pedagogical research in such settings (Creswell & Creswell, 2023). The research design is summarized in Table 1.

Table 1. *Research Design*

| Group        | Pretest | Treatment | Posttest |
|--------------|---------|-----------|----------|
| Experimental | O1      | X         | O2       |
| Control      | O3      | —         | O4       |

*Note: O1 = experimental-group pretest; O2 = experimental-group posttest; O3 = control-group pretest; O4 = control-group posttest; X = implementation of the syntax-integration model in Arabic writing instruction.*

### **Participants**

This study involved 108 students, comprising 53 students in the experimental class and 55 students in the control class. All participants were students in Arabic Language Education programs at three universities in Indonesia representing three different geographic regions: Java, Sumatra, and Kalimantan. The three institutions were selected on the basis of curricular similarity—all offer syntax (*Nahwu*) and writing (*Insyā'*) courses as compulsory components of the study program—allowing for a more equivalent comparison across institutional contexts.

The sampling technique used was cluster sampling based on naturally formed (intact) classes, without altering the existing academic structure. All students in the sample had completed a basic *Nahwu* course and had relatively equivalent baseline competence in fundamental syntactic structures, as indicated by the pretest results and the variance-homogeneity test reported in the Findings section. All participants, both the instructors and the students, consented to the collection of their data for this research. Consent was obtained through a written request to instructors and students so that the classes they taught or attended could be used as a source of research data.

### **Instructional Intervention**

The intervention was administered systematically to the experimental group through a syntax-integration approach to the writing process. It was implemented over 8 weeks, with one 100-minute session per week. The intervention comprised four interrelated core components. The first component was syntactic-structure modeling through authentic texts, in which students closely observed how syntactic structure functions to convey meaning and build cohesion in academic Arabic texts. The second component was sentence- and paragraph-transformation practice, in which students converted simple word or sentence structures into complex sentences and organized them into coherent paragraphs. The third component was collaborative peer-editing, in which students evaluated their peers' writing with a focus on identifying syntactic errors and understanding the relationship between sentence structure and text comprehensibility. The fourth component was explicit syntactic feedback, provided by the instructor on each draft of students' writing, allowing them to see directly how linguistic structure affected the quality of their text.

Table 2 summarizes how these components were sequenced across the eight intervention weeks, mapped onto the specific *Nahwu* patterns targeted at each stage, drawn from the course module used in this study. The weekly grouping reflects the module's internal sequencing from simpler to more complex patterns; actual pacing was adjusted as needed to student progress.

Table 2. *Weekly Intervention Schedule for the Experimental Group*

| Week | Nahwu Pattern(s)<br>Covered                                                                                                                        | Writing Material /<br>Focus                                               | Integration Activity                                                                                                                                              |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | <i>Jumlah fi'liyah with zarf al-makān, zarf al-zamān, and combined place+time adverbials</i>                                                       | Simple verbal sentences describing daily actions with place/time settings | Authentic-text modeling of place/time-adverbial sentences; guided sentence completion; combining sentences with <i>tsumma/ba'da dzālika</i> into short paragraphs |
| 2    | <i>Jumlah fi'liyah with maf'ūl bih (object), alone and combined with place/time adverbials</i>                                                     | Transitive-verb sentences (verb–subject–object, with optional adverbials) | Sentence-transformation practice (simple → object-bearing sentences); peer-editing focused on verb–object agreement and adverbial placement                       |
| 3    | <i>Jumlah ismiyah with mubtada' as demonstrative (ism al-ishārah) or pronoun (ḍamīr) + khabar as isim</i>                                          | Introductory nominal sentences (this/that is...; he/she is...)            | Authentic-text modeling of nominal-sentence structure; explicit feedback on <i>mibtada'</i> – <i>khabar</i> gender/number agreement                               |
| 4    | <i>Jumlah ismiyah with mubtada' as definite noun (isim ma'rifah) or construct phrase (muḍāf–muḍāf ilayh) + khabar as isim</i>                      | More complex nominal-subject constructions                                | Sentence-transformation practice converting simple nominal subjects into definite-noun/construct-phrase subjects; peer-editing                                    |
| 5    | <i>Jumlah ismiyah with khabar as a quasi-sentence (syibh al-jumlah / jār–majrūr), across pronoun, definite-noun, and construct-phrase subjects</i> | Nominal sentences with prepositional-phrase predicates                    | Modeling and guided practice converting <i>isim-khabar</i> sentences into <i>syibh-al-jumlah-khabar</i> sentences; explicit feedback                              |

| Week | Nahwu Pattern(s)<br>Covered                                                                                                              | Writing Material /<br>Focus                                              | Integration Activity                                                                                                                                                                                 |
|------|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6    | <i>Jumlah ismiyah</i> with inverted word order, <i>khobar</i> preceding <i>mubtada'</i>                                                  | Stylistic variation in nominal-sentence word order                       | Collaborative peer-editing comparing standard vs. inverted word order for cohesion/emphasis; paragraph-combination practice                                                                          |
| 7    | <i>Jumlah ismiyah</i> with <i>khobar</i> as a full <i>jumlah fi'liyah</i> (pronoun or definite-noun subject + verbal-sentence predicate) | Complex nominal sentences embedding a verbal clause as predicate         | Sentence- and paragraph-transformation practice combining <i>jumlah ismiyah</i> and <i>jumlah fi'liyah</i> structures within one paragraph; explicit feedback on structural integration              |
| 8    | Integrated review                                                                                                                        | Free paragraph composition drawing on the full range of patterns covered | Authentic-text modeling of a complete paragraph; independent composition under guided prompts; final peer-editing and instructor feedback cycle, mirroring the study's pretest/posttest writing task |

