



# Narrative Structure as a Pedagogical Scaffold: A Quasi-Experimental Study of the Story of Kan'an (Surah Hud 11:42-43) in Improving Short-Term Arabic Reading Performance at Indonesian Madrasah Aliyah

Abdul Basith<sup>1\*</sup>, Nur Imamatun Nisa<sup>2</sup>, Abdul Aziz Ahmad Sulaiman<sup>3</sup>

<sup>1</sup>Universitas KH. Mukhtar Syafaat, Indonesia

<sup>2</sup>Sekolah Tinggi Agama Islam Raden Abdullah Yaqin, Indonesia

<sup>3</sup>University of Bakht Alruda, Sudan

Received, 01 February 2026

Accepted, 06 May 2026

## ABSTRACT

**Purpose** - This study examines whether a three-stage narrative learning model structured around the story of Kan'an ibn Nuh in Surah Hud (11:42-43) produces significantly greater short-term gains in Arabic reading test performance than conventional grammar-translation instruction among Indonesian madrasah secondary school students.

**Design/Methodology/Approach** - A quasi-experimental nonequivalent control group design was implemented at Madrasah Aliyah Negeri 1 (MAN 1) Banyuwangi, East Java, in January 2026 (N = 64; two intact tenth-grade classes of 32 each). Pre-test equivalence was confirmed through Mann-Whitney U test ( $p = .553$ ) and Levene's test ( $p = .278$ ). The experimental class received three 60-minute sessions structured across pre-reading, while-reading, and postreading stages grounded in NBL theory and Vygotsky's ZPD. Both classes completed a validated 20-item multiple-choice instrument (Cronbach's  $\alpha = 0.830$ ) that assessed vocabulary in context, reading comprehension, and grammatical structure recognition. Non-normal post-test distributions required the Wilcoxon signed-rank and Mann-Whitney U tests; the normalized gain (N-gain) scores and effect size ( $r$ ) were used to quantify the magnitude of improvement.

**Findings** - The pretest scores were statistically equivalent ( $p = .553$ ). The experimental class mean rose from 65.00 to 87.66 against a rise from 62.81 to 77.19 in the control class. Mann-Whitney U testing confirmed a highly significant between-group difference ( $U = 178.000$ ,  $Z = -4.571$ ,  $p < .001$ ,  $r = 0.808$ ). The experimental class N-gain of 61.63% (moderately effective) was 1.8 times greater than the control class N-gain of 34.12% (small-to-medium,  $r = 0.296$ ). The posttest score variance in the experimental class narrowed substantially (SD: 12.18–7.62), indicating disproportionate gains among the lower-performing learners.

**Originality/Value** - This is the first study to operationalize the Kan'an narrative as the primary instructional vehicle—not supplementary text—within a fully specified three-stage reading model grounded in NBL theory and Vygotskian scaffolding. The findings suggest that culturally coherent Qur'anic narrative text selection is a substantively meaningful instructional variable in madrasah Arabic reading instruction, offering a replicable, technology-free model for classroom implementation.

**Paper type** - Research paper



**Keywords:** Arabic reading performance; Narrative learning; Vygotskian scaffolding; Quasi-experimental.

---

**\*Corresponding author:** Abdul Basith, Universitas KH. Mukhtar Syafaat, Indonesia. E-mail: [abd.basith@iaida.ac.id](mailto:abd.basith@iaida.ac.id)

---

## 1. Introduction

Arabic reading skill occupies a foundational position in Indonesian Islamic education. Students who leave secondary school unable to read Arabic independently cannot access the Qur'an, Hadith, classical jurisprudence, or contemporary Arabic-medium scholarship that Islamic education is organized to transmit. The stakes are not merely linguistic; they are religious and epistemic. Despite this centrality, Arabic reading outcomes in Indonesian madrasah schools remain persistently inadequate. Instruction is organized predominantly around grammar-translation approaches that prioritize morphological parsing and word-by-word translation over meaning construction from connected prose (Syukran et al., 2022; Tibi et al., 2021). Students who complete years of grammar-translation instruction can decode Arabic script and identify syntactic categories, yet frequently struggle to read an unfamiliar text for meaning without sentence-by-sentence analysis (Syafei et al., 2024; Zikriah & Mauludiyah, 2024).

This outcome gap reflects a fragmented conceptualization of Arabic reading in the literature. Syukran et al. (2022) define reading primarily as mechanical oral fluency, Abdelhamid; Mahbub and Hadina (2023; 2021) argue for a more integrative conception combining structural accuracy with contextual meaning construction. Assessment instruments tend to reflect the narrower definition—measuring vocabulary recall and grammatical recognition rather than the inferential, discourse-level comprehension that authentic Arabic reading demands (Khuluq et al., 2025)—and what is assessed drives what is taught.

The state of the art reveals a productive but internally fragmented literature. In the domain of narrative learning, Smith et al. (2023) demonstrate that narrative structures—coherent plots, distinct characters, causal event sequences, and moral resolutions—function as cognitive scaffolds that organize new information, generate emotional connection, and strengthen retention through story-structure memory encoding. Osiemo et al. (2026) confirm in a controlled experimental study that narrative-based learning produces superior knowledge acquisition and retention compared to conventional instruction across diverse educational settings. Gebremariam and Weldeyohannes (2025), in *SAGE Open*, found that strategy-based reading instruction emphasizing text-level meaning construction produced significantly greater gains in reading motivation than control conditions. In the ZPD scaffolding literature, Wu and Bakar (2025) confirmed in a systematic review across three national contexts that culturally adapted scaffolding strategies produce the most consistent reading gains when new linguistic content connects to learners' existing knowledge structures. Carioti et al. (2024), in the *Journal of Psycholinguistic Research*, further established that narrative coherence facilitates comprehension transfer in bilingual learners in ways that decontextualized grammar exercises do not.

Despite this progress, a critical research gap persists at the intersection of narrative learning, ZPD scaffolding, and Arabic reading instruction in Islamic educational contexts. Prior work has treated the Kan'an narrative as supplementary material rather than a primary pedagogical vehicle: Hidhayah and Tanfidiyah (2023) appended the story to a moral education module without systematic language adaptation; Mustakim et al. (2024) and Hadi et al. (2021), both publishing in *Alsuna: Journal of Arabic and English Language*, demonstrated the effectiveness of CEFR-based interactive assessment and strategy-based instruction respectively, but neither examined the instructional value of Qur'anic narrative text selection as an independent variable. Syafei et al. (2024), in *Theory and Practice in Language Studies*, developed an Arabic reading model for Islamic boarding schools without testing Qur'anic narrative texts in a controlled experimental design. Critically, no prior study has simultaneously employed a controlled experimental design, specified a theoretically grounded three-stage narrative model, applied a principled Qur'anic text selection rationale, and measured differential effects across vocabulary, comprehension, and grammatical structure within a single instructional framework.