As shown in Table 2 above, the experimental group progressed from basic verbal sentences (*jumlah fi'liyah*) with adverbials and direct objects during the first two weeks to nominal structures (*jumlah ismiyah*) featuring complex subjects and prepositional predicates across weeks three through five. The final stages introduced advanced stylistic variations, such as inverted word order and embedded clauses, culminating in an integrated review and free paragraph composition. Throughout these stages, students actively engaged in authentic text modelling, sentence transformation, peer editing, and explicit instructor feedback, whereas the control group received conventional composition (*Insyā'*) instruction centered on general writing practice without explicit syntactic scaffolding.

### **Instrument and Validation**

Students' writing ability was assessed using an analytic scoring rubric measuring five main aspects: (1) syntactic accuracy, (2) vocabulary accuracy, (3) inter-paragraph cohesion and coherence, (4) sentence-structure complexity, and (5) overall text readability. The rubric was developed based on the academic literature on foreign-language writing assessment. It was reviewed

by an Arabic linguistics expert to ensure the substantive and constructive adequacy of the instrument.

Empirical validity was tested using the corrected item-total correlation. The results showed that all indicators had correlation values exceeding the minimum threshold, indicating that each indicator had adequate correlational strength and could accurately measure the writing-ability construct (see Table 2). Reliability testing using Cronbach's alpha yielded  $\alpha = .962$ , far exceeding the minimum threshold of .70 (see Table 3).

The writing task used for the pretest and post-test was a guided paragraph-writing task modelled on the exercise patterns in the teaching material. This material develops writing ability progressively, from arranging words into sentences, completing paragraphs, forming sentences from prompt words, and joining sentences using connectors such as *tsumma*, *ba'da dzālika*, and *wa*, through to independent composition. On both tests, students were asked to produce a short Arabic composition of at least one paragraph consisting of a minimum of five interrelated sentences, using a variety of *jumlah fi'liyah* (verbal sentences) and *jumlah ismiyah* (nominal sentences)—including combinations of *fi'l-fā'il* (verb-subject), *maf'ul bih* (object), *zarf al-makān/al-zamān* (adverbial of place/time) or *jār-majrūr* (prepositional phrase), as well as *muftada'-khabar* (subject-predicate) patterns realized as *khabar mufrad* (single-word predicate), *syibh al-jumlah* (quasi-sentence), and *jumlah fi'liyah* (verbal-sentence predicate).

The pretest and post-test prompts differed but were designed to be parallel. On the pretest, students wrote a paragraph about their daily activities on a class day, while on the post-test they wrote about their daily activities on a day off. The two prompts were equated in terms of task type, required writing length, range of syntactic structures, content demands, topic familiarity for students, and time allotted.

The task was completed individually in class within 60 minutes. Students were permitted to use a printed Arabic-Indonesian dictionary. However, they were not permitted to use automatic translators, artificial-intelligence applications, digital grammar checkers, the internet, class notes, or assistance from other participants. Identical conditions, time allocation, and task instructions were applied to both the experimental and control groups, and at both measurement points. All writing samples were scored using the five-dimension analytic rubric described above.

### **Data Analysis**

Data analysis was conducted in four stages. First, a Shapiro-Wilk normality test was used to determine the data distribution. Second, a homogeneity test using Levene's test examined the equality of variances between groups. Third, because the normality test indicated a non-normal distribution ( $p < .05$ ), within-group differences were analyzed using the Wilcoxon signed-rank test, and between-group differences were analyzed using the Mann-Whitney U test applied to the gain scores. Fourth, effect size was calculated using the rank-biserial correlation ( $r$ ) with the formula  $r = Z / \sqrt{N}$ , interpreted according to Cohen's (1988) convention:  $r = .10$  (small),  $r = .30$  (medium),  $r = .50$  (large). All analyses were conducted using SPSS version 27.

### 3. RESULTS AND DISCUSSION

#### 3.1 Result

This section presents the results of the data analysis in sequence: instrument validity and reliability, descriptive statistics, statistical assumption testing, and hypothesis testing addressing both research questions. The first question is addressed through the comparison of gain scores between groups, while the second question is addressed through an analysis of the improvement pattern within each group.

##### *Instrument Validity and Reliability*

The validity test results show that all indicators on the Arabic writing-ability scoring rubric had corrected item-total correlation values exceeding the r-table value. Full results are presented in Table 3 and Table 4.

Table 3. *Validity Test Results for the Arabic Writing-Ability Rubric*

| Dimension              | Parameter | R Table | R Count | Note  |
|------------------------|-----------|---------|---------|-------|
| Sentence Structure     | Y1        | 0.361   | 0.961   | Valid |
| Vocabulary Mastery     | Y2        | 0.361   | 0.955   | Valid |
| Grammar                | Y3        | 0.361   | 0.885   | Valid |
| Cohesion and Coherence | Y4        | 0.361   | 0.954   | Valid |
| Readability            | Y5        | 0.361   | 0.919   | Valid |

Table 4. *Reliability Test Results for the Arabic Writing-Ability Rubric*

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .962             | 5          |

As presented in Table 3 and Table 4 above, the Arabic writing ability rubric demonstrates robust construct validity and exceptional internal consistency. The validity test results indicate that all five dimensions achieved calculated correlation values ranging between 0.885 and 0.961, substantially exceeding the critical threshold value of 0.361. Additionally, the reliability analysis yielded a Cronbach's Alpha coefficient of 0.962 across all five items, confirming that the assessment instrument is statistically valid and highly reliable for evaluating writing performance.