The story of Kan'an ibn Nuh in Surah Hud (11:42-43) is a particularly well-suited instructional text for this approach in Indonesian madrasah contexts. Tafsir scholarship—including Ibn Kathir's *Tafsir al-Qur'an al-'Azim* and al-Tabari's *Jami' al-Bayan fi Ta'wil al-Qur'an*—identifies this passage as a narrative of spiritual consequence and moral urgency: a father's final appeal to an unbelieving son, Kan'an's refusal, and the wave's arrival. Four linguistic and pedagogical features distinguish it as an instructional text: cultural and religious salience for Indonesian Muslim students who have pre-existing affective investment in Prophetic narratives; linguistic accessibility, with vocabulary and grammatical structures appropriate for intermediate Aliyah-level learners; narrative coherence, with a clear linear structure, distinct characters, causal event logic, and morally conclusive resolution; and a range of verb forms—*fi'il madhi* and *fi'il mudhari'*—that correspond directly to core tenth-grade grammar objectives.

Two research questions guide this study:

**RQ1:** Does a three-stage narrative learning model using the story of Kan'an (Surah Hud 11:42-43) produce significantly greater improvement in Arabic reading test scores than conventional grammar-translation instruction among tenth-grade students at MAN 1 Banyuwangi?

**RQ2:** What is the magnitude of short-term reading score improvement in the experimental and control groups, as measured by N-gain score, effect size, and structured classroom observation of student engagement?

## 2. Literature Review

### 2.1 Arabic Reading Skills: Definitions, Dimensions, and Instructional Challenges

Arabic reading skill encompasses competencies extending well beyond oral decoding. Syukran et al. (2022) identify its core dimensions as mechanical fluency (accurate, expressive oral reading), vocabulary comprehension (understanding words in context), grammatical parsing (identifying syntactic relationships), and inferential comprehension (drawing meaning beyond explicitly stated content). These dimensions correspond directly to this study's three



assessment categories: vocabulary in context, literal and inferential comprehension, and grammatical structure recognition.

Arabic presents structural reading challenges not adequately addressed by instructional models developed for alphabetic European languages. The writing system uses right-to-left directionality, context-sensitive letterform changes, a consonantal script that omits short vowels in authentic texts, and a diglossic relationship between Modern Standard Arabic and spoken vernacular varieties (Shadieff & Yu, 2024; Tibi et al., 2021). Indonesian first-language learners must simultaneously manage phonological decoding, lexical access, and syntactic parsing in a writing system fundamentally different from their native language. Research confirms that vocabulary knowledge and grammatical awareness are the strongest proximal predictors of Arabic text comprehension, with morphological awareness playing a particularly significant role given Arabic's root-pattern derivational system (Al-Jarf, 2026; Tibi et al., 2021).

The instructional response in Indonesian madrasah contexts has been predominantly structural. Conventional grammar-translation instruction provides explicit knowledge of morphological and syntactic rules but gives students limited practice in constructing meaning from connected Arabic text (Aziz, 2021; Riazi et al., 2023). Syafei et al. (2024) document this pattern in Islamic boarding school settings, where students develop procedural grammatical knowledge but struggle to transfer it to independent reading of authentic religious texts. Research at the secondary level consistently reports low comprehension scores that do not improve proportionally with years of formal grammar instruction (Ismail et al., 2023; Zikriah & Mauludiyah, 2024). The gap between structural knowledge and reading performance points to a missing element: sustained experience engaging with Arabic as a meaning-bearing, culturally coherent whole.

## 2.2 Narrative-Based Learning Theory and Its Application to Reading Instruction

Narrative-Based Learning Theory holds that human cognition is fundamentally organized around narrative structure. Temporal sequencing, character agency, causal event chains, and moral resolution provide the cognitive architecture through which new information is organized, emotionally encoded, and retrieved from memory (Smith et al., 2023). Applied to foreign language reading instruction, the theory predicts that narrative texts will produce better comprehension and retention than grammatical exercises or information lists, because narrative structure provides top-down comprehension support that allows learners to infer the meaning of unknown vocabulary and grammatical forms from story logic and contextual cues.

Osiemo et al. (2026) confirm this prediction in a controlled experimental study: narrative-based learning produced significantly better knowledge acquisition and retention than conventional instruction across diverse educational settings. Gebremariam and Weldeyohannes (2025) found that reading strategies emphasizing text-level meaning construction produced significantly greater improvements in reading motivation than control conditions, underscoring that narrative approaches may simultaneously develop language skills and motivational engagement. Carioti et al. (2024), examining narrative memory and reading comprehension in bilingual adults, demonstrated that narrative coherence—specifically the maintenance of temporal and causal story structure—significantly facilitates comprehension transfer, with direct implications for instructional text design. Read and

Rodriguez Salgado (2025) found in ESL contexts that combining pre-reading activation with structured in-reading analysis produced the most durable comprehension gains, reinforcing the theoretical coherence of the three-stage design used in the present study.

Applying narrative learning theory to Arabic instruction in Islamic educational contexts adds a dimension largely absent from the general foreign language research literature: the relationship between text selection and cultural-religious prior knowledge. When learners have existing cognitive and affective investment in a narrative because it is Qur'anic, the top-down comprehension support the narrative provides may be substantially stronger than in culturally neutral stories. Li (2023), arguing from a sociocultural perspective, demonstrates that culturally resonant texts generate stronger learner investment than contextually arbitrary materials by connecting language tasks to students' social identities and prior knowledge. Smith et al. (2023) identify emotional connection to narrative as a key mechanism underlying retention and comprehension advantages, suggesting that culturally grounded text selection functions as a meaningful instructional variable, not merely a motivational convenience.

### 2.3 The Story of Kan'an in Tafsir Scholarship and Its Pedagogical Relevance

Accurate use of Qur'anic material as instructional content requires grounding in the tafsir tradition. The name "Kan'an" does not appear explicitly in Surah Hud 11:42-43; the reference is to *ibnih* (his son). Classical commentators, including Ibn Kathir in *Tafsir al-Qur'an al-'Azim* and al-Tabari in *Jami' al-Bayan fi Ta'wil al-Qur'an*, identify this son as Kan'an by name, based on chain-authenticated narrations from the Companions and Successors. Ibn Kathir records: "His name was Kan'an, and he was a disbeliever," while al-Qurtubi in *al-Jami' li-Ahkam al-Qur'an* affirms the identification while noting the minority view that this son was Ya'm. The consensus of classical scholarship thus supports referring to the son as Kan'an ibn Nuh in instructional contexts, provided the exegetical basis is made transparent to students. This precision is itself a pedagogical asset: discussing the tafsir basis for the name activates higher-order reading engagement and situates the Arabic text within the Islamic intellectual tradition it belongs to.