##### *Descriptive Statistics*

Descriptive statistics for the writing-ability scores of both groups are presented in Table 5. Descriptively, the experimental class had a slightly higher mean pretest score ( $M = 80.23$ ,  $SD = 9.40$ ) than the control class ( $M = 76.45$ ,  $SD = 7.86$ ); this initial difference is examined further in the Initial Group Equivalence section below. On the posttest, the experimental class achieved a mean of  $M = 90.72$  ( $SD = 8.08$ ), while the control class achieved  $M = 83.44$  ( $SD = 6.03$ ). The mean gain score (the difference between posttest and pretest) was 10.49 ( $SD = 6.92$ ) for the experimental class, higher than the control class's 6.98 ( $SD = 6.00$ ).

Table 5. *Descriptive Statistics for Writing-Ability Scores*

| Group        | Measurement | N  | Mean  | SD   | Min | Max |
|--------------|-------------|----|-------|------|-----|-----|
| Experimental | Pretest     | 53 | 80.23 | 9.40 | 65  | 97  |
| Experimental | Posttest    | 53 | 90.72 | 8.08 | 67  | 99  |
| Experimental | Gain Score  | 53 | 10.49 | 6.92 | —   | —   |
| Control      | Pretest     | 55 | 76.45 | 7.86 | 63  | 91  |
| Control      | Posttest    | 55 | 83.44 | 6.03 | 66  | 98  |
| Control      | Gain Score  | 55 | 6.98  | 6.00 | —   | —   |

**Normality Test**

The Shapiro–Wilk normality test results show that none of the study variables were normally distributed. In the experimental class, the pretest score yielded a Shapiro–Wilk statistic of .951 with  $p = .031$ . In contrast, the posttest score yielded a statistic of .875 with  $p = .000$ . A similar pattern was found in the control class, with a pretest statistic of .942 ( $p = .013$ ) and a posttest statistic of .950 ( $p = .028$ ). Because all  $p$ -values were below .05, the normality assumption was not met for any of the four variables, and subsequent analyses therefore used a nonparametric approach. Full results are presented in Table 6.

Table 6. *Shapiro–Wilk Normality Test Results*

| Variable                     | Statistic | df | Sig. |
|------------------------------|-----------|----|------|
| Pretest, Experimental Class  | .951      | 53 | .031 |
| Posttest, Experimental Class | .875      | 53 | .000 |
| Pretest, Control Class       | .942      | 55 | .013 |
| Posttest, Control Class      | .950      | 55 | .028 |

*a. Lilliefors Significance Correction***Homogeneity Test**

The variance-homogeneity test using Levene's test showed that the pretest score variances of the experimental and control classes were homogeneous. Significance values from all four computational approaches were above the critical threshold of .05, indicating no significant difference in variance between the two groups. Full results are presented in Table 7.

Table 7. *Levene's Test of Homogeneity of Variance*

| Basis                         | Levene Statistic | df1 | df2     | Sig. |
|-------------------------------|------------------|-----|---------|------|
| Based on Mean                 | 2.896            | 1   | 106     | .092 |
| Based on Median               | 2.331            | 1   | 106     | .130 |
| Based on Median (adjusted df) | 2.331            | 1   | 104.274 | .130 |
| Based on Trimmed Mean         | 2.997            | 1   | 106     | .086 |

**Initial Group Equivalence**

The homogeneity test in the previous section tested only the equivalence of pretest variances between groups, not the equivalence of their central tendencies. To confirm that both groups had equivalent baseline writing ability before the intervention, a Mann-Whitney U test was conducted on the pretest scores. The experimental class obtained a mean rank of 60.50, while the control class obtained a mean rank of 48.72. The test yielded  $U = 1775.500$  and  $Z = 1.954$ ,  $p = .051$ , with an effect

size of  $r = .19$  (small effect). For comparison, an independent-samples t-test on the same data yielded  $t = 2.265$ ,  $p = .026$ . Full results are presented in Table 8.

The two tests yield different conclusions right around the conventional significance threshold of  $\alpha = .05$ : the nonparametric test (which is appropriate given this study's data distribution) shows a marginal, non-significant difference ( $p = .051$ ), while the parametric test shows a significant difference ( $p = .026$ ). The experimental class's pretest score ( $M = 80.23$ ) is, descriptively, somewhat higher than the control class's ( $M = 76.45$ ). Because these results cannot conclusively establish baseline equivalence between the two groups, the possibility that this baseline difference contributed to the observed gain-score difference should be considered an additional limitation in interpreting this study's results. Though it is also worth noting that the experimental class started from a higher baseline yet still showed a larger gain score, a pattern that, given the potential ceiling effect associated with a higher starting score, would, if anything, tend to work against this study's hypothesis rather than in its favor.

Table 8. *Pretest Equivalence Test Results (Mann-Whitney U Test)*

| Group        | N  | Mean Rank | Test Statistic         | Value              |
|--------------|----|-----------|------------------------|--------------------|
| Experimental | 53 | 60.50     | Mann-Whitney U         | 1775.500           |
| Control      | 55 | 48.72     | Z                      | 1.954              |
|              |    |           | Asymp. Sig. (2-tailed) | .051               |
|              |    |           | Effect Size (r)        | .19 (small effect) |

#### ***Within-Group Pretest-Posttest Comparison***

To address the second research question, a Wilcoxon signed rank test was conducted to evaluate the writing ability gains within each cohort, as summarized in Table 9.

Table 9. *Wilcoxon Signed-Rank Test Results*

|                                     | N         | Mean Rank | Sum of Ranks |
|-------------------------------------|-----------|-----------|--------------|
| Experimental Class — Negative Ranks | 3         | 12.33     | 37.00        |
| Experimental Class — Positive Ranks | 50        | 27.88     | 1394.00      |
| Experimental Class — Ties           | 0         | —         | —            |
| <b>Total (Experimental)</b>         | <b>53</b> |           |              |
| Control Class — Negative Ranks      | 7         | 12.43     | 87.00        |
| Control Class — Positive Ranks      | 47        | 29.74     | 1398.00      |
| Control Class — Ties                | 1         | —         | —            |
| <b>Total (Control)</b>              | <b>55</b> |           |              |

Table 10. *Wilcoxon Signed-Rank Test Results*

| Test Statistic                | Experimental Class | Control Class      |
|-------------------------------|--------------------|--------------------|
| <b>Z</b>                      | -6.010             | -5.648             |
| <b>Asymp. Sig. (2-tailed)</b> | .000               | .000               |
| <b>Effect Size (r)</b>        | .83 (large effect) | .76 (large effect) |

As detailed in Table 9 and Table 10, the statistical results indicate significant positive progress across both groups, with the experimental class demonstrating a higher rate of overall improvement. In the experimental group, 50 out of 53 students (94.3%) achieved increased post-test scores (mean rank = 27.88; sum of ranks = 1394.00), whereas only 3 students showed a decrease (mean rank = 12.33; sum of ranks = 37.00). This gain was statistically significant ( $Z = -6.010$ ,  $p = .000$ ) with a large effect size ( $r = .83$ ).

In the control group, 47 out of 55 participants (85.5%) increased their scores (mean rank = 29.74; sum of ranks = 1398.00), while 7 students showed a decrease (mean rank = 12.43; sum of ranks = 87.00) and 1 student recorded no change. While the control group also improved significantly ( $Z = -5.648$ ,  $p = .000$ ,  $r = .76$ ), the experimental condition yielded a higher proportion of advancing students (94.3% versus 85.5%) as well as a larger overall effect size.

### ***Between-Group Gain Score Comparison***

To address the first research question, the Mann-Whitney U test was used to compare gain scores between the experimental and control classes. The experimental class obtained a mean rank of 63.05 (sum of ranks = 3341.50), while the control class obtained a mean rank of 46.26 (sum of ranks = 2544.50). Statistically, the Mann-Whitney U test yielded  $U = 1004.500$  and  $Z = -2.788$ ,  $p = .005$  ( $p < .05$ ), confirming that the gain-score difference between the two groups was statistically significant—that is, unlikely to have arisen by chance. Statistical significance on its own, however, indicates only that an effect is unlikely to be due to chance; it does not indicate the size of that effect in practical terms, so the two must be interpreted separately. The effect size was  $r = .27$ , which, following Cohen's (1988) convention ( $r = .10$  small,  $.30$  medium,  $.50$  large), falls within the small-to-moderate range—closer to the medium end of that range than the small end, though still below the conventional threshold for a medium effect. This indicates that, despite the statistically significant  $p$ -value, the practical magnitude of the experimental group's advantage over the control group, while real, was moderate rather than large. Full results are presented in Table 11.

Table 11. *Mann-Whitney U Test Results*

|                        | Class        | N          | Mean Rank                             | Sum of Ranks |
|------------------------|--------------|------------|---------------------------------------|--------------|
| Gain Score             | Experimental | 53         | 63.05                                 | 3341.50      |
|                        | Control      | 55         | 46.26                                 | 2544.50      |
| <b>Total</b>           |              | <b>108</b> |                                       |              |
| Statistic              |              |            | Gain Score                            |              |
| Mann-Whitney U         |              |            | 1004.500                              |              |
| Wilcoxon W             |              |            | 2544.500                              |              |
| Z                      |              |            | -2.788                                |              |
| Asymp. Sig. (2-tailed) |              |            | .005                                  |              |
| <b>Effect Size (r)</b> |              |            | <b>.27 (small-to-moderate effect)</b> |              |

## 3.2 Discussion

### ***Effectiveness of the Syntax Integration Model***

With respect to the first research question, the Mann-Whitney U test results show that the experimental class's gain score was statistically significantly different from the control class's ( $Z = -2.788$ ,  $p = .005$ ), with a rank-biserial correlation effect size of  $r = .27$ , within the small-to-moderate range according to Cohen's (1988) convention. This finding indicates that the experimental group showed statistically significantly greater writing-ability improvement than the control group. However, the magnitude of this between-group advantage was small to moderate rather than large, and should accordingly be read as a modest rather than substantial practical difference.

It is important to interpret this finding proportionately. Both groups showed statistically significant improvement with large effect sizes on the Wilcoxon test (experimental:  $r = .83$ ; control:  $r = .76$ ). This indicates that general learning factors—exposure to writing tasks, instructor feedback, and repeated practice—contributed significantly to improvement in both groups. The experimental class's advantage over the control class, as reflected in the small-to-moderate between-group effect size ( $r = .27$ ), indicates that the added value of syntactic integration is real but moderate. This finding is consistent with the literature showing that grammar-based intervention effects in quasi-experimental contexts generally produce medium effect sizes (Brown et al., 2026).