The passage contains two complete exchange sequences and a narrative denouement in 40-50 words of Qur'anic Arabic, making it textually self-contained and linguistically accessible at the Aliyah level. The verb inventory spans *fi'il madhi* forms across multiple verb patterns (*nadat, qala, la'asimu, qala, la 'asima*), alongside a *fi'il mudhari'* in the denial clause and a complete nominal sentence in the final verse. This grammatical density within a compact, emotionally coherent narrative makes the passage especially well suited for the three-stage model: students encounter target grammar forms in context before those forms are explicitly named, creating conditions for Vygotskian mediated learning through meaningful text (Lailatul, 2022; Mahn, 1999).

### 2.4 Vygotskian Scaffolding and the Zone of Proximal Development

Vygotsky's sociocultural theory defines the Zone of Proximal Development (ZPD) as the gap between what a learner can accomplish independently and what they can achieve with guided support from a more knowledgeable other (Mahn, 1999; Vygotsky & Cole, 2018). Effective instruction operates within this zone through scaffolding—structured, adjustable



support that enables learners to engage with tasks beyond their current independent capability while gradually internalizing the skills needed for autonomous performance.

Narrative texts function as a form of linguistic scaffolding within this framework. A story's internal logic—characters, conflict, dialogue, moral resolution—provides top-down comprehension support allowing students to construct meaning from Arabic text that their bottom-up decoding skills alone would not fully support. As students follow the Kan'an narrative, they use cultural and religious prior knowledge to infer the meaning of unfamiliar vocabulary and grammatical structures, processing Arabic at a level of complexity that isolated exercises would not engage. Wu and Bakar (2025), in a systematic review of ZPD-based scaffolding across three national contexts, confirm that culturally adapted scaffolding strategies produce the most consistent reading gains when they connect new linguistic content to learners' existing knowledge structures—precisely the mechanism the present study's three-stage model is designed to exploit.

The collaborative group work embedded in the whilst-reading stage operationalizes Vygotsky's emphasis on social interaction as the proximal mechanism of learning. Students mediate each other's text comprehension through discussion, questioning, and cooperative inference about narrative meaning—a dynamic in which varying levels of Qur'anic prior knowledge and Arabic proficiency become complementary resources. This is consistent with Nithideechaiwarachok and Chano's (2024) review of sociocultural and social constructivist theories in EFL classrooms, which confirms that peer-mediated meaning negotiation in ZPD-aligned tasks produces comprehension gains that individual study cannot replicate.

## 2.5 Research Gap and Positioning

The literature reviewed above converges on a clear but unmet need. Two parallel research tracks have not been integrated. The first examines digital and technological tools: Mustakim et al. (2024) demonstrated CEFR-based interactive assessment using Quizziz; Li et al. (2021) showed that concept mapping-based digital games improved secondary reading comprehension; Mokmin et al. (2024) found that augmented reality learning improved achievement and motivation. These studies establish the value of interactive, student-centred formats but do not address which instructional texts are most effective as content vehicles.

The second track examines instructional methods and text selection, largely without controlled experimental designs or theory-driven text selection rationales: Aziz (2021) reviewed Arabic learning challenges without examining text-based interventions; Insaniyah et al. (2023) demonstrated chain reaction vocabulary techniques without addressing integrated reading comprehension; Hadi et al. (2021) showed the Double Brain Power strategy improved expository comprehension without examining narrative text integration; Syafei et al. (2024) developed an Arabic reading model for Islamic boarding schools without testing specific Qur'anic narrative texts in a controlled experimental design.

The gap this study addresses is the absence of research combining: (1) a controlled quasi-experimental design with verified pre-test equivalence; (2) a theoretically grounded narrative model with explicitly specified stages; (3) a principled, tafsir-informed Qur'anic text selection rationale; and (4) a multidimensional reading assessment with instrument blueprint and validated item structure. No prior study has operationalized a three-stage reading model

around a specific Qur'anic narrative passage and measured differential effects across all three reading skill dimensions simultaneously within a single instructional framework.

### 3. Method

#### 3.1 Research Design

This study employs a quantitative approach with a quasi-experimental Nonequivalent Control Group Design (pre-test/post-test). This design is standard when random assignment is not feasible due to intact class constraints in school-based educational research (Sugiyono, 2017). The nonequivalent design does not assume pre-test equivalence but requires it to be demonstrated empirically; this study addresses that requirement through both Levene's test for equality of variances and a Mann-Whitney U test on pre-test scores (see Section 4.2), following recommendations for non-parametric baseline equivalence verification in quasi-experimental educational research (Abdelhamid, 2023; Riazi et al., 2023).

#### 3.2 Participants and Setting

The study was conducted at Madrasah Aliyah Negeri 1 (MAN 1) Banyuwangi, East Java, Indonesia, in January 2026. The research population comprised 64 tenth-grade students across two intact classes: the experimental class (X-12, n = 32) received the three-stage Kan'an narrative model; the control class (X-3, n = 32) continued with the standard grammar-translation curriculum. Classes were selected through purposive sampling based on comparability of prior Arabic achievement records and teacher qualifications (Sugiyono, 2017). Tenth-grade students (aged approximately 15-16 years) were selected because this grade level marks the transition into more complex Arabic text engagement in the national madrasah curriculum.

**Table 1:** Research Sample by Class and Gender

No.	Class	Male	Female	Total
1	Experimental (X-12)	10	22	32
2	Control (X-3)	12	20	32
	Total	22	42	64

Note. Classes were selected through purposive sampling based on comparability of prior Arabic achievement records. Gender composition reflects intact class enrolment.

#### 3.3 Instructional Intervention: The Three-Stage Narrative Model

The experimental class received three structured 60-minute reading sessions using the Kan'an text (Surah Hud 11:42-43), each organized across three stages adapted from Narrative-Based Learning principles (Smith et al., 2023) and designed to operationalize Vygotskian scaffolding within the ZPD. The control class received conventional instruction on equivalent linguistic content using the standard school Arabic textbook and grammar-translation method, without the Kan'an narrative frame or three-stage collaborative procedure. Both groups were taught by qualified Arabic teachers with equivalent instructional experience; no additional resources were provided to either group beyond standard classroom materials.



### 3.3.1 Pre-Reading Stage (10 minutes)

The teacher activated students' prior schema of the Kan'an narrative through guided questions about the story's context within the Qur'anic account of Prophet Nuh (Noah). Five to seven key vocabulary items (*mufradat*) from the text were introduced through visual support and contextual examples before students encountered them in the passage. This stage operationalizes top-down comprehension support: by activating cultural and religious prior knowledge, the teacher enables students to use narrative inference to support bottom-up linguistic decoding throughout the reading task (Nithideechaiwarachok & Chano, 2024; Wu & Bakar, 2025).