### ***Improvement Patterns Within Each Group***

In response to the second research question, the comparison of Wilcoxon patterns between the two groups shows that the proportion of students who improved was higher in the experimental class (94.3%) than in the control class (85.5%). This difference indicates that the syntax-integration model produced more widespread improvement among students. It should be emphasized, however, that the analysis in this study was conducted on total rubric scores rather than at the level of specific morphosyntactic structures. While the intervention and writing task targeted particular *Nahwu* structures—including *i'rāb*, *jumlah fi'liyah*, and *jumlah ismiyah* (see Table 2 and the Instructional Intervention section)—students' writing was not coded or scored for accuracy on these structures individually. Identifying which specific structures showed the greatest improvement would require a structure-level error analysis of students' texts, which was beyond the scope of the current rubric-based design and therefore cannot be confirmed from the available data.

### ***Theoretical Implications***

Theoretically, this finding is consistent with the sociocognitive framework that views writing as an interaction between linguistic knowledge, cognitive strategies, and social context (Nishino & Atkinson, 2015). More specifically, this finding can be interpreted through the lens of skill acquisition theory (DeKeyser & Suzuki, 2025), which explains foreign-language learning as a gradual transition from declarative knowledge—knowing the rules—to procedural knowledge—the ability to use those rules automatically in text production. The intervention designed for this study operationally created conditions intended to support the proceduralization of syntactic knowledge: by design, students were required to repeatedly apply *Nahwu* rules in the context of authentic text production rather than rely on memorization alone. This study, however, measured writing-ability outcomes only and did not directly assess the underlying learning process; claims about proceduralization itself, as

distinct from its instructional design rationale, would require process-level evidence—such as think-aloud protocols or response-time measures—beyond the scope of this study.

This finding also supports the form-focused instruction (FFI) position put forward by Ellis (2024), namely that attention to linguistic form develops most effectively when integrated into tasks demanding meaningful language use. The instructional model in this study operationalizes FFI principles by situating syntax instruction within the context of authentic academic writing. This is also consistent with Zhou and Lü's (2022) findings, who showed that FFI significantly increased syntactic complexity in L2 writing, indicating that similar effects may apply across languages with complex morphosyntactic systems.

This study also contributes to framing syntax as a literacy practice (Pölc, 2022). The writing-ability improvement observed in the experimental class shows that when syntax is positioned as a rhetorical and cognitive tool within the writing process, students can develop a greater capacity to organize ideas and build discourse effectively. The traditional separation between *Nahwu* and *Insyā'* in the Arabic-education curriculum in Indonesia thus appears to be pedagogically less productive (Prihantoro, 2019; Safrullah et al., 2022).

### ***Illustrative Evidence from Student Writing***

To illustrate how this shift from declarative to procedural knowledge manifested at the clause level, Table 12 presents matched pretest–posttest data from experimental-group students, selected to show patterns of change consistent with the sentence-and-paragraph reconstruction component of the intervention.

Table 12. *Illustrative Pretest–Posttest Clause-Level Changes (Experimental Group)*

| Pretest Excerpt                                                                                                          | Posttest Excerpt                                                                                          | Change                                            |
|--------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|---------------------------------------------------|
| إِلَى الْمَكْتَبَةِ<br>("to the library," accusative/nominative marking used where the genitive is required after إِلَى) | إِلَى الْمَدْرَسَةِ<br>("to school," correct genitive marking)                                            | Case marking ( <i>i'rāb</i> ) after preposition   |
| أَنَا تَقْرَأُ<br>("I" + 2nd/3rd-person verb form, person mismatch)                                                      | أَقْرَأُ<br>(used consistently with other correctly conjugated 1st-person verbs throughout the paragraph) | Subject–verb person agreement                     |
| فِي الْمَطْبَخِ<br>("in the kitchen," nominative marking used where the genitive is required after فِي)                  | فِي الْمَطْبَخِ<br>(correct genitive marking)                                                             | Case marking ( <i>i'rāb</i> ) after a preposition |
| أَنَا يَذْهَبُ<br>("I" + 3rd-person-masculine verb form, person mismatch)                                                | أَذْهَبُ<br>(correct 1st-person form, no redundant/mismatched pronoun)                                    | Subject–verb person agreement                     |

| Pretest Excerpt                                                                                               | Posttest Excerpt                             | Change                                             |
|---------------------------------------------------------------------------------------------------------------|----------------------------------------------|----------------------------------------------------|
| إِلَى السُّوقِ<br>("to the market," nominative marking<br>used where the genitive is required<br>after إِلَى) | إِلَى السُّوقِ<br>(correct genitive marking) | Case marking ( <i>i'rāb</i> )<br>after preposition |
| أَنَا تَدْرُسُ<br>("I" + 2nd-person-masculine/3rd-<br>person-feminine verb form, person<br>mismatch)          | أَدْرُسُ<br>(correct 1st-person form)        | Subject-verb person<br>agreement                   |

As illustrated in Table 12, across all cases, two recurring error types present at pretest were substantially reduced at posttest. The first was case marking (*i'rāb*) following a preposition, corrected in each of the examples shown. The second was subject-verb person agreement, in which the first-person pronoun *anā* ("I") was combined with a second- or third-person verb form at pretest; at posttest, the students used correctly conjugated first-person verbs consistently throughout the paragraph, without the mismatched pronoun. Both error types involve rules that are explicitly taught early in *Nahwu* instruction and that students could, in principle, state correctly if asked in isolation; their frequent appearance in pretest paragraph writing is consistent with the declarative-procedural gap described in the Introduction, in that students appeared to know the rules but could not yet apply them reliably under the real-time demands of continuous text production. Their reduced frequency at posttest, following repeated practice transforming and combining sentences into paragraphs—the specific focus of the sentence-and-paragraph reconstruction component—is consistent with, though does not on its own prove, progress toward proceduralization in DeKeyser and Suzuki (2025) sense. As with the claims discussed above, this illustrative analysis is based on a small, purposively selected set of excerpts rather than a systematic structure-level coding of the full dataset, and should be read as suggestive rather than confirmatory evidence.

### Limitations

Several important limitations must be acknowledged in order to situate this study's findings accurately within the broader literature. This study used a quasi-experimental design with intact classes and without random assignment. Although the homogeneity test showed equivalent pretest variances between groups, the test of central-tendency equivalence produced results right at the conventional significance threshold (Mann-Whitney U,  $p = .051$ ; t-test,  $p = .026$ ), with the experimental class starting from a slightly higher mean pretest score; the possibility of unmeasured pre-treatment variables—such as learning motivation or the intensity of independent study—or the baseline difference itself cannot be fully ruled out as a threat to internal validity. Relatedly, the lecturer effect could not be fully controlled, since differences in the quality or experience of the instructors teaching the experimental and control classes may have contributed to the observed differences in outcomes, and a possible Hawthorne effect should also be considered, as the experimental class may have performed better partly because of an awareness that they were receiving special treatment rather than solely because of the intervention's effectiveness. Future

research should consider standardizing instructor training or explicitly controlling for the instructor variable to address these concerns.

In terms of measurement, the analysis was conducted on total rubric scores rather than at the dimension level, which limits the study's ability to identify which specific aspects of writing were most affected by the intervention. The small-to-moderate between-group effect size ( $r = .27$ ) further indicates that, although the gain-score difference was statistically significant, its practical significance should be interpreted with caution, as factors outside the intervention design likely also contributed to the improvement observed. In addition, the very high Cronbach's alpha value ( $\alpha = .962$ ) for the five-dimension rubric should be read with caution: for an instrument intended to measure conceptually distinct aspects of writing ability (syntactic accuracy, vocabulary, cohesion, complexity, readability), an inter-dimension correlation this high ( $r = .885-.961$ ) raises the possibility of a halo effect in the scoring process—raters tending to assign similar scores across all dimensions for a given student rather than rating each dimension independently—rather than purely reflecting the instrument's strength. Future research should consider stricter inter-rater training or having different raters score each dimension separately to reduce this risk.

Finally, the pretest and posttest prompts used different topics, though designed to be parallel (class-day activities on the pretest; day-off activities on the posttest), and this order was applied consistently to all participants without counterbalancing. Because the topic order was the same for both the experimental and control groups, this potential bias does not threaten the validity of the between-group gain-score comparison (RQ1); however, some of the score improvement observed in both groups (RQ2) may have been influenced in part by differences in topic ease or familiarity between measurement points, rather than solely by learning that occurred.

## **Implications**

### **Implications for Teaching Practice**

At the classroom level, this finding indicates that syntax instruction tends to be more productive when integrated into meaningful writing activities than when taught as a stand-alone formal component. The four intervention components—authentic-text modeling, sentence-and-paragraph reconstruction, collaborative editing, and explicit syntactic feedback—appear to work together to create a learning experience in which students come to experience syntax not as rules to be memorized, but as a tool for organizing meaning in text.

The most immediate practical implication is that Arabic instructors might consider embedding syntax-based tasks into regular writing-instruction sessions, rather than waiting for students to complete a separate *Nahwu* course first. This study, however, did not compare different integration models against one another, so further investigation into the relative contribution of each intervention component is needed before more specific recommendations can be made.

### **Implications for Curriculum**

At the curricular level, this finding provides an initial empirical basis for reconsidering the conventional separation between *Nahwu* and *Insyā'* courses in Arabic Language Education programs. This study does not claim that this separation should be eliminated, or that the tested model is ready to be adopted as a new curricular framework. What can be suggested is that closer coordination

between syntax instruction and writing practice has the potential to help students connect structural knowledge with academic text production. This claim requires further empirical testing, including studies that explicitly analyze how differences in students' syntactic background moderate the effectiveness of an integrative approach.

### **Implications for Future Research**

This study opens up several questions that warrant deeper investigation. First, future studies should conduct a structure-level error analysis—for example, coding students' texts for accuracy in specific *Nahwu* structures such as *i'rāb*, *jumlah fi'liyah*, and *jumlah ismiyah*—to identify which specific structures are most responsive to integrated syntax instruction, a question this study's holistic rubric-based design could not address. Second, a delayed-posttest design is needed to assess the long-term stability of improvement. Third, research with a larger sample size and a true experimental design would yield stronger evidence regarding the causal relationship. Fourth, qualitative or mixed-methods studies exploring students' cognitive and rhetorical processes could provide deeper insight into how syntax integration works. Fifth, replicating this study in other language-teaching contexts with complex morphosyntactic systems would contribute to the generalizability of these findings within the broader foreign-language pedagogy literature.