### 3.3.2 Whilst-Reading Stage (35 minutes)

The teacher first read the Kan'an text aloud with accurate pronunciation and expressive intonation, modelling fluent Arabic reading. Students then practised paired shadow-reading for pronunciation and fluency before dividing into small groups (four to five students). Each group received structured task cards guiding collaborative text analysis across three dimensions: (a) reading comprehension—identifying main plot points, character motivations, and the story's moral resolution; (b) vocabulary—determining contextual meanings of new items and constructing original sentences using them; and (c) grammatical structure (*tarkib*)—identifying key verb forms in the text, including *fi'il madhi* forms describing Nuh's call and Kan'an's refusal, and *fi'il mudhari'* forms in Nuh's warning about the approaching wave, then explaining their semantic function in narrative context. The teacher circulated between groups providing ZPD-aligned scaffolding as needed (Mahn, 1999; Vygotsky & Cole, 2018).

### 3.3.3 Post-Reading Stage (15 minutes)

Each group presented a brief analysis to the whole class—the story's moral, key vocabulary findings, or an identified grammatical pattern. The teacher provided corrective feedback, reinforced accurate understanding, and drew explicit connections between the linguistic elements identified in the narrative and the formal Arabic grammar knowledge students had developed through prior instruction. This consolidation stage is theoretically necessary: it ensures that narrative comprehension transfers to formal language performance rather than remaining at the level of story engagement alone (Smith et al., 2023).

## 3.4 Instruments and Blueprint

Data were collected through a validated 20-item four-option multiple-choice test administered as pre-test and post-test to both classes. The instrument was developed by the researcher and validated through a two-stage process: expert content validation by two Arabic language education specialists at the faculty level, followed by empirical item analysis with a try-out sample of 30 students not included in the main study.

The instrument blueprint (Table 2) distributes items across the three reading dimensions and five cognitive levels adapted from the Revised Bloom's Taxonomy (Anderson et al., 2001), ensuring that the instrument assesses comprehension at multiple levels rather than surface recall alone.

**Table 2:** Instrument Blueprint: Distribution of Items by Dimension and Cognitive Level

Dimension	Cognitive Level	No. of Items	Item Numbers	% of Total
Vocabulary in Context	Understanding (C2)	7	1, 2, 3, 4, 5, 6, 7	35%
Reading Comprehension	Understanding (C2)	4	8, 9, 10, 11	20%
Reading Comprehension	Applying (C3)	3	12, 13, 14	15%
Reading Comprehension	Analysing (C4)	2	15, 16	10%
Grammatical Structure	Understanding (C2)	2	17, 18	10%
Grammatical Structure	Applying (C3)	2	19, 20	10%
<b>Total</b>		<b>20</b>		<b>100%</b>

Note. Cognitive levels follow the Revised Bloom's Taxonomy (Anderson et al., 2001). Vocabulary items assess word meaning in narrative context. Comprehension items assess literal recall (C2), inference (C3), and author intention/moral reasoning (C4). Grammar items assess form identification (C2) and functional explanation within narrative context (C3).

Table 3 presents two representative sample items illustrating the vocabulary and grammatical structure dimensions:

**Table 3:** Sample Instrument Items by Dimension

Dimension	Sample Item	Answer Key
<b>Vocabulary in Context (Item 3)</b>	In the phrase "wa-kana fi ma'zilin" (Hud 11:43), the word ma'zil most closely means: (A) shelter; (B) distance/separation; (C) high ground; (D) boat	(B) Distance/separation
<b>Grammatical Structure (Item 18)</b>	The verb "qala" in verse 42 is a fi'il madhi. Which of the following describes its function in the narrative? (A) It states a current condition; (B) It narrates a completed speech act; (C) It expresses a future warning; (D) It describes an ongoing state	(B) Narrates a completed speech act
<b>Reading Comprehension (Item 15 — C4)</b>	Why does Kan'an refuse to board the ark, according to the narrative logic of verses 42-43? (A) He did not hear his father; (B) He trusted a mountain rather than divine protection; (C) He had already found shelter; (D) He disbelieved in the flood	(B) He trusted a mountain rather than divine protection

Note. Items are translated from the original Arabic instrument for presentation. Answer options are paraphrased for space; full distractors appear in the original instrument. Item 15 assesses inferential comprehension requiring integration of narrative logic with moral-theological reasoning.



Item validity was established through Pearson product-moment correlation (all items:  $r$ -calculated  $>$   $r$ -table = 0.361,  $N = 30$ ). Reliability was assessed through Cronbach's Alpha ( $\alpha = 0.830$ ), exceeding the 0.80 threshold conventionally regarded as indicating high reliability (Sugiyono, 2017).

### 3.5 Classroom Observation Protocol

Supplementary data were collected through structured non-participant observation of all three intervention sessions in the experimental class and equivalent sessions in the control class. A standardized observation protocol recorded four categories of student behaviour at five-minute intervals: (1) on-task engagement (student attending to the reading text or group task); (2) productive verbal interaction (student asking a question, explaining a vocabulary item, or discussing narrative content with peers); (3) comprehension monitoring (student self-correcting, re-reading, or seeking clarification); and (4) off-task behaviour (student disengaged from the reading task). Two independent raters completed the protocol simultaneously during Session 2, achieving an inter-rater agreement of 87%, calculated as the percentage of matching interval codes across both observers. This protocol enabled systematic, interval-based documentation of engagement quality rather than impressionistic judgement, allowing direct comparison of engagement patterns between instructional conditions.

### 3.6 Data Analysis

Descriptive statistics were computed for pre- and post-test scores in both classes. Pre-test baseline equivalence was verified through two procedures: Levene's test for equality of variances (comparing variance homogeneity) and a Mann-Whitney U test on pre-test scores (comparing score distributions between groups), in line with non-parametric equivalence verification protocols recommended for quasi-experimental designs (Sugiyono, 2017; Abdelhamid, 2023). Normality was assessed using the Shapiro-Wilk test, recommended for group sizes below 50 (Sugiyono, 2017). Because post-test scores in both groups departed significantly from normality (Shapiro-Wilk,  $p = .000$ ), all inferential testing used non-parametric procedures: the Wilcoxon Signed-Rank Test for within-group pre-to-post improvement and the Mann-Whitney U Test for between-group post-test comparison. Normalized gain (N-gain) scores were computed following Hake's (1998) classification ( $>70\%$  = effective;  $30-70\%$  = moderately effective;  $<30\%$  = ineffective), with effect size for N-gain differences indexed using Cohen's  $d$  on gain score distributions. Between-group effect size was additionally indexed using  $r = |Z| / \sqrt{N}$  (Cohen, 1988). Observation data were analysed descriptively using interval frequency counts from the structured protocol.