## **4. CONCLUSION**

This study examined the effectiveness of a syntax-integration model in Arabic-as-a-foreign-language writing instruction through a multi-institutional quasi-experimental design involving 108 students from three universities in Indonesia. Based on Wilcoxon signed-rank test and Mann-Whitney U test analyses, both groups were found to show statistically significant improvement in writing ability. However, the experimental class showed a statistically higher gain score than the control class ( $Z = -2.788$ ,  $p = .005$ ,  $r = .27$ ), with an effect size in the small-to-moderate range.

This finding indicates that integrating syntax instruction into the writing process—through authentic-text modeling, sentence-and-paragraph reconstruction, collaborative editing, and explicit syntactic feedback—was associated with a measurable, though moderate, added value in writing-ability improvement compared with conventional instruction.

It should be emphasized that these findings come from a single study with a quasi-experimental design lacking random assignment, long-term follow-up measurement, and dimension-level analysis of writing ability. Generalizing these findings to other contexts should therefore be done with caution and requires further empirical replication. While acknowledging these limitations, this study offers an initial, empirically grounded, and contextually specific contribution toward understanding how syntax integration can function as a more functional pedagogical approach in Arabic-as-a-foreign-language writing instruction, particularly within Indonesia's multilingual higher-education settings.

## **DECLARATION OF AI-ASSISTED TECHNOLOGIES**

The authors use Claude AI to ensure readability and sentence coherence, DeepL to assist with translation, and Grammarly to correct grammar, punctuation, and writing style in order to improve the clarity of the text.

## FUNDING INFORMATION

The authors declare that no funding was received for this study.

## AUTHOR CONTRIBUTIONS STATEMENT

| Name of Author        | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|-----------------------|---|---|----|----|----|---|---|---|---|---|----|----|---|----|
| Aidillah Suja         | ✓ | ✓ |    | ✓  | ✓  | ✓ | ✓ | ✓ | ✓ | ✓ | ✓  | ✓  | ✓ |    |
| Syarifaturrahmatullah | ✓ | ✓ |    | ✓  | ✓  | ✓ | ✓ | ✓ | ✓ | ✓ | ✓  | ✓  | ✓ |    |
| Moh. Syaifudin        | ✓ | ✓ |    | ✓  | ✓  | ✓ | ✓ | ✓ | ✓ | ✓ |    | ✓  | ✓ |    |
| Renti Yasmar          | ✓ | ✓ | ✓  |    | ✓  | ✓ |   |   | ✓ | ✓ |    |    |   |    |
| Ade Destri Deviana    | ✓ | ✓ | ✓  |    | ✓  | ✓ |   |   | ✓ | ✓ |    |    |   |    |
| Achmad Tito Rusady    | ✓ | ✓ | ✓  |    | ✓  | ✓ |   |   | ✓ | ✓ |    |    |   |    |