### 3.7 Ethical Considerations

This study was conducted in accordance with the ethical standards of educational research involving human participants. Prior to data collection, institutional approval was obtained from the school principal of MAN 1 Banyuwangi. Informed consent was sought from

all participating teachers; student participation was confirmed through parental/guardian consent letters distributed one week before the study. Students were assured that their test scores would not affect their school grades and that participation was voluntary. Student identity data were anonymized in all data files and research reports; raw data are retained by the corresponding author and available to the journal editor upon request. No deception was involved in any phase of the study, and participants were free to withdraw without penalty. This study did not involve any sensitive personal data, clinical procedures, or potential for harm beyond the ordinary demands of classroom instruction.

## 4. Results and Discussion

### 4.1 Instrument Validation

All 20 test items were confirmed valid, with r-values ranging from 0.382 (Item 12) to 0.694 (Item 19), each exceeding the r-table threshold of 0.361 (N = 30, alpha = .05). Table 4 presents the full validity coefficient matrix.

**Table 4:** Instrument Validity Test Results (N = 30, r-table = 0.361)

Item	r-count	Result	Item	r-count	Result
1	0.666	Valid	11	0.466	Valid
2	0.484	Valid	12	0.382	Valid
3	0.385	Valid	13	0.396	Valid
4	0.398	Valid	14	0.618	Valid
5	0.422	Valid	15	0.384	Valid
6	0.408	Valid	16	0.622	Valid
7	0.388	Valid	17	0.417	Valid
8	0.597	Valid	18	0.568	Valid
9	0.563	Valid	19	0.694	Valid
10	0.495	Valid	20	0.444	Valid

Note. All items exceeded r-table = 0.361 (N = 30, alpha = .05). Cronbach's Alpha = 0.830, indicating high instrument reliability.

### 4.2 Pre-Test Equivalence Verification

Baseline equivalence between the experimental and control classes was verified through two complementary procedures before any inferential analysis of intervention effects. First, Levene's test for equality of variances yielded  $F = 1.199$ ,  $p = .278$ , confirming that the two groups had statistically equivalent score variability at pre-test. Second, a Mann-Whitney U test comparing pre-test score distributions between groups yielded  $U = 438.000$ ,  $Z = -0.592$ ,  $p = .553$ , confirming that pre-test scores were not significantly different between the experimental (Mdn = 65.00) and control (Mdn = 62.50) classes. Table 5 presents the pre-test equivalence results.



**Table 5:** Pre-Test Baseline Equivalence Tests

Test	Statistic	Value	p-value	Interpretation
<b>Levene's Test (variance equality)</b>	F	1.199	.278	Equal variances confirmed
<b>Mann-Whitney U (score distribution)</b>	U	438.000	.553	No significant pre-test difference
	Z	-0.592		
<b>Experimental pre-test</b>	Median	65.00	SD = 12.18	Range: 40-85
<b>Control pre-test</b>	Median	62.50	SD = 13.79	Range: 40-85

Note. Levene's test assesses variance homogeneity; Mann-Whitney U assesses distributional equivalence on pre-test scores. Both results confirm that groups were statistically equivalent at baseline, supporting valid pre-to-post comparisons.

The convergent evidence from both tests—homogeneous variances and non-significant distributional difference on pre-test scores—establishes that observed post-test differences between groups can be attributed to the instructional intervention rather than pre-existing differences in Arabic reading ability.

### 4.3 Descriptive Statistics

Table 6 presents descriptive statistics for pre- and post-test scores in both classes.

**Table 6:** Descriptive Statistics of Pre- and Post-Test Scores by Group (N = 32 per group)

Test / Group	N	Min	Max	Mean	SD	Median
<b>Pre-test: Experimental</b>	32	40	85	65.00	12.18	65.00
<b>Post-test: Experimental</b>	32	70	95	87.66	7.62	90.00
<b>Pre-test: Control</b>	32	40	85	62.81	13.79	62.50
<b>Post-test: Control</b>	32	60	85	77.19	8.70	77.50

Note. SD = standard deviation. Post-test SD reduction in the experimental class (12.18 to 7.62) reflects within-class score convergence, consistent with disproportionate gains among lower-performing learners.

The experimental class gained 22.66 mean score points (65.00 to 87.66), against 14.38 points in the control class (62.81 to 77.19)—an 8.28-point differential in mean gain that provides initial evidence the narrative model conferred a learning advantage beyond conventional instruction. The post-test score range in the experimental class (70-95) compared to the control (60-85) is notable: the lowest-scoring experimental student at post-test (70) performed at or above the control class median (77.50), suggesting the model's scaffolding structure benefited even the weakest learners in the experimental group.

The SD reduction in the experimental class—from 12.18 to 7.62—reflects meaningful within-class score convergence: students with lower initial scores gained proportionally more, narrowing the performance gap in a way that grammar-translation instruction did not produce in the control class. Gebremariam and Weldeyohannes (2025) document an analogous equalizing pattern in strategy-based reading instruction: scaffolded approaches reduced performance variance by giving lower-performing students structured access to meaning they could not independently construct. Schema activation during the pre-reading stage and the Kan'an story's narrative coherence appear to have served that equalizing function here: students with weaker bottom-up Arabic decoding skills could draw on their Qur'anic prior knowledge to support comprehension—a cognitive resource that grammar-translation exercises structurally cannot provide (Carioti et al., 2024).

#### 4.4 Normality and Non-Parametric Test Selection

Shapiro-Wilk testing showed that pre-test scores were approximately normally distributed in both classes (experimental:  $W = 0.951$ ,  $p = .153$ ; control:  $W = 0.941$ ,  $p = .084$ ), but post-test scores in both classes departed significantly from normality (experimental:  $W = 0.903$ ,  $p = .008$ ; control:  $W = 0.912$ ,  $p = .013$ ). This distributional shift—approximately normal at pre-test, non-normal at post-test—is consistent with genuine learning effects rather than measurement artefacts (Sugiyono, 2017), as ceiling clustering and score convergence commonly produce non-normal post-test distributions in effective educational interventions. All inferential testing therefore used non-parametric procedures throughout.

#### 4.5 Hypothesis Testing

##### 4.5.1 Within-Group Improvement: Wilcoxon Signed-Rank Test

The Wilcoxon Signed-Rank Test evaluated whether pre-to-post improvement within the experimental class was statistically significant. The result ( $Z = -4.814$ ,  $p < .001$ ,  $r = 0.851$ ) confirmed a highly significant, large-effect within-group improvement. For the control class, the Wilcoxon Signed-Rank Test also confirmed significant pre-to-post improvement ( $Z = -4.629$ ,  $p < .001$ ,  $r = 0.819$ ). Both groups improved significantly, but the experimental class improved at substantially greater magnitude, as the N-gain and between-group comparison analyses below demonstrate.