C : Conceptualization

I : Investigation

Vi : Visualization

M : Methodology

R : Resources

Su : Supervision

So : Software

D : Data Curation

P : Project administration

Va : Validation

O : Writing - Original Draft

Fu : Funding acquisition

Fo : Formal analysis

E : Writing - Review & Editing

## CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

## REFERENCES

- Akbari, Z. (2014). The role of grammar in second language reading comprehension: Iranian ESP context. *Procedia - Social and Behavioral Sciences*, 98, 122–126. <https://doi.org/10.1016/j.sbspro.2014.03.397>
- Awang, N. A., Ali, N., Mustafa, Z., Nik Din, N. M., & Ismail, W. (2025). Bridging linguistic gaps: Syntax and morphology as key drivers in enhancing Arabic writing skills. *International Journal of Academic Research in Business and Social Sciences*, 15(11). <https://doi.org/10.6007/IJARBS/v15-i11/26838>
- Aziza, Y. U., & Sutanto, A. V. (2024). Hubungan kemampuan sintaksis dengan kemampuan literasi awal pada anak kelas 2 SD Negeri 1 Pilang Randublatung Blora. *Jurnal Terapi Wicara dan Bahasa*, 3(1), 167–175. <https://doi.org/10.59686/jtwb.v3i1.151>
- Batalipu, A., I Miolo, M., & Sarif, S. (2026). Learning Arabic in heterogeneous classrooms: Experiences of non-pesantren students in Indonesian Islamic higher education. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya*, 14(1), 81–96. <https://doi.org/10.23971/altarib.v14i1.11341>
- Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading*, 40(1), 57–74. <https://doi.org/10.1111/1467-9817.12050>
- Brown, D., Liu, Q., & Norouzian, R. (2026). Effectiveness of written corrective feedback in developing L2 accuracy: A Bayesian meta-analysis. *Language Teaching Research*, 30(3), 1357–1389. <https://doi.org/10.1177/13621688221147374>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Creswell, J. W., & Creswell, D. J. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
- DeKeyser, R., & Suzuki, Y. (2025). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (4th ed.). Routledge.
- Ellis, R. (2024). Rod Ellis's essential bookshelf: Focus on form. *Language Teaching*, 57(2), 246–261. <https://doi.org/10.1017/S026144482200012X>
- Fahrurrozi, A. (2014). Pembelajaran bahasa Arab: Problematika dan solusinya. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 1(2). <https://doi.org/10.15408/a.v1i2.1137>
- Ferguson, C. A. (1959). Diglossia. *WORD*, 15(2), 325–340. <https://doi.org/10.1080/00437956.1959.11659702>
- Hidayat, N. S. (2020). Problematika pembelajaran bahasa Arab. *Jurnal Naskhi: Jurnal Kajian Pendidikan dan Bahasa Arab*, 2(1), 40–58. <https://doi.org/10.47435/naskhi.v2i1.290>
- Ishak, M. S. (2018). Hubungan antara penguasaan sintaksis dan penguasaan nahwu terhadap kemampuan menulis teks bahasa Arab pada mahasiswa Jurusan Pendidikan Bahasa Arab IAIN Pontianak. *Alfazuna: Jurnal Pembelajaran Bahasa Arab dan Kebahasaaraban*, 3(1), 51–74. <https://doi.org/10.15642/alfazuna.v3i1.267>
- Jaya, S., Susyla, D., & Melati. (2025). Developing syntactic complexity and accuracy in EFL students' essays through sentence combining. *Indonesian Journal of Educational Development*, 6(2), 408–421. <https://doi.org/10.59672/ijed.v6i2.4802>
- Khalaf, J. R. (2025). The impact of syntax surgery technique on EFL intermediate school students' writing skill. *Journal of Language Studies*, 9(4, Part 2), 189–121. <https://doi.org/10.25130/Lang.9.4.P2.10>
- Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspectives* (pp. 39–52). John Benjamins. <https://doi.org/10.1075/sibil.2.07lon>
- Marjokorpi, J. (2023). The relationship between grammatical understanding and writing skills in Finnish secondary L1 education. *Reading and Writing*, 36(10), 2605–2625. <https://doi.org/10.1007/s11145-022-10405-z>
- Marjokorpi, J., & van Rijt, J. (2025). Grammatical understanding predicts reading comprehension in secondary-level students: Insights from a Finnish national survey. *Language and Education*, 39(4), 924–943. <https://doi.org/10.1080/09500782.2024.2376280>
- Nishino, T., & Atkinson, D. (2015). Second language writing as sociocognitive alignment. *Journal of Second Language Writing*, 27, 37–54. <https://doi.org/10.1016/j.jslw.2014.11.002>
- Pölcz, Á. (2022). A mondattani elemzés életszerűségéről. *Magyar Nyelvőr*, 146(2), 177–185. <https://doi.org/10.38143/Nyr.2022.2.177>
- Prihantoro, S. (2019). Analisis kesalahan bahasa pada taksonomi linguistik dalam penulisan insya'. *Al Mahāra: Jurnal Pendidikan Bahasa Arab*, 5(1), 41–62. <https://doi.org/10.14421/almahara.2019.051-03>

- Ryding, K. C. (2005). *A reference grammar of modern standard Arabic*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486975>
- Safrullah, D. Y., Sari, N. S. A., Julia, J., Ali, E. Y., & Widiawati, N. (2022). Enhancing students' understanding of Arabic syntax among high school students in Indonesia. *Cypriot Journal of Educational Sciences*, 17(3), 702–718. <https://doi.org/10.18844/cjes.v17i3.6876>
- Suja, A., Muassomah, M., Sutaman, S., Hamid, M. A., & Rosyidi, A. W. (2025). Examining the relationship between teachers' theoretical beliefs and classroom practices in teaching Arabic. *Australian Journal of Applied Linguistics*, 8(1), Article 102580. <https://doi.org/10.29140/ajal.v8n1.102580>
- Suja, A., Syarifaturrahmatullah, Anwar, K., Fachrurrazi, Khuluq, M., & Eka Retnaningsih, L. (2025). Islamic kindergarten teachers' beliefs about early Arabic learning in the Indonesian context: Advantaged learning, teachers' ability, and their methods in teaching Arabic. *Journal of Early Childhood Teacher Education*. Advance online publication. <https://doi.org/10.1080/10901027.2025.2581022>
- Sutaman, Suja, A., Syarifaturrahmatullah, Andaru Bahy, M. B., Shobirin, & Marajo. (2026). Cultural representation in Arabic textbooks: Unveiling intercultural gaps and pedagogical implications in Indonesian Islamic higher education. *Multicultural Education Review*, 18(3), 404–423. <https://doi.org/10.1080/2005615X.2026.2705691>
- Tong, X., Yu, L., & Deacon, S. H. (2025). A meta-analysis of the relation between syntactic skills and reading comprehension: A cross-linguistic and developmental investigation. *Review of Educational Research*, 95(3), 385–426. <https://doi.org/10.3102/00346543241228185>
- Wang, J., Kim, Y.-S. G., Lam, J. H. Y., & Leachman, M. A. (2025). A meta-analysis of relationships between syntactic features and writing performance and how the relationships vary by student characteristics and measurement features. *Assessing Writing*, 63, Article 100909. <https://doi.org/10.1016/j.asw.2024.100909>
- Xie, Q., & Yeung, S. S. (2022). Do vocabulary, syntactic awareness, and reading comprehension in second language facilitate the development of each other in young children? *Learning and Instruction*, 82, Article 101682. <https://doi.org/10.1016/j.learninstruc.2022.101682>
- Zheng, Y., & Barrot, J. S. (2024). Syntactic complexity in second language (L2) writing: Comparing students' narrative and argumentative essays. *System*, 123, Article 103342. <https://doi.org/10.1016/j.system.2024.103342>
- Zhou, J., & Lü, C. (2022). Enhancing syntactic complexity in L2 Chinese writing: Effects of form-focused instruction on the Chinese topic chain. *Frontiers in Psychology*, 13, Article 843789. <https://doi.org/10.3389/fpsyg.2022.843789>