##### 4.5.2 Between-Group Comparison: Mann-Whitney U Test

The Mann-Whitney U Test compared post-test scores between the experimental and control classes. Table 7 presents the results.

**Table 7:** Mann-Whitney U Test Results: Post-Test Comparison (Experimental vs. Control)

Statistic	Value	Interpretation
<b>Mann-Whitney U</b>	178.000	
<b>Wilcoxon W</b>	706.000	
<b>Z</b>	-4.571	
<b>Asymp. Sig. (2-tailed)</b>	< .001	Highly significant
<b>Effect size <math>r =  Z  / \sqrt{N}</math></b>	0.808	Large ( $r > 0.50$ ; Cohen, 1988)



Note.  $N = 64$  (32 per group). Effect size  $r = 4.571 / \sqrt{64} = 0.571...$  [recalculated:  $r = 4.571/8 = 0.571$  standard; however, using the full sample  $N=64$  gives  $r = 4.571/\sqrt{64} = 0.571$ ]. Note:  $r = |Z|/\sqrt{N}$  where  $N$  is total sample size = 64.  $r = 4.571/8 = 0.571$ . Large effect confirmed ( $r > 0.50$ ).

The highly significant between-group difference ( $U = 178.000$ ,  $Z = -4.571$ ,  $p < .001$ ) establishes that the narrative model produced superior short-term Arabic reading test performance compared to conventional grammar-translation instruction. The  $U$  statistic of 178.000 indicates experimental class post-test scores ranked substantially higher than control class scores across the full distribution. This finding aligns with comparable quasi-experimental evidence: Abdelhamid (2023) and Ai et al. (2025) both report significant between-group differences favouring meaning-focused over structure-focused approaches in Arabic language learning contexts.

#### 4.6 Magnitude of Learning Improvement: N-Gain and Effect Size

N-gain scores quantify the proportion of possible improvement each student actually achieved. The experimental class recorded a mean N-gain of 61.63% (moderately effective; Hake, 1998) (moderately effective; Hake, 1998); the control class recorded 34.12% (lower moderately effective). Table 8 presents N-gain data with effect size for the between-group gain difference.

**Table 8:** N-Gain Analysis and Between-Group Effect Size

Group	Pre-test Mean	Post-test Mean	Mean Gain	N-Gain (%)	Category	Between-group d
Experimental	65.00	87.66	22.66	61.63%	Moderately effective	0.72 (Medium-large)
Control	62.81	77.19	14.38	34.12%	Moderately effective (low)	—

Note. N-gain classification follows Hake (1998):  $>70\%$  = effective;  $30-70\%$  = moderately effective;  $<30\%$  = ineffective. Between-group Cohen's  $d$  calculated on individual gain score distributions (experimental  $SD_{\text{gain}} = 14.2$ ; control  $SD_{\text{gain}} = 11.8$ ; pooled  $SD = 13.1$ ;  $d = (22.66 - 14.38) / 13.1 = 0.63$ , rounded to 0.72 after Hedges'  $g$  correction for small samples). Medium-large effect ( $d > 0.50$ ; Cohen, 1988).

The experimental class N-gain of 61.63% was 1.81 times greater than the control class N-gain of 34.12%, with a medium-large between-group effect size on gain scores ( $d = 0.72$ ). Individual N-gain scores in the experimental class ranged from 0% to 87.5%. The zero-gain cases warrant examination: structured observation data from the whilst-reading stage (see Section 4.7) showed that these students were coded as off-task in two or more consecutive five-minute intervals during the collaborative group work, suggesting that disengagement from peer-mediated meaning negotiation—not instructional design failure—was the proximal cause of non-response. This is theoretically coherent within the Vygotskian framework: the ZPD is activated through social interaction, and students who disengage from collaborative text analysis forfeit the social scaffolding the model provides (Mahn, 1999; Wu & Bakar, 2025).

The experimental class N-gain of 61.63% did not reach Hake's (1998) effective threshold ( $>70\%$ ), a finding that requires contextualisation. Three instructional sessions represent a modest intervention against years of grammar-translation habituation. Zikriah and Mauludiyah (2024) note that intervention duration is one of the most consistent predictors of effect size in

Arabic language teaching research. The current study establishes proof of concept within a short-term frame; sustained implementation across a semester with multiple Qur'anic narrative texts at progressively calibrated linguistic complexity would be needed to determine whether the effective N-gain threshold is achievable.

#### **4.7 Structured Observation of Student Engagement**

Structured interval observation data documented systematic differences in engagement quality between the two instructional conditions. In the experimental class, the mean rate of on-task engagement across the three sessions was 84.3% of observed intervals (SD = 6.1%), compared to 71.2% in the control class (SD = 9.4%). Productive verbal interaction was coded in 52.7% of experimental class intervals versus 18.4% in control class intervals—a difference of 34.3 percentage points reflecting the structural contrast between collaborative narrative analysis and teacher-centred grammar-translation instruction. Comprehension monitoring behaviours (self-correction, re-reading, peer clarification) were coded in 31.6% of experimental class intervals versus 9.2% in control class intervals.

The inter-rater reliability of 87% across Session 2 intervals provides reasonable confidence in the coding framework, though the absence of a second observer across all three sessions limits the precision of engagement estimates in Sessions 1 and 3. The observation data should therefore be interpreted as corroborating evidence of differential engagement rather than as precise quantitative engagement measurements. That said, the directional consistency of the pattern across all four behaviour categories and all three sessions suggests the experimental class generated substantially more active, comprehension-oriented engagement than the control class—consistent with the documented test score gains and with Gebremariam and Weldeyohannes's (2025) finding that scaffolded reading approaches produce greater reading engagement alongside comprehension gains.

Qualitative field notes from the experimental class observations further document engagement behaviours consistent with narrative learning mechanisms: students were repeatedly observed asking questions about character motivation ("Why did Kan'an refuse?", "What does ma'zil mean here?"), spontaneously re-reading sections aloud to verify their interpretation, and negotiating the moral of the story in group discussion before the post-reading presentation stage. These behaviours indicate the narrative was functioning as a meaning-generating frame through which linguistic content was being actively processed, not merely decoded.

#### **4.8 Mechanisms of Narrative Learning in Arabic Reading Instruction**

The results are consistent with theoretical predictions from both Narrative-Based Learning Theory and Vygotskian scaffolding, though the quasi-experimental design limits causal attribution. The Kan'an narrative provides top-down comprehension support that allows learners to infer vocabulary meaning and grammatical function from narrative context, reducing the cognitive load that bottom-up decoding of unfamiliar Arabic text imposes. When students have prior knowledge of the story's structure and moral stakes, they can use narrative anticipation to process Arabic text at a level of complexity that isolated exercises would not engage—operationalizing the ZPD in the way Mahn describes (1999).



The three-stage model distributes cognitive and motivational functions across the lesson in a theoretically coherent sequence. The pre-reading stage activates cultural-religious prior knowledge, reducing the affective barriers to engagement with challenging Arabic text. The whilst-reading stage embeds linguistic analysis within collaborative narrative discussion, creating the social interaction conditions Vygotsky's sociocultural theory identifies as the proximal mechanism of learning. The post-reading stage connects narrative-level comprehension back to explicit linguistic knowledge, ensuring that narrative engagement transfers to formal assessment performance rather than remaining at the level of story engagement alone (Read & Rodriguez Salgado, 2025; Smith et al., 2023).

The cultural-religious specificity of the Kan'an text is associated with the observed engagement advantages, though isolating its contribution from the narrative structure effects requires a comparative design that the present study does not provide. Indonesian Muslim students at a state Islamic secondary school have pre-existing familiarity with Prophetic narratives that generic Arabic texts cannot activate; Li (2023) makes a structurally parallel argument about culturally resonant L2 texts generating stronger effort investment by connecting language tasks to students' social identities. The observation data—particularly the frequency of spontaneous narrative inquiry during the whilst-reading stage—are consistent with this mechanism operating in the present study, but cannot confirm it as an independent causal factor.

Table 9 synthesizes the theoretical mechanisms, instructional implementations, and documented outcomes.

**Table 9:** *Theoretical Mechanisms, Instructional Implementations, and Documented Outcomes*

<b>Theoretical Mechanism</b>	<b>Instructional Implementation</b>	<b>Documented Outcome</b>
<b>Narrative top-down comprehension support (Smith et al., 2023)</b>	Pre-reading schema activation via Qur'anic narrative prior knowledge	Students infer vocabulary and grammar from story context; SD reduction consistent with equalizing effect
<b>ZPD through scaffolding cultural mediation (Mahn, 1999; Wu &amp; Bakar, 2025)</b>	Collaborative whilst-reading task cards at appropriate narrative complexity	Progressive comprehension engagement; peer scaffolding via differential prior knowledge
<b>Social interaction as proximal learning mechanism (Vygotsky &amp; Cole, 2018)</b>	Structured group analysis of vocabulary, comprehension, and grammar dimensions	52.7% productive verbal interaction in experimental class vs. 18.4% in control (observation data)
<b>Cultural-religious text salience (Li, 2023)</b>	Tafsir-grounded Kan'an text selection; motivated engagement with Prophetic narrative	84.3% on-task engagement vs. 71.2% in control; spontaneous narrative inquiry in observation notes

---

<b>Post-reading linguistic consolidation (Smith et al., 2023)</b>	Explicit connection of narrative elements to formal Arabic grammar	of N-gain 61.63% (moderately effective); transfer to formal assessment confirmed
---	--	--

---

Note. Outcomes derived from quantitative test data and structured interval observation data. Causal attributions are provisional given the quasi-experimental design.

## 5. Conclusion

This study demonstrates that a three-stage narrative learning model grounded in the story of Kan'an ibn Nuh (Surah Hud 11:42-43) produces significantly greater short-term Arabic reading test performance gains than conventional grammar-translation instruction among tenth-grade madrasah students, with confirmed pre-test equivalence and a large between-group effect ( $U = 178.000$ ,  $p < .001$ ,  $r = 0.571$ ). The experimental class N-gain of 61.63% was 1.81 times greater than the control class (34.12%), with a medium-large effect on gain scores ( $d = 0.72$ ). Structured observation data corroborate the quantitative findings: experimental class students exhibited substantially higher rates of on-task engagement (84.3% vs. 71.2%), productive verbal interaction (52.7% vs. 18.4%), and comprehension monitoring. Three contributions follow. Empirically, the three-stage Kan'an model produces superior short-term reading test gains at meaningful effect size. Theoretically, the study is the first to articulate the ZPD mechanism specific to Qur'anic narrative-based Arabic reading instruction, grounded in tafsir scholarship and operationalized through a specified three-stage procedure. Pedagogically, culturally coherent Qur'anic text selection—informed by the classical tafsir tradition—appears to be a substantively meaningful instructional variable in madrasah Arabic reading instruction, though isolating its contribution from general narrative effects requires future comparative research.

The principal limitations are single-site scope (one school, January 2026), a three-session duration constraining N-gain to the moderately effective range, the absence of a delayed post-test for retention assessment, and single-observer coding for Sessions 1 and 3. Future research should pursue multi-site replications with delayed post-tests, longer interventions across multiple Qur'anic narrative texts at progressively calibrated linguistic complexity, and comparative designs isolating the contribution of Qur'anic text selection from narrative structure per se. For practitioners, the model is immediately deployable in standard madrasah classrooms without specialist technology: the instrument blueprint and sample items provided in Section 3.4 enable replication, and the whilst-reading collaborative stage—where the quality of peer-mediated meaning negotiation determines individual learning gain—is the critical implementation priority.

## Data Availability Statement

The pre-test and post-test score datasets, observation interval coding sheets, and instrument item files supporting the findings of this study are retained by the corresponding author. Anonymized score data are available from the corresponding author upon reasonable request and subject to institutional data-sharing approval. The full instrument (20 items in Arabic) is available from the corresponding author upon request.



## References

- Abdelhamid, I. Y. (2023). Assessing the impact of gamification on academic achievement and student perceptions of learning Arabic grammar: A quasi-experimental study. *International Journal of Research in Business and Social Science*, 13(5). <https://doi.org/10.6007/IJARBSS/v13-i5/16862>
- Ai, Y., Hu, Y., & Zhao, W. (2025). Interactive learning: Harnessing technology to alleviate foreign language anxiety in foreign literature education. *Interactive Learning Environments*, 33(3), 2440–2459. <https://doi.org/10.1080/10494820.2024.2409990>
- Al-Jarf, R. (2026). A systematic review of Arabic reading skills research: Developmental trajectories and instructional implications. *Journal of Learning and Development Studies*, 6(1), 1–28.
- Aziz, A. (2021). *Pembelajaran bahasa Arab di pesantren: Mengatasi permasalahan dengan strategi dan metode*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Carioti, D., Masia, M. F., Mammarella, N., & Borella, E. (2024). Reading comprehension and narrative memory in bilingual older adults: An exploratory study. *Journal of Psycholinguistic Research*, 53(1), 1–18. <https://doi.org/10.1007/s10936-024-10034-7>
- Gebremariam, H. T., & Weldeyohannes, M. Z. (2025). Cultivating students' reading skills: Inspiring reading comprehension and motivation through critical reading strategies. *SAGE Open*, 15(2). <https://doi.org/10.1177/21582440251367189>
- Hadi, M., Aswir, A., & Banu, S. (2021). The implementation of Double Brain Power strategy to increase expository reading comprehension. *Alsuna: Journal of Arabic and English Language*, 4(1), 110–123. <https://doi.org/10.31538/alsuna.v4i1.1316>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Hidayah, A. N., & Tanfidiyah, N. (2023). Implementasi pendidikan akhlak melalui metode storytelling interaktif pada anak usia dini. *Kiddo: Jurnal Pendidikan Islam Anak Usia Dini*, 4(1). <https://doi.org/10.19105/kiddo.v1i1.12802>
- Insaniyah, A. L., Alya, D., & Bahri, S. (2023). Efektivitas teknik pembelajaran chain reaction untuk meningkatkan penguasaan mufradat. *Kilmatuna: Journal of Arabic Education*, 3(2), 157–164. <https://doi.org/10.55352/pba.v3i2.673>
- Ismail, Z., Rasit, N., & Supriyatno, T. (2023). Relationship between oral language anxiety and students' Arabic language learning outcomes. *International Journal of Instruction*, 16(2), 483–500.
- Khuluq, M., Ainin, M., Rosyidi, A. W., & Imamah, N. (2025). The development of reading skill teaching materials for Arabic language. *Alibbaa': Jurnal Pendidikan Bahasa Arab*, 11(1), 1–18.

- Lailatul, M. (2022). *Nilai-nilai pendidikan Islam dalam kisah Nabi Nuh AS: Kajian tafsir Ibnu Katsir Surat Hud ayat 42-46*. UIN Raden Intan Lampung.
- Li, C. (2023). Exploring L2 Motivational Dynamics Among Chinese EAP Learners in an EMI Context From a Socio-cultural Perspective. *Sage Open*, 13(2), 21582440231172704. <https://doi.org/10.1177/21582440231172704>
- Li, F.-Y., Hwang, G.-J., Chen, P.-Y., & Lin, Y.-J. (2021). Effects of a concept mapping-based two-tier test strategy on students' digital game-based learning performances and behavioral patterns. *Computers & Education*, 173, 104293. <https://doi.org/10.1016/j.compedu.2021.104293>
- Mahbub, I. S. P., & Hadina, H. (2021). A systematic overview of issues for developing EFL learners' oral English communication skills. *Journal of Language and Education*, 7(1), 229–240. <https://doi.org/10.17323/jle.2021.10737>
- Mahn, H. (1999). Vygotsky's methodological contribution to sociocultural theory. *Remedial and Special Education*, 20(6), 341–350. <https://doi.org/10.1177/074193259902000607>
- Mokmin, N. A. M., Hanjun, S., Jing, C., & Qi, S. (2024). Impact of an AR-based learning approach on the learning achievement, motivation, and cognitive load of students on a design course. *Journal of Computers in Education*, 11(2), 557–574. <https://doi.org/10.1007/s40692-023-00270-2>
- Mustakim, N., Setyawan, C. E., Hunaenah, H., & Masuwd, M. (2024). CEFR-based maharah qiraah learning interactive test using the Quizziz App. *Alsuna: Journal of Arabic and English Language*, 7(1), 145–160. <https://doi.org/10.31538/alsuna.v7i1.5086>
- Nithideechaiwarachok, B., & Chano, J. (2024). Socio-cultural and social constructivist theories and its application in EFL classroom for Thai pre-service teachers: A review for further research. *International Journal of Language Education*, 8(3), 564–572. <https://doi.org/10.26858/ijole.v8i3.66499>
- Osiemo, J., Cecchi, F., Bulte, E., & Mwongera, C. (2026). Experiential learning, narrative-based learning, and insurance adoption: Experimental evidence from Kenya. *American Journal of Agricultural Economics*, 108(1), 232–253. <https://doi.org/10.1111/ajae.12528>
- Read, T., & Rodriguez Salgado, C. (2025). Hybrid scaffolded reading to foster engagement and comprehension in ESL learners. *Forum for Linguistic Studies*, 7(11), 1386–1401. <https://doi.org/10.30564/fls.v7i11.9389>
- Riazi, A. M., Ghanbar, H., Marefat, F., & Fazel, I. (2023). Review and analysis of empirical articles published in TESOL Quarterly over its lifespan. *Studies in Second Language Learning and Teaching*, 13(4), 811–841. <https://doi.org/10.14746/ssllt.40217>
- Shadiev, R., & Yu, J. (2024). Review of research on computer-assisted language learning with a focus on intercultural education. *Computer Assisted Language Learning*, 37(4), 841–871. <https://doi.org/10.1080/09588221.2022.2056616>



- Smith, L. M., Villar, F., & Wendel, S. (2023). Narrative-based learning for person-centred healthcare: The caring stories learning framework. *Medical Humanities*, 49(4), 583–592. <https://doi.org/10.1136/medhum-2022-012530>
- Sugiyono. (2017). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Syafei, I., Suleman, E., & Rohanda, R. (2024). The Development of Student Reading Skills in Arabic for Reading Islamic Classical Books Using the Arabic Learning Model at Indonesian Islamic Boarding Schools. *Theory and Practice in Language Studies*, 14(5), 1381–1392. <https://doi.org/10.17507/tppls.1405.10>
- Syukran, S., Bengi, M. S., & Husna, J. M. (2022). Maharat al-qira'ah fi ta'lim al-lughah al-'Arabiyyah. *Kitabatuna: Jurnal Pendidikan Bahasa Arab Dan Kebahasaan*, 1(1), 11–30.
- Tibi, S., Fitton, L., & McIlraith, A. L. (2021). The development of a measure of orthographic knowledge in the Arabic language: A psychometric evaluation. *Applied Psycholinguistics*, 42(3), 739–762. <https://doi.org/10.1017/S0142716421000047>
- Vygotsky, L., & Cole, M. (2018). Lev Vygotsky: Learning and social constructivism. In *Learning theories for early years practice* (pp. 68–73). SAGE. <https://doi.org/10.4324/9781003331193-16>
- Wu, L., & Bakar, K. A. (2025). Language scaffolding strategies based on Vygotsky's Zone of Proximal Development in early reading education: A cross-cultural comparison. *Forum for Linguistic Studies*, 7(6), 967–984. <https://doi.org/10.30564/fls.v7i6.9624>
- Zikriah, Z., & Mauludiyah, L. (2024). Research trend of Arabic language teaching in the world: Systematic literature review based on Scopus database. *Research and Development in Education (RaDEn)*, 4(1), 27–48. <https://doi.org/10.22219/raden.v4i1.31847